



**UNIVERSIDAD NACIONAL DE LA PLATA**

**FACULTAD DE CIENCIAS EXACTAS**

**DEPARTAMENTO DE QUÍMICA**

---

***Trabajo de Tesis Doctoral***

***ESTUDIO Y APLICACIÓN DE TÉCNICAS MATEMÁTICAS  
DE LA TEORÍA QSAR-QSPR EN QUÍMICA ANALÍTICA Y  
QUÍMICA DE LOS ALIMENTOS***

---

***Tesista: Cristian Xavier Rojas Villa***

***Director: Reinaldo Pis Diez***

***Codirector: Pablo R. Duchowicz***

**2017**



El presente trabajo de Tesis se desarrolló en el Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), dependiente del Departamento de Química de la Facultad de Ciencias Exactas de la Universidad Nacional de la Plata (UNLP).

Se presenta en consideración de las autoridades de dicha Facultad para acceder al grado académico de Doctor de la Facultad de Ciencias Exactas Área Química.



# AGRADECIMIENTOS

Deseo expresar mi agradecimiento a todas las personas que hicieron posible el desarrollo de la presente tesis doctoral:

A mis directores Reinaldo Pis Diez y Pablo R. Duchowicz, por dirigir el trabajo de tesis y por su invaluable enseñanza y aporte tanto en lo científico como en lo personal.

A todas las personas que he conocido en el INIFTA, mi lugar de trabajo, particularmente a los integrantes del Grupo de Estudio Teórico de Sistemas Químicos, Físicos y Biológicos: Eduardo A. Castro, Francisco M. Fernández, Pablo R. Duchowicz, Andrew G. Mercader, Ofelia Oña, Javier García y José F. Aranda, con quienes he compartido gratos momentos en el transcurso de estos años.

A Roberto Todeschini y su Grupo de Quimiometría y QSAR de la Universidad de Milán–Bicocca: Davide Ballabio, Andrea Mauri y Viviana Consonni, por la cálida acogida durante mi estancia doctoral de investigación, por todo el soporte científico y acceso a los diversos programas quimioinformáticos que facilitaron el desarrollo del trabajo. También a las personas con quienes compartí en el laboratorio: Matteo Cassotti, Francesca Grisoni, Serena Nembri, y Agustín Herrera. Asimismo, al Ministerio de Asunto Exteriores y Cooperación Internacional del Gobierno Italiano (FARNESINA) por la beca otorgada.

A las autoridades de la Universidad del Azuay por el permiso institucional y apoyo económico brindado dentro del programa de formación y capacitación del recurso humano.

A Piercosimo Tripaldi por haberme inmerso en el mundo del análisis multivariado de los datos y enseñarme desde joven el valor de investigar, siendo un constante respaldo en todos los trabajos científicos que hemos realizado. Asimismo a Andrés Pérez, con quien he discutido y ejecutado algunas ideas de investigación durante este tiempo.

A los integrantes del Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB) de la Universidad Nacional de La Plata: Lucas Alberca, Juan Francisco Morales, Alan Talevi, Luciana Gavernet y Carolina Carrillo; con quienes tuve el placer de intercambiar varias ideas de investigación durante el curso de Posgrado Descubrimiento de Fármacos Asistido por Computadora y el Workshop Internacional de Diseño de Fármacos y Enfermedades Tropicales Desatendidas.

A Wayne Hanson, por su desinteresada contribución en la revisión técnica del inglés de todos los trabajos científicos y por todas sus valiosas sugerencias para mejorar la calidad técnica de los mismos. De igual manera, a Adrián Rojas por el valioso aporte en la edición de figuras de los artículos científicos, pósters y de la presente tesis.

A mis amigos Francesco Piazza y Elisa Accornero por su hospitalidad durante mi estadía en su departamento. De igual manera, a mis amigos Diego Suárez, Fernando Cárdenas y Juan de Dios Alvarado, quienes

durante estos años de formación me han animado a seguir adelante y no decaer. A mis amigos de la Universidad Nacional de Córdoba: Rommel, Ricardo, Viviana y Jorge, así como a todos los demás amigos a quienes he conocido durante estos años y con quienes he compartido inolvidables momentos y reuniones, particularmente a José, Irene, Christian, Elizabeth, Santiago, Salvador, Alejandro, Julie, Zuly, Efrén, Nasly.

A la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación de la República del Ecuador (SENESCYT), por la Beca Doctoral otorgada dentro del Programa “Convocatoria Abierta 2013 Primera Fase”.

Finalmente, mi mayor gratitud va para mi esposa Fer y mi familia: Mariana, María Fernanda, Freddy, Marcelo, José Alejandro, Francisco, Blanca, Rodrigo, David y Miguel. Ustedes han sido mi apoyo en todo sentido durante estos años y han entendido mi ausencia en momentos importantes de sus vidas.

*Esta tesis está dedicada con profundo  
cariño a toda mi familia*



*A la memoria de mi padre José Tomás*



*Un buen científico debe tener la imaginación de un niño, la determinación de un joven, la racionalidad de un adulto y la experiencia de un anciano. La dificultad radica en tener todas estas cualidades al mismo tiempo*

*Roberto Todeschini en Molecular Descriptors for Chemoinformatics: Wiley-VCH (2009)*



# ÍNDICE

|  |            |
|--|------------|
| <b>LISTA DE ABREVIATURAS</b> .....   | <b>i</b>   |
| <b>MATERIAL ANEXO</b> .....  | <b>iii</b> |
| <b>CAPÍTULO 1: RELACIONES CUANTITATIVAS ESTRUCTURA-<br/>ACTIVIDAD/PROPIEDAD (QSAR/QSPR)</b>    |            |
| 1.1    Introducción .....  | 1          |
| 1.2    Definición y formalismo.....  | 2          |
| 1.3    Objetivos de un modelo QSAR.....  | 4          |
| 1.4    Aplicaciones de un modelo QSAR.....   | 4          |
| 1.4.1.    Naturaleza de la respuesta.....  | 4          |
| 1.4.2.    Naturaleza de los compuestos químicos.....   | 5          |
| 1.4.3.    Áreas de aplicación .....  | 5          |
| 1.5    Componentes y etapas de un modelo QSAR.....   | 5          |
| 1.5.1.    Preparación de los datos.....  | 6          |
| 1.5.2.    Procesamiento de los datos .....   | 6          |
| 1.5.3.    Validación y predicción de los datos .....   | 7          |
| 1.5.4.    Interpretación de los datos .....  | 7          |
| 1.6    Enfoques específicos QSAR .....   | 7          |
| 1.6.1.    Enfoque de Hansch .....  | 7          |
| 1.6.2.    Enfoque de Free–Wilson .....   | 8          |
| 1.6.3.    Métodos de Contribución de Grupos.....   | 8          |
| 1.6.4.    Cribado de alto rendimiento .....  | 9          |
| 1.6.5.    Barrido virtual .....  | 9          |
| 1.7    Principios del modelado QSAR .....  | 10         |
| 1.7.1.    Actividad/propiedad definida .....   | 10         |
| 1.7.2.    Algoritmo inequívoco .....   | 10         |
| 1.7.3.    Dominio de aplicabilidad definido .....  | 11         |
| 1.7.4.    Medida apropiada de la bondad de ajuste, robustez y<br>predictividad.....            | 11         |
| 1.7.5.    Interpretación del mecanismo de acción de los descriptores (de<br>ser posible) ..... | 11         |
| 1.8    Referencias.....  | 12         |
| <b>CAPÍTULO 2: DESCRIPTORES MOLECULARES</b>  |            |
| 2.1    Introducción .....  | 15         |
| 2.2    Historia.....   | 16         |
| 2.3    Definición y requisitos .....   | 17         |

|         |  |    |
|---------|--|----|
| 2.4     | Representación de la estructura molecular .....  | 19 |
| 2.4.1.  | Grafos moleculares .....   | 20 |
| 2.4.2.  | Representación matricial .....   | 21 |
| 2.5     | Codificación de la estructura molecular .....  | 23 |
| 2.6     | Curado de las estructuras moleculares .....  | 24 |
| 2.6.1.  | KNIME .....  | 25 |
| 2.7     | Principales tipos de descriptores moleculares.....   | 26 |
| 2.7.1.  | Descriptores constitucionales .....  | 26 |
| 2.7.2.  | Descriptores de anillo .....   | 27 |
| 2.7.3.  | Índices topológicos .....  | 27 |
| 2.7.4.  | Descriptores tipo P_VSA .....  | 30 |
| 2.7.5.  | Propiedades moleculares .....  | 30 |
| 2.7.6.  | Fragmentos centrados en el átomo y número de grupos<br>funcionales.....                        | 31 |
| 2.7.7.  | Descriptores geométricos .....   | 31 |
| 2.7.8.  | Descriptores topo-geométricos .....  | 34 |
| 2.8     | Descriptores para moléculas desconectadas .....  | 35 |
| 2.9     | Huellas digitales moleculares .....  | 36 |
| 2.9.1.  | Huellas digitales moleculares de conectividad ampliada.....                                    | 37 |
| 2.10    | Programas para el cálculo de descriptores moleculares y huellas<br>digitales moleculares ..... | 39 |
| 2.10.1. | DRAGON .....   | 40 |
| 2.11    | Referencias .....  | 43 |

### **CAPÍTULO 3: MÉTODOS QUIMIOMÉTRICOS EN QSAR/QSPR**

|        |  |    |
|--------|--|----|
| 3.1    | Introducción .....   | 49 |
| 3.2    | Estructura multivariada de los datos.....                  | 50 |
| 3.3    | Pretratamiento de los datos .....                          | 50 |
| 3.3.1. | Escalado de los datos .....                                | 51 |
| 3.4    | Técnicas de exploración de la estructura de los datos..... | 52 |
| 3.4.1. | Análisis de componentes principales .....                  | 52 |
| 3.4.2. | Escalado multidimensional .....                            | 52 |
| 3.4.3. | Análisis de conglomerados.....                             | 53 |
| 3.5    | Técnicas de reducción de variables .....                   | 54 |
| 3.5.1. | Correlación entre pares de variables.....                  | 54 |
| 3.5.2. | Método V-WSP .....   | 55 |
| 3.5.3. | Otros métodos basados en técnicas multivariable .....      | 55 |

|        |   |    |
|--------|---|----|
| 3.6    | Técnicas de selección de variables .....    | 55 |
| 3.6.1. | Búsqueda exacta .....                       | 55 |
| 3.6.2. | Métodos de regresión a pasos .....          | 56 |
| 3.6.3. | Algoritmos Genéticos .....                  | 56 |
| 3.6.4. | Método de Reemplazo .....                   | 57 |
| 3.7    | Técnicas de modelado de datos .....         | 58 |
| 3.7.1. | Métodos de regresión .....                  | 58 |
| 3.7.2. | Métodos de clasificación .....              | 63 |
| 3.7.3. | Análisis de consenso .....                  | 68 |
| 3.8    | Técnicas de validación .....                | 68 |
| 3.8.1. | Validación cruzada o interna .....          | 69 |
| 3.8.2. | Validación externa .....                    | 72 |
| 3.9    | Dominio de Aplicabilidad .....              | 73 |
| 3.9.1. | AD basado en el valor de influencia .....   | 74 |
| 3.9.2. | AD basado en la similitud <i>k</i> NN ..... | 75 |
| 3.10   | Referencias .....                           | 76 |

#### **CAPÍTULO 4: APLICACIONES EN QUÍMICA ANALÍTICA**

|        |  |     |
|--------|--|-----|
| 4.1    | Modelo QSPR para índices de retención medidos en la columna apolar OV-101 .....                      | 82  |
| 4.1.1. | Introducción .....   | 82  |
| 4.1.2. | Materiales y métodos .....   | 82  |
| 4.1.3. | Resultados y discusión .....   | 84  |
| 4.1.4. | Conclusiones .....   | 90  |
| 4.2    | Modelo QSPR para índices de retención medidos en la columna polar Carbowax 20M.....                  | 90  |
| 4.2.1. | Introducción .....   | 90  |
| 4.2.2. | Materiales y métodos .....   | 91  |
| 4.2.3. | Resultados y discusión .....   | 91  |
| 4.2.4. | Conclusiones .....   | 100 |
| 4.3    | Modelo QSPR para índices de retención de aromas medidos en la columna polar DB-225MS.....            | 100 |
| 4.3.1. | Introducción .....   | 100 |
| 4.3.2. | Materiales y métodos .....   | 100 |
| 4.3.3. | Resultados y discusión.....  | 103 |
| 4.3.4. | Conclusiones .....   | 106 |
| 4.4    | Modelo QSPR para índices de retención de aromas medidos en la columna ligeramente polar HP5-MS ..... | 107 |

|        |   |     |
|--------|---|-----|
| 4.4.1. | Introducción .....  | 107 |
| 4.4.2. | Materiales y métodos .....  | 109 |
| 4.4.3. | Resultados y discusión .....  | 109 |
| 4.4.4. | Conclusiones .....  | 112 |
| 4.5    | Modelo QSPR para índices de retención de aromas medidos en la columna apolar HP-1 ..... | 112 |
| 4.5.1. | Introducción .....  | 112 |
| 4.5.2. | Materiales y métodos .....  | 113 |
| 4.5.3. | Resultados y discusión .....  | 114 |
| 4.5.4. | Conclusiones .....  | 116 |
| 4.6    | Referencias.....  | 117 |

## **CAPÍTULO 5: APLICACIONES EN QUÍMICA DE LOS ALIMENTOS**

|        |  |     |
|--------|--|-----|
| 5.1    | Modelo QSPR basado en índices de retención para el control de calidad de arroz ..... | 125 |
| 5.1.1. | Introducción .....   | 125 |
| 5.1.2. | Materiales y métodos .....   | 126 |
| 5.1.3. | Resultados y discusión .....   | 129 |
| 5.1.4. | Conclusiones .....   | 135 |
| 5.2    | Modelo QSPR para el dulzor relativo .....  | 135 |
| 5.2.1. | Introducción .....   | 135 |
| 5.2.2. | Materiales y métodos .....   | 137 |
| 5.2.3. | Resultados y discusión .....   | 139 |
| 5.2.4. | Conclusiones .....   | 143 |
| 5.3    | Modelo QSAR para discriminar los gustos dulce y amargo.....                          | 144 |
| 5.3.1. | Introducción .....   | 144 |
| 5.3.2. | Materiales y métodos .....   | 145 |
| 5.3.3. | Resultados y discusión .....   | 147 |
| 5.3.4. | Conclusiones .....   | 151 |
| 5.4    | Modelo QSAR para discriminar los gustos dulce e insípido .....                       | 151 |
| 5.4.1. | Introducción .....   | 151 |
| 5.4.2. | Materiales y métodos .....   | 152 |
| 5.4.3. | Resultados y discusión .....   | 152 |
| 5.4.4. | Conclusiones .....   | 155 |
| 5.5    | Modelo QSAR basado en un sistema experto para predecir el dulzor .....               | 156 |
| 5.5.1. | Introducción .....   | 156 |
| 5.5.2. | Materiales y métodos .....   | 157 |

|  |            |
|--|------------|
| 5.5.3. Resultados y discusión .....  | 161        |
| 5.5.4. Conclusiones .....  | 171        |
| 5.6 Referencias.....   | 172        |
| <b>CONCLUSIONES GENERALES Y PROYECCIONES FUTURAS .....</b>   | <b>181</b> |
| <b>PUBLICACIONES Y TRABAJOS PRESENTADOS EN EVENTOS CIENTÍFICOS .....</b>   | <b>185</b> |
| <b>TABLAS ANEXAS (en CD) .....</b>   | <b>189</b> |
| Tabla 1A. Nombres químicos e índices de retención experimentales de 1208 sabores y aromas medidos en la columna apolar OV-101 .....                                | 189        |
| Tabla 2A. Nombres químicos e índices de retención experimentales de 1184 sabores y aromas medidos en la columna polar Carbowax 20M .....                           | 220        |
| Tabla 3A. Nombres químicos, número de registro CAS e índices de retención experimentales de las 269 fragancias medidas en la columna polar DB-225MS .....          | 250        |
| Tabla 4A. Nombres químicos, número de registro CAS e índices de retención experimentales de las 266 fragancias medidas en la columna ligeramente polar HP5-MS..... | 257        |
| Tabla 5A. Nombres químicos, número de registro CAS e índices de retención experimentales de las 262 fragancias medidas en la columna apolar HP-1 .....             | 264        |
| Tabla 6A. Nombres químicos, número de registro CAS e índices de retención experimentales de 137 compuestos volátiles de arroz medidos en SPME/GC/MS .....          | 271        |
| Tabla 7A. Nombres químicos y dulzor relativo experimental de 233 compuestos dulcificantes.....   | 275        |
| Tabla 8A. Nombres químicos y respuesta experimental de los 508 compuestos dulces y amargos .....   | 282        |
| Tabla 9A. Nombres químicos y respuesta experimental de los 566 compuestos dulces e insípidos .....   | 293        |
| Tabla 10A. Nombres químicos y respuesta experimental de los 649 compuestos dulces y no dulces.....   | 305        |

---

# LISTA DE ABREVIATURAS

---

|          |  |
|----------|--|
| 3D-MoRSE | Representación molecular 3D de estructuras basada en la difracción de electrones |
| AC       | Exactitud  |
| AD       | Dominio de aplicabilidad   |
| ADME     | Absorción, distribución, metabolismo y eliminación                               |
| ANN      | Red neuronal artificial  |
| BE       | Eliminación hacia atrás  |
| BNN      | Vecinos más cercanos agrupados   |
| BPANN    | red neuronal artificial de retropropagación                                      |
| BSM      | Método de subconjuntos balanceados   |
| CART     | Árboles de clasificación y regresión   |
| CATS     | Búsqueda de plantillas químicamente avanzadas                                    |
| CoMFA    | Análisis comparativo de campos moleculares                                       |
| CoMSIA   | análisis comparativo del índice de similitud molecular                           |
| CIR      | Solucionador de identificación química   |
| DA       | Análisis discriminante   |
| ECFPs    | Huellas digitales moleculares de conectividad ampliada                           |
| FPs      | Huellas digitales moleculares  |
| FS       | Selección hacia adelante   |
| GAs      | Algoritmos genéticos   |
| GC       | Cromatografía de gases   |
| GC-MS    | Cromatografía de gases-espectrometría de masa                                    |
| GCM      | Métodos de contribución de grupos  |
| GETAWAY  | Descriptor de ensamblado de pesos de átomos, geometría y topología               |
| GFA      | Algoritmo funcional genético   |
| GPCR     | Receptor acoplado a proteínas G  |
| HOMO     | Orbital molecular de más alta energía ocupado                                    |
| HTS      | Cribado de alto rendimiento  |
| I        | Índice de retención cromatográfico   |
| kNN      | k-vecinos más cercanos   |
| L-M ANN  | red neuronal artificial de Levenberg-Marquardt                                   |
| LDA      | Análisis discriminante lineal  |
| LLA      | Máquina de aprendizaje lineal  |
| LMO      | Dejar-varios-fuera   |
| LOO      | Dejar-uno-fuera  |
| LSE      | relaciones lineales para la energía de solvatación                               |
| LUMO     | Orbital molecular no ocupado de más baja energía                                 |
| LVs      | Variables latentes   |
| MCIs     | Índices de conectividad molecular  |
| MDS      | Escalado multidimensional  |
| MLR      | Regresión lineal múltiple  |
| MPA      | Teoría de enlace multipunto  |
| MSS      | Suma de los cuadrados del modelo   |
| N3       | N-vecinos más cercanos   |
| NER      | Tasa de aciertos   |
| OECD     | Organización para la Cooperación Económica y el Desarrollo                       |
| PAHs     | hidrocarburos aromáticos policíclicos  |
| PCA      | Análisis de componentes principales  |

|       |   |
|-------|---|
| PSO   | optimización por enjambre de partículas                     |
| PPP   | Presencia de potenciales farmacóforos                       |
| OLS   | Mínimos cuadrados ordinarios                                |
| PLS   | Mínimos cuadrados parciales                                 |
| PLSDA | Análisis discriminante de mínimos cuadrados parciales       |
| PRESS | Suma de los cuadrados de predicción                         |
| QAAR  | Relaciones cuantitativas actividad–actividad                |
| QDA   | Análisis discriminante cuadrático                           |
| QPAR  | Relaciones cuantitativas propiedad–actividad                |
| QPPR  | Relaciones cuantitativas propiedad–propiedad                |
| QSAR  | Relación cuantitativa estructura–actividad                  |
| QSPR  | Relación cuantitativa estructura–propiedad                  |
| QSRR  | Relaciones cuantitativas estructura–retención               |
| QSRSR | Relaciones cuantitativas estructura–dulzor relativo         |
| QSTR  | Relaciones cuantitativas estructura–toxicidad               |
| QSCR  | Relaciones cuantitativas estructura–citotoxicidad           |
| QShAR | Relaciones cuantitativas forma–actividad                    |
| QSiAR | Relaciones cuantitativas similitud–actividad                |
| QSRR  | Relaciones cuantitativas estructura–retención               |
| QSRR  | Relaciones cuantitativas estructura–reactividad             |
| QTTR  | Relaciones cuantitativas toxicidad–toxicidad                |
| RDF   | Función de distribución radial                              |
| RF    | bosques al azar   |
| RM    | Método de reemplazo   |
| RS    | Dulzor relativo   |
| RSS   | Suma de cuadrados de los residuos                           |
| SIMCA | Modelado suave independiente por analogía de clases         |
| SVM   | Máquinas de soporte vectorial                               |
| SWR   | Regresión a pasos   |
| TIs   | Índices topológicos   |
| TSS   | Suma total de los cuadrados                                 |
| VS    | Barrido virtual   |
| VSA   | Área superficial de van der Waals                           |
| VSI   | Identificador del estado de valencia                        |
| V–WSP | Método de reducción de variables basado en el algoritmo WSP |
| WHIM  | Descriptores moleculares invariantes holísticos ponderados  |

---

## **MATERIAL ANEXO**

---

En el CD que se adjunta al ejemplar de tesis se encuentran disponibles todas las Tablas que detallan las bases de datos utilizadas en las diversas aplicaciones QSAR/QSPR. Asimismo, se incluye una copia de todos los artículos científicos publicados.

# Capítulo 1

---

## RELACIONES CUANTITATIVAS ESTRUCTURA–ACTIVIDAD/PROPIEDAD (QSAR/QSPR)

---

### 1.1 Introducción

Las relaciones cuantitativas estructura–actividad/propiedad (QSAR/QSPR) (Hansch *et al.*, 1995; Todeschini & Consonni, 2009; Dearden, 2016), también denominados estudios *in silico*, nacieron a inicios de los años sesenta, con el trabajo de Corwin Hansch *et al.* Con su trabajo estimularon un gran interés en la predicción de propiedades, especialmente en el área de la química medicinal y el diseño de fármacos. Sin embargo, la teoría QSAR/QSPR ha evolucionado a través de los años a partir de modelos de regresión simple basados en pocos descriptores, los cuales eran seleccionados de manera intuitiva, hasta transformarse en una herramienta importante de amplia aplicación, particularmente en las ciencias químicas, biológicas, toxicológicas, ambientales, medicinales, farmacológicas y de alimentos. En la actualidad, la metodología QSAR/QSPR continúa en desarrollo y expansión, como lo indica la gran variedad de métodos QSAR modernos que han sido propuestos para la predicción de las propiedades/actividades de compuestos desconocidos, generando una fuerte tendencia al incremento en el número de publicaciones en este campo de investigación (Galvez & Garcia-Domenech, 2010; Gangwal *et al.*, 2016).

La metodología QSAR busca predecir las actividades/propiedades de las moléculas, con las cuales se podrá realizar inferencias sobre información del fenómeno o mecanismo involucrado. Los estudios QSAR complementan otros estudios teóricos o experimentales que buscan elucidar racionalmente las interrogantes de tipo químico involucradas en dichos problemas (Kubinyi, 2008). En la actualidad el desarrollo de modelos QSAR predictivos ha sido propuesto por diferentes organismos internacionales como una herramienta útil para explorar mediante técnicas racionales la información contenida en la estructura química de las moléculas (Todeschini *et al.*, 2009).

El estudio sobre relaciones estructura/actividad publicado en 1962 por Hansch, Maloney y Fujita, estableció la dependencia de los reguladores

de crecimiento de plantas con las constantes de Hammett y la hidrofobicidad. Este enfoque, en el que se relacionan las propiedades fisicoquímicas con las actividades biológicas, se lo denominó “*Análisis de Hansch*” (Hansch *et al.*, 1995; Franke & Gruska, 2003). En 1964 Free y Wilson presentaron un modelo de contribuciones aditivas de los sustituyentes químicos para actividades biológicas en base a la presencia/ausencia de los mismos. Este modelo se lo llamó “*Enfoque de novo*” y se basa en el supuesto de que cada sustituyente en la molécula brinda un efecto aditivo y constante a la actividad biológica, independientemente de la presencia de los demás sustituyentes (Kubinyi, 1988; Franke & Gruska, 2003). Estos dos trabajos pioneros son los que sentaron las bases para el desarrollo del campo de conocimiento que actualmente se conoce como QSAR/QSPR.

## 1.2 Definición y formalismo

El modelado QSAR se refiere al desarrollo de correlaciones matemáticas entre una respuesta (actividad/propiedad/toxicidad) y atributos químicos definidos (descriptores moleculares, huellas digitales moleculares) de las moléculas analizadas. El nombre específico que toma el estudio depende de la naturaleza de la respuesta a ser modelada; así, se pueden diferenciar dos grandes clases, denominadas relaciones cuantitativas estructura–actividad/propiedad (QSAR/QSPR). Sin embargo, esta nomenclatura también se puede aplicar cuando se estudian respuestas más específicas, por ejemplo (Kaliszan, 2007; Todeschini & Consonni, 2009; Roy *et al.*, 2015a; Rojas *et al.*, 2016):

- Relaciones cuantitativas estructura–retención (QSRR).
- Relaciones cuantitativas estructura–dulzor relativo (QSRSR).
- Relaciones cuantitativas estructura–toxicidad (QSTR).
- Relaciones cuantitativas estructura–citotoxicidad (QSCR).
- Relaciones cuantitativas estructura–reactividad (QSRR).
- Relaciones cuantitativas propiedad–propiedad (QPPR).
- Relaciones cuantitativas propiedad–actividad (QPAR).
- Relaciones cuantitativas actividad–actividad (QAAR).
- Relaciones cuantitativas toxicidad–toxicidad (QTTR).
- Relaciones cuantitativas similitud–actividad (QSiAR).
- Relaciones cuantitativas forma–actividad (QShAR).

El término QSAR se usa frecuentemente para referirse a todos estos tipos específicos de estudios, los cuales permiten predecir la actividad de las moléculas (nuevas o hipotéticas) en función de características moleculares de las mismas. Así, el formalismo básico de la teoría QSAR se puede representar matemáticamente de la siguiente manera (Todeschini & Consonni, 2009; Hongmao, 2015; Roy *et al.*, 2015a):

$$\text{Actividad / Propiedad Biológica} = f(\text{Estructura Química}) \quad (1.1)$$

Las propiedades fisicoquímicas son aquellos atributos de las moléculas que se extraen de forma experimental (punto de fusión, tiempo de degradación, etc.); mientras que la estructura química denota a la información que se extrae de forma teórica (descriptores moleculares y/o huellas digitales moleculares). En la Figura 1.1. se presenta el rol de los descriptores moleculares y/o huellas digitales moleculares en la formulación de modelos matemáticos QSAR/QSPR.

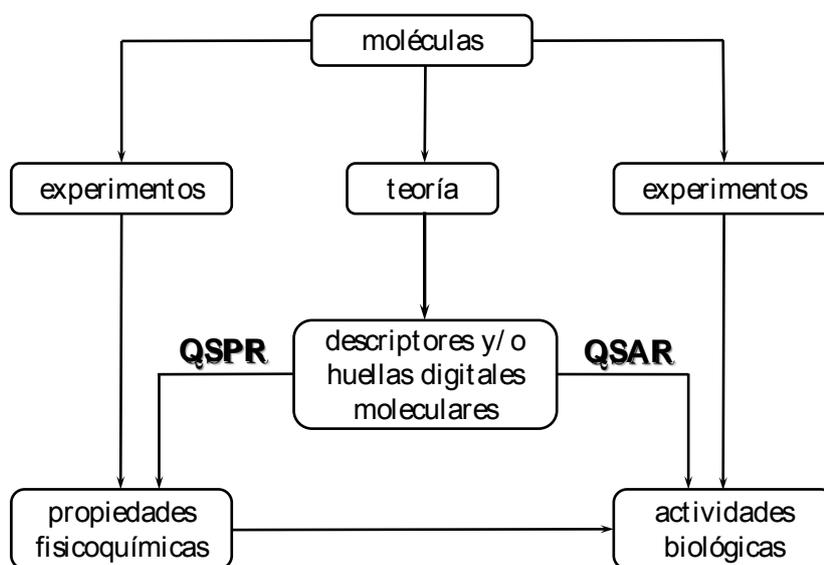


Figura 1.1. Esquema general de un estudio QSAR/QSPR

El formalismo presentado en la Ec. 1.1 se puede aplicar tanto a modelos continuos (regresión) como a modelos discretos (clasificación) (Chaudhry *et al.*, 2007; Todeschini & Consonni, 2009; Hongmao, 2015; Liaw & Svetnik, 2015). En los modelos de regresión, la actividad/propiedad es una variable cuantitativa continua; es decir, toma cualquier valor dentro de la escala de medida de dicha respuesta; por ejemplo el dulzor relativo o el índice de retención. Por otro lado, los modelos de clasificación consideran a la respuesta como una variable cualitativa nominal; es decir, la actividad se presenta en forma de categorías no ordenadas, por ejemplo, compuestos dulces (clase 1), compuestos amargos (clase 2) y compuestos insípidos (clase 3). En clasificación, el balance entre las clases tiene gran influencia en la calidad del modelo QSAR/QSPR, debido a que los modelos tienden a sesgarse hacia la clase más numerosa. Por este motivo, es común que el grupo de calibración sea formado tomando en consideración las numerosidad de las clases (Hongmao, 2015).

### 1.3 Objetivos de un modelo QSAR

El objetivo principal de cualquier estudio QSAR es el desarrollo racional de un modelo matemático predictivo acompañado de una interpretación de la información química involucrada. Dicho modelo puede ser desarrollado a partir de los compuestos para los cuales se disponga de una determinada respuesta de actividad/propiedad y permite realizar la predicción de la actividad/propiedad para un número mayor de moléculas (Hamzeh-Mivehroud *et al.*, 2015). Por lo tanto, el desarrollo de modelos QSAR es de utilidad en procesos de investigación; así como en diversos campos donde es importante la predicción de actividades/propiedades de los compuestos químicos (Cronin, 2010; Roy *et al.*, 2015a). El formalismo de la teoría QSAR también tiene otros objetivos que se enlistan a continuación (Cronin, 2010; Roy *et al.*, 2015b):

1. Predicción de una actividad/propiedad química de interés.
2. Reducción y reemplazo de la experimentación de laboratorio usando animales. La experimentación normalmente es larga y costosa.
3. Barrido virtual de bibliotecas de datos.
4. Comprensión de los mecanismos de acción dentro de un grupo de sustancias químicas.
5. Optimización de moléculas con actividades/propiedades deseadas y optimización de la síntesis química (minimizar la eliminación de desechos y reducir el costo).
6. Refinamiento estructural de moléculas objetivo sintéticas.
7. Identificación de compuestos peligrosos en las etapas iniciales del diseño de los mismos.
8. Predicción de la toxicidad de los compuestos en seres humanos mediante la exposición ocasional u ocupacional. Así también, la predicción de la toxicidad sobre especies ambientales.
9. Aplicaciones con fines regulatorios por parte de agencias gubernamentales.

### 1.4 Aplicaciones de un modelo QSAR

El modelado QSAR es una opción adecuada para el monitoreo racional de la actividad o propiedad de los compuestos químicos; por lo que resulta de utilidad para diversos campos científicos. Las áreas de aplicación de la teoría QSAR han crecido a lo largo del tiempo y son muy variadas. De esta forma, las aplicaciones de los modelos QSAR se agrupan de tres maneras distintas (Roy *et al.*, 2015a):

#### 1.4.1. Naturaleza de la respuesta

Desde el punto de vista de la naturaleza de la respuesta se pueden agrupar en actividades, propiedades y toxicidades. Ejemplos de actividades biológicas son las antioxidantes, antibacteriales, antimalaria, antihipertensivas, antihistamínicos, antiepilépticos, antidepresivos, anti-VIH, antidiuréticos, etc. Por otro lado, las propiedades incluyen: coeficiente de partición octanol/agua, temperatura de transición vítrea, índices de retención en cromatografía, hidrólisis, oxidación atmosférica, biodegradación, bioacumulación, etc. Finalmente, las toxicidades: aguda en peces, acuática persistente, aguda por inhalación, aguda oral, de órganos (por ejemplo, hepatotoxicidad, cardiotoxicidad, nefrotoxicidad), irritación/corrosión de la piel y ojos, carcinogenicidad, etc.

### **1.4.2. Naturaleza de los compuestos químicos**

Considerando la naturaleza de los compuestos químicos se pueden diferenciar entre compuestos químicos de beneficio para la salud (fármacos y aditivos alimentarios), compuestos químicos usados en procesos industriales y de laboratorio (solventes, perfumería, surfactantes y reactivos en general) y compuestos químicos que tienen efectos nocivos para la salud (pesticidas, toxinas, xenobióticos, agentes carcinógenos, compuestos orgánicos volátiles, contaminantes orgánicos persistentes, etc.).

### **1.4.3. Áreas de aplicación**

Las áreas de aplicación de la teoría QSAR pueden subdividirse en tres tipos: diseño de fármacos, ciencia de los materiales y toxicología predictiva. El diseño de fármacos involucra el modelado ADME (Absorción, Distribución, Metabolismo y Excreción), el cual permite monitorear el perfil farmacocinético de un potencial fármaco antes de su síntesis, permitiendo mejorar la eficacia en el diseño de compuestos dentro de un sistema biológico. Por otra parte, el modelado QSAR se aplica en ciencia de materiales para el estudio de polímeros, líquidos iónicos, catálisis, nanomateriales, fullerenos, surfactantes, biomateriales, cerámicos, etc. Finalmente, la toxicología predictiva comprende la evaluación de la toxicidad sistémica (toxicidad sobre órganos específicos) y el control de riesgos ecotoxicológicos (daños en el ecosistema).

## **1.5 Componentes y etapas de un modelo QSAR**

El modelado QSAR consiste en datos cuantitativos y el uso de técnicas quimiométricas apropiadas. Los datos cuantitativos de los compuestos químicos se obtienen a partir de dos fuentes: 1) medición de la actividad/propiedad de interés e 2) información química codificada en los descriptores moleculares, que se obtienen mediante diversas

teorías que se aplican a la representación de las moléculas. Debido a la gran cantidad de información que se obtiene, el uso de computadores es necesario para llevar a cabo el desarrollo de los modelos QSAR. Existen 4 pasos básicos que se siguen en los estudios QSAR (Roy *et al.*, 2015b; Golbraikh *et al.*, 2017):

### **1.5.1. Preparación de los datos**

Los datos deben ser ordenados de una manera conveniente y útil. Para iniciar un estudio QSAR es necesario poseer un conjunto de compuestos químicos con una respuesta de interés (actividad/propiedad). Algunos ejemplos de actividades/propiedades son: EC50 (concentración efectiva en el 50% de la población), IC50 (concentración requerida para el 50% de inhibición), LD50 (dosis requerida para matar la mitad del total de la población, coeficiente de partición octanol/agua, índices de retención, dulzor relativo.

Cuando se trabaja con respuestas que tienen una amplia escala, es necesario realizar una transformación (normalmente logarítmica), de tal forma que los valores numéricos de la actividad sean lo más cercanos posibles (al menos de 3 o 4 unidades logarítmicas). Otro aspecto importante en la preparación de los datos es la representación de las estructuras químicas de tal forma que sea posible el cálculo de los descriptores moleculares (variables independientes), los cuales serán relacionados con la actividad/propiedad. De esta manera, se puede construir la matriz de datos QSAR con la cual se trabajará en las etapas posteriores.

### **1.5.2. Procesamiento de los datos**

Antes de iniciar el desarrollo de un modelo QSAR, es importante el pretratamiento de los datos. Esto incluye:

- Eliminación de moléculas duplicadas. Por ejemplo, moléculas que tienen la misma estructura (por ejemplo, la misma notación de cadena SMILES) cuando se trabaja con modelos independientes de la conformación.
- Eliminación de descriptores con valores faltantes.
- Eliminación de descriptores constantes o casi constantes; así como los que se encuentran correlacionados más arriba de un cierto umbral.

Para el desarrollo de un modelo predictivo QSAR es indispensable la división de la matriz de datos QSAR en dos conjuntos o grupos: grupo de entrenamiento o calibración (training set) y grupo de predicción (test

set). Este tipo de división se puede realizar de forma casual (usando números aleatorios) o usando diversos métodos quimiométricos (análisis de conglomerados, Kennard–Stone, etc.). El grupo de calibración se usa para el desarrollo del modelo, para lo cual se usan técnicas de regresión o clasificación acopladas con métodos de selección de variables.

### **1.5.3. Validación y predicción de los datos**

El modelo desarrollado sobre el grupo de calibración es validado usando métodos de validación interna o cruzada (cross-validation) y una validación externa mediante los valores de respuesta obtenidos con el grupo de predicción. También se tiene que definir el dominio de aplicabilidad (AD), y en algunos casos se pueden utilizar otros criterios disponibles para la validación de modelos QSAR.

### **1.5.4. Interpretación de los datos**

Una vez que el modelo QSAR ha sido validado, la interpretación de la información codificada en los descriptores moleculares se usa para explorar el mecanismo de acción del conjunto de compuestos químicos considerados. Asimismo, es posible el uso del modelo para el diseño de nuevas sustancias químicas con una actividad/propiedad deseada.

## **1.6 Enfoques específicos QSAR**

Actualmente existen varios enfoques QSAR reportados en la literatura con una amplia gama de nombres, lo cual dificulta establecer un sistema de clasificación bien definido. Un intento por clasificar los enfoques QSAR podría hacerse considerando (Todeschini *et al.*, 2009):

1. El objetivo que debe ser alcanzado: diseño de fármacos, análisis de similitud molecular y cribado de alto rendimiento.
2. El tipo de propiedad molecular que se modela: análisis ADME, QSAR ambiental, LSERs y QSAR binario.
3. El tipo de descriptores moleculares que componen el modelo: 2D-QSAR, 3D-QSAR y 4D-QSAR.
4. El método matemático o algoritmo computacional usado para estimar los parámetros del modelo: Métodos de contribución de grupos, análisis estructural y técnicas QSAR basadas en grillas.

A continuación algunos de los enfoques QSAR se describirán con mayor detalle.

### **1.6.1. Enfoque de Hansch**

En términos generales se puede pensar que la actividad biológica de un grupo de compuestos químicos congénicos pueden ser descritos por el siguiente modelo matemático (Hansch *et al.*, 1962):

$$\log \frac{1}{C_{50}} = a\pi + b\varepsilon + cS + d \quad (1.2)$$

donde  $C_{50}$  es la actividad toxicológica de interés, correlacionada a un término hidrofóbico ( $\pi$ ), un término electrónico ( $\varepsilon$ ), un término estérico ( $S$ ) y un término que depende del tipo de actividad que se modela ( $d$ ).

El término electrónico era originalmente la constante de sustituyente de Hammett y el término estérico la constante de sustituyente de Taft. Estas dos contribuciones juntas fueron las que sentaron las bases para el desarrollo QSAR por parte de Hansch y Fujita.

### 1.6.2. Enfoque de Free–Wilson

El enfoque de Free–Wilson (Free & Wilson, 1964) se basa en el supuesto de que para un grupo de compuestos congénicos, la actividad/propiedad puede ser modelada utilizando en modo aditivo la contribución de los sustituyentes. Es decir, se considera que los efectos de tales sustituyentes presentes en la molécula sean independientes entre sí y que puedan ser simplemente sumados. Una vez identificado un esqueleto común para los compuestos congénicos de la base de datos, se realiza una regresión en busca de un modelo:

$$\bar{y}_i = b_0 + \sum_{s=1}^S \sum_{k=1}^{Ns} b_{ks} I_{i,ks} \quad (1.3)$$

donde  $S$  es el número de sitios de sustitución y  $Ns$  es el número de sustituyentes diferentes posibles. Los descriptores del  $i$ -ésimo compuesto se indican con la variable  $I_{i,ks}$ , siendo  $I_{i,ks} = 1$  si el sustituyente se encuentra presente en el  $s$ -ésimo sitio e  $I_{i,ks} = 0$  en caso contrario. Los coeficientes  $b_{ks}$  indican la importancia del  $k$ -ésimo sustituyente en cada  $s$ -ésimo sitio para la respuesta en estudio  $\bar{y}_i$ , donde  $b_0$  es la ordenada al origen del modelo, que corresponde a la respuesta media calculada de los compuestos considerados para construir el modelo.

### 1.6.3. Métodos de Contribución de Grupos

Los métodos de contribución de grupos (GCM) (Reinhard & Drefahl, 1999; Duchowicz *et al.*, 2006) se basan en la búsqueda de una relación cuantitativa entre la propiedad estructural y una respuesta fisicoquímica o biológica:

$$P = f(G_1, G_2, \dots, G_m; n_1, n_2, \dots, n_m) \quad (1.4)$$

donde  $P$  es la propiedad de interés, la cual es función de la contribución de  $m$  grupos funcionales  $G_i$  y de su frecuencia  $n_j$ . La contribución de los grupos funcionales se expresa mediante cantidades numéricas asociadas a los elementos estructurales de la molécula en forma de átomos individuales: pares de átomos, fragmentos moleculares y grupos funcionales. El modo en el cual cada grupo estructural se representa, depende del esquema elegido. La contribución de los grupos se estima utilizando métodos de regresión multivariable a partir de un grupo de moléculas cuyas propiedades son conocidas. El modelo lineal se puede escribir como:

$$y_i = k_0 + \sum_{j=1}^m G_j I_{ij} \quad (1.5)$$

donde  $k_0$  es la constante del modelo,  $j$  recorre los  $m$  grupos presentes (definidos en base al esquema GCM elegido),  $G_j$  es la contribución del  $j$ -ésimo grupo e  $I_{ij}$  es una variable que considera la presencia o ausencia del  $j$ -ésimo grupo en la  $i$ -ésima molécula.

#### 1.6.4. Cribado de alto rendimiento

El cribado de alto rendimiento o high throughput screening (HTS) (Pereira & Williams, 2007; Young, 2009) es un enfoque ampliamente usado en la industria farmacéutica para el descubrimiento de fármacos mediante el uso de la robótica. Los experimentos del HTS se realizan en placas de microvaloración que contienen cientos o miles de platos con un formato definido, donde se colocan las soluciones de los compuestos a ser evaluados. La información obtenida de los experimentos es recopilada mediante sensores sensibles para formar grandes bases de datos, las cuales serán posteriormente controladas y analizadas. El proceso de automatización permite el ensayo de la actividad/propiedad de miles de compuestos tipo fármaco en un tiempo reducido, pudiéndose identificar rápidamente los compuestos activos de interés.

#### 1.6.5. Barrido virtual

El barrido virtual (VS) (Vogt & Bajorath, 2010; Talevi *et al.*, 2011; Taft & da Silva, 2014) es un grupo de técnicas computacionales heterogéneas que permiten explorar las representaciones digitales de compuestos químicos registrados en bases de datos (públicas o privadas) con la finalidad de identificar potenciales candidatos con ciertas restricciones estructurales o determinados requisitos, los cuales

están de alguna manera vinculados a las propiedades/actividades deseadas. Es una técnica teórica debido a que los compuestos químicos a ser analizados han sido ya sintetizados y registrados en las bases de datos. Esto permite que el barrido virtual sea eficiente con respecto a otras técnicas que requieren mayor tiempo y dinero para diseñar un compuesto (por ejemplo el HTS). Asimismo, se la puede usar con la finalidad de reposicionar fármacos ya existentes, evitando de esta manera todo el estudio que conlleva su aprobación por parte de los organismos reguladores (Talevi & Bruno-Blanch, 2016).

## **1.7 Principios del modelado QSAR**

Para garantizar que las predicciones de los modelos QSAR sean confiables, se deben cumplir algunas condiciones. Las autoridades reguladoras necesitan asegurarse que un modelo QSAR haya sido estrictamente validado previo a su aplicación para la evaluación reglamentaria de compuestos químicos. Para que un modelo QSAR pueda ser aceptado para usos regulatorios debe:

1. Demostrar la validez del modelo.
2. La molécula a ser predicha (grupo de predicción) debe caer dentro del dominio de aplicabilidad del modelo.
3. La confiabilidad en el enfoque del modelado tiene que ser bien documentada de tal forma que se proporcione transparencia en el algoritmo subyacente.

Para abordar el procedimiento de validación de los modelos QSAR, existen principios referenciados y adoptados por la Organización para la Cooperación Económica y el Desarrollo (OECD, 2007). Así, cualquier modelo QSAR debe cumplir con 5 principios básicos internacionalmente reconocidos y que destacan varios aspectos relevantes para su aceptación regulatoria.

### **1.7.1. Actividad/propiedad definida**

Debido a que una determinada actividad/propiedad puede ser medida siguiendo protocolos experimentales diferentes y bajo diferentes condiciones experimentales, el objetivo de este primer principio es el de garantizar claridad en la definición de la actividad/propiedad que será modelada y predicha por el modelo.

### **1.7.2. Algoritmo inequívoco**

Debido a que varios enfoques de modelización han sido propuestos a lo largo del tiempo, este objetivo busca asegurar transparencia en el algoritmo utilizado para el desarrollo del modelo QSAR. En el caso de modelos desarrollados con fines comerciales, esta información no

siempre está disponible de forma pública, lo que genera un inconveniente para su aceptación regulatoria; es decir, la calidad de un modelo no puede ser establecida independientemente de la información del algoritmo utilizado para su desarrollo.

### **1.7.3. Dominio de aplicabilidad definido**

Este principio se refiere al hecho de que los modelos QSAR son modelos reduccionistas, los cuales están inevitablemente asociados con las limitaciones en términos de los tipos de estructuras químicas consideradas, de las propiedades fisicoquímicas y los mecanismos de acción para los cuales los modelos pueden generar predicciones confiables. En teoría, la aplicabilidad de un modelo QSAR está limitado a los compuestos químicos que son estructuralmente similares a los que fueron usados para calibrar el modelo.

### **1.7.4. Medida apropiada de la bondad de ajuste, robustez y predictividad**

Para una mejor evaluación de la calidad de un modelo, es esencial saber si el modelo es robusto, si no está sobreajustado y si es capaz de predecir confiablemente la actividad/propiedad para moléculas externas. Este principio brinda toda la información necesaria para un modelo QSAR obtenido mediante validación interna (bondad de ajuste y robustez) y validación externa (capacidad predictiva) mediante el uso de un conjunto de calibración y de predicción, respectivamente.

### **1.7.5. Interpretación del mecanismo de acción de los descriptores (de ser posible)**

Este quinto principio sugiere proporcionar interpretación de los descriptores moleculares del modelo y su importancia para la actividad/propiedad que se modela (siempre que sea posible). Esta interpretación mecanicista entre los descriptores y la actividad/propiedad puede añadir mayor confianza en el modelo desarrollado; sin embargo, no siempre es posible obtener tal interpretación desde un punto de vista científico. El hecho que no se pueda dar una interpretación mecanicista al modelo, no significa que tal modelo no sea potencialmente útil dentro de un contexto regulatorio y que tenga que ser rechazado.

## 1.8 Referencias

- Cronin, M. T. (2010). Quantitative Structure-Activity Relationships (QSARs)-Applications and Methodology. In T. Puzyn, J. Leszczynski & M. T. Cronin (Eds.), *Recent Advances in QSAR Studies*, (pp. 3-11): Springer.
- Chaudhry, Q., Chrétien, J., Craciun, M., Guo, G., Lemke, F., Müller, J.-A., Neagu, D., Piclin, N., Pintore, M., & Trundle, P. (2007). Algorithms for (Q) SAR Model Building. In E. Benfenati (Ed.), *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*, (pp. 111-147): Elsevier.
- Dearden, J. C. (2016). The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relationships*, 1(1), 1-44.
- Duchowicz, P. R., Castro, E. A., Toropov, A., & Benfenati, E. (2006). Applications of Flexible Molecular Descriptors in the QSPR-QSAR Study of Heterocyclic Drugs. In S. P. Gupta (Ed.), *QSAR and Molecular Modeling Studies in Heterocyclic Drugs I*, (pp. 1-38).
- Franke, R., & Gruska, A. (2003). General Introduction to QSAR. In R. Benigni (Ed.), *Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens*, (pp. 1-40): CRC press.
- Free, S. M., & Wilson, J. W. (1964). A Mathematical Contribution to Structure-Activity Studies. *Journal of Medicinal Chemistry*, 7(4), 395-399.
- Galvez, J., & Garcia-Domenech, R. (2010). Molecular Topology in QSAR and Drug Design Studies. In E. Castro (Ed.), *QSPR-QSAR Studies on Desired Properties for Drug Design*, (pp. 63-94): Research Signpost.
- Gangwal, R. P., Damre, M. V., & Sangamwar, A. T. (2016). Overview and Recent Advances in QSAR Studies. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research: QSAR in Medicinal Chemistry*, (pp. 1-32): CRC Press.
- Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment. In J. Leszczynski, A. Kaczmarek-Kędziera, T. Puzyn, M. G. Papadopoulos, H. Reis & M. K. Shukla (Eds.), *Handbook of Computational Chemistry*, (pp. 2303-2340): Springer.
- Hamzeh-Mivehroud, M., Sokouti, B., & Dastmalchi, S. (2015). An Introduction to the Basic Concepts in QSAR-Aided Drug Design. In K. Roy (Ed.), *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*: IGI Global.
- Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of Biological Activity of Phenoxyacetic Acids with Hammett

- Substituent Constants and Partition Coefficients. *Nature*, 194, 178–180.
- Hansch, C., Leo, A., & Hoekman, D. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*: American Chemical Society.
- Hongmao, S. (2015). *A Practical Guide to Rational Drug Design*: Elsevier/Woodhead Publishing.
- Kaliszan, R. (2007). QSRR: Quantitative Structure-(Chromatographic) Retention Relationships. *Chemical reviews*, 107, 3212-3246.
- Kubinyi, H. (1988). Free Wilson Analysis. Theory, Applications and Its Relationship to Hansch Analysis. *Molecular Informatics*, 7(3), 121-133.
- Kubinyi, H. (2008). *QSAR: Hansch Analysis and Related Approaches* (Vol. 1): John Wiley & Sons.
- Liaw, A., & Svetnik, V. (2015). QSAR Modeling: Prediction of Biological Activity from Chemical Structure. In A. L. Gould (Ed.), *Statistical Methods for Evaluating Safety in Medical Product Development*, (pp. 66-83): Wiley.
- OECD. (2007). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. In: OECD Publishing.
- Pereira, D., & Williams, J. (2007). Origin and Evolution of High Throughput Screening. *British journal of pharmacology*, 152(1), 53-61.
- Reinhard, M., & Drefahl, A. (1999). *Handbook for Estimating Physicochemical Properties of Organic Compounds*: Wiley.
- Rojas, C., Duchowicz, P. R., Pis Diez, R., & Tripaldi, P. (2016). Applications of Quantitative Structure-Relative Sweetness Relationships in Food Chemistry. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research: QSAR in Medicinal Chemistry*, (pp. 317-339): Apple Academic Press.
- Roy, K., Kar, S., & Das, R. N. (2015a). *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*: Springer.
- Roy, K., Kar, S., & Das, R. N. (2015b). *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*: Academic Press.
- Taft, C. A., & da Silva, C. H. T. d. P. (2014). Current State-of-the-art for Virtual Screening and Docking Methods. In C. A. Taft & C. H. T. d. P. da Silva (Eds.), *New Developments in Medicinal Chemistry*, vol. 2 (pp. 3-169).
- Talevi, A., Castro, E. A., & Bruno-Blanch, L. E. (2011). Virtual Screening: An Emergent, Key Methodology for Drug Development in an Emergent Continent-A Bridge Towards Patentability. In E. A. Castro & A. K. Haghi (Eds.), *Advances Methods and Applications*

- in Chemoinformatics: Research Progress and New Applications*, (pp. 229-245).
- Talevi, A., & Bruno-Blanch, L. E. (2016). Virtual Screening Applications in the Search of Novel Antiepileptic Drug Candidates. In A. Talevi & L. Rocha (Eds.), *Antiepileptic Drug Discovery: Novel Approaches*, (pp. 237-258).
- Todeschini, R., Consonni, V., & Gramatica, P. (2009). Chemometrics in QSAR. In R. Tauler, B. Walczak & S. D. Brown (Eds.), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, (pp. 129-170): Elsevier.
- Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*: Wiley-VCH.
- Vogt, M., & Bajorath, J. (2010). Virtual Screening Methods Based on Bayesian Statistics. In H. Lodhi & Y. Yamanishi (Eds.), *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, (pp. 190-211): Medical Information Science Reference.
- Young, D. C. (2009). *Computational Drug Design: A Guide for Computational and Medicinal Chemists*: John Wiley & Sons.

# Capítulo 2

---

## DESCRIPTORES MOLECULARES

---

### 2.1 Introducción

Los descriptores moleculares son de gran importancia en el campo de las relaciones cuantitativas estructura–actividad/propiedad, debido a que constituyen las variables independientes con las cuales se predicen las actividades/propiedades de interés. De esta forma, es imprescindible poder representar de forma apropiada la estructura molecular, para con el uso de herramientas quimiométricas adecuadas desarrollar los modelos QSAR/QSPR. Así, los científicos se han enfocado en cómo capturar y convertir (de forma teórica) la información codificada en una estructura química en números que se puedan usar para relacionarlos cuantitativamente con actividades/propiedades biológicas u otras propiedades experimentales de interés (Gasteiger & Engel, 2006).

Los descriptores moleculares son un conjunto de parámetros capaces de describir una estructura molecular de forma cuantitativa. Sin embargo, este conjunto de parámetros no es único, debido a que se los puede extraer a partir de diferentes formas de representación de la molécula. En consecuencia, se pueden distinguir dos grandes clases de descriptores moleculares: los experimentales y los teóricos. Los descriptores experimentales son aquellos que se obtienen mediante experimentos estandarizados, por ejemplo, coeficiente de partición octanol/agua, refractividad molar, polarizabilidad, etc.; mientras que los descriptores teóricos son aquellos que se obtienen aplicando algoritmos matemáticos bien definidos a una representación inequívoca de la estructura molecular (Todeschini & Consonni, 2009).

El creciente interés de los científicos en el desarrollo de descriptores moleculares, junto con los avances en la quimiometría y quimioinformática, han permitido relacionar de diversas formas la estructura molecular con las actividades/propiedades de los compuestos químicos. Por lo tanto, las áreas de aplicación de los descriptores es variada: fisicoquímica, química medicinal, farmacología, protección ambiental, toxicología y ecotoxicología, ciencia de los alimentos, barrido virtual, cribado de alto rendimiento, entre otras. Asimismo, existen aplicaciones más recientes de los descriptores en sistemas más complicados; tales como sales (estructuras desconectadas), nanoestructuras, péptidos, polímeros y líquidos iónicos (Mauri *et al.*, 2016).

## 2.2 Historia

El desarrollo histórico de los descriptores moleculares se encuentra estrechamente relacionado con el concepto de estructura molecular y se considera como uno de los conceptos más importantes del siglo XIX. De esta forma, el desarrollo histórico de los descriptores refleja la creatividad de los científicos basado en sus capacidades para imaginar cómo representar una estructura química y al mismo tiempo obtener información a partir de dicha representación. Los descriptores moleculares pueden considerarse el desarrollo más importante de la idea de Alexander Crum Brown plasmado en su tesis titulada “*On the Theory of Chemical Combination*”, presentada en el año 1861 en la Universidad de Edimburgo. La tesis de Crum Brown indica que fue un investigador pionero en el campo de la química matemática, debido a que desarrolló un sistema de representación gráfica de los compuestos químicos el cual es muy similar al que se usa en la actualidad. Su representación gráfica fue la primera que permitió visualizar claramente la valencia y los enlaces de los átomos en compuestos orgánicos (Consonni & Todeschini, 2009).

Más adelante, Crum Brown publicó el trabajo titulado “*Theory of Isomeric Compounds*”, en el cual relaciona la matemática y la química mediante el análisis de varios tipos de isómeros usando una representación gráfica. En 1868 y en colaboración con T. R. Fraser propusieron que la acción de un compuesto en un determinado sistema biológico está definida como una función de su constitución química (Franke & Gruska, 2003; Kubinyi, 2008):

$$\Phi = f(C) \quad (2.1)$$

En esta relación, ellos indican que una modificación de la constitución química ( $\Delta C$ ) se reflejaría en un efecto sobre la actividad biológica ( $\Delta \Phi$ ), y se puede considerar la primera proposición general de una relación cuantitativa estructura–actividad.

Por otro lado, Wilhelm Körner en 1874 publicó los resultados de su trabajo relacionados a la síntesis de compuestos bencénicos di-sustituidos, donde estableció el efecto de las posiciones *orto*, *meta* y *para* de los sustituyentes (descriptores). De esta manera, Körner pensó que los diferentes colores que mostraban tales compuestos estarían relacionados a diferencias en la estructura molecular (Consonni & Todeschini, 2009). Alrededor de 1935, Hammett también estableció el efecto de los sustituyentes en reacciones orgánicas (particularmente en la ionización de ácidos benzóicos), donde buscó relaciones lineales para la energía libre mediante el uso de constantes estéricas, electrónicas e hidrofóbicas para diversos sustituyentes y el uso de un modelo aditivo para determinar la actividad congénica de los compuestos (Hansch *et al.*, 1995).

En 1947 se introducen los primeros descriptores basados en la teoría de grafos moleculares, el índice de Wiener y el número de Platt, con la finalidad de modelar los puntos de ebullición de hidrocarburos. Más adelante, en la década de 1950, Taft estableció relaciones lineales para la energía de solvatación (LSER) basadas en parámetros estéricos, polares y de resonancia para derivados congenéricos. A inicios de la década de 1960 se propusieron otros descriptores moleculares basados principalmente en la teoría de grafos, y casi una década después se comenzaron a utilizar los descriptores cuánticos en los estudios QSAR. Asimismo, la representación topológica de la estructura molecular llevada a cabo en los trabajos de Alexandru Balaban, Milan Randić, Lemont Kier y Lowell Hall (Roy & Das, 2012), permitió realizar avances significativos en el desarrollo de nuevos descriptores; así como extender estos conceptos para buscar una representación geométrica de la molécula para el desarrollo de descriptores 3D y huellas digitales moleculares (Doucet & Panaye, 2010).

Por otra parte, a finales de la década de 1980 se propusieron estudios QSAR basados en el análisis comparativo de campos moleculares (CoMFA) y en el análisis comparativo del índice de similitud molecular (CoMSIA) (Kubinyi, 1998; da Silva *et al.*, 2010; Doucet & Panaye, 2010; Martin, 2010). El primer tipo de análisis mide las energías de interacción entre una molécula y grupos de interacción (“*probes*”) en diversos puntos del espacio 3D. Los grupos de interacción pueden ser la molécula de agua, el grupo metilo o el átomo de hidrógeno; los cuales se usan para medir las energías de interacción en miles de puntos de una malla cuadrículada en la que se introduce la molécula. Así, se obtiene un campo escalar de valores de energía de interacción que caracterizan la molécula analizada. De forma similar, CoMSIA mide la similitud de las moléculas en función de sus propiedades fisicoquímicas; es decir, los resultados muestran regiones de la molécula con afinidad o falta de afinidad a la presencia de un grupo de interacción con propiedades fisicoquímicas específicas.

### 2.3 Definición y requisitos

Los descriptores moleculares son formalmente representaciones matemáticas de una molécula, los cuales se obtienen a través de la aplicación de un algoritmo específico a una representación molecular definida, o a un procedimiento experimental específico (Consonni & Todeschini, 2010; Mauri *et al.*, 2016).

La definición formal de descriptor molecular fue propuesta por Todeschini y Consonni en el año 2000: “*el descriptor molecular es el resultado final de un procedimiento lógico y matemático que transforma la información química codificada en una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado*” (Todeschini & Consonni, 2009). Debido a esta definición

amplia, el campo de los descriptores moleculares es multidisciplinario y engloba diferentes teorías en diferentes niveles. Así, para el desarrollo de nuevos descriptores es necesario poseer conocimientos de álgebra, teoría de grafos, teoría de la información, química computacional, teorías de reactividad química y fisicoquímica. Por otro lado, para el análisis e interpretación de los mismos son necesarios conocimientos de estadística, quimiometría, principios básicos de modelado QSAR/QSPR, conocimiento específico del problema, bases de programación y entrenamiento en el uso de computadores y programas específicos (Gasteiger & Engel, 2006).

En la actualidad existen miles de descriptores que han sido propuestos en la literatura; por tal motivo se han definido reglas básicas que deben cumplir para ser aceptados como tal (Randić, 1996; Guha & Willighagen, 2012):

1. Ser invariante al etiquetado y enumerado de los átomos.
2. Ser invariante a la roto-traslación de la molécula.
3. Ser definido por un algoritmo inequívoco.
4. Poseer una aplicabilidad bien definida en las estructuras moleculares.

Además, para ser potencialmente útil, un descriptor molecular debería cumplir los siguientes requisitos:

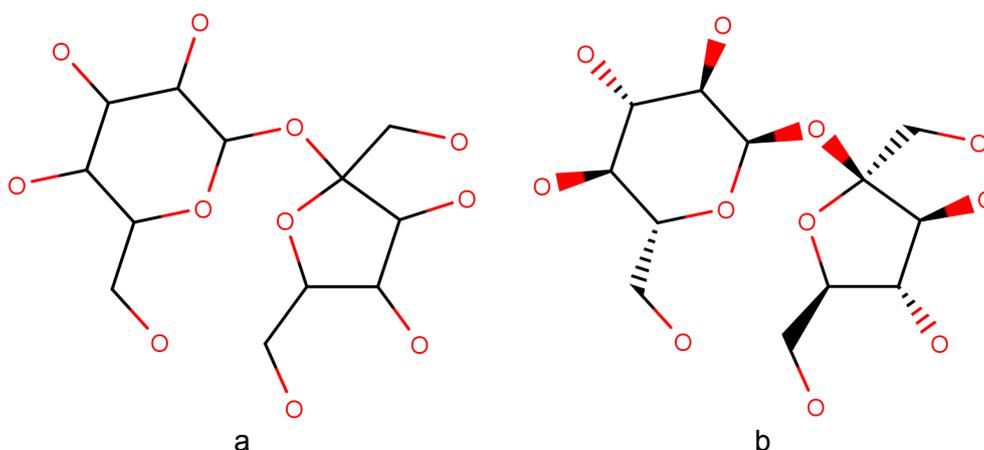
5. Tener una interpretación estructural.
6. Tener buena correlación al menos con una propiedad experimental.
7. No tener relación trivial con otros descriptores.
8. No estar basado en propiedades experimentales.
9. Ser preferencialmente continuo.
10. Preferencialmente poseer mínima degeneración.
11. Ser preferencialmente simple.
12. Preferencialmente ser aplicable a una amplia clase de moléculas.
13. Preferencialmente ser capaz de discriminar isómeros.
14. Preferencialmente poseer valores calculados en un rango numérico adecuado para el conjunto de moléculas sobre las cuales se utilizará.

Las primeras cuatro reglas se usan para saber si un descriptor molecular está bien definido, pero no permiten identificar si un descriptor molecular será o no será apropiado para predecir una propiedad/actividad específica. Las siguientes diez reglas se relacionan con el uso de un descriptor, el cual debe ser interpretable, relacionado al menos con una propiedad experimental pero no estrechamente

relacionado con los demás descriptores. La continuidad y baja degeneración indica que un descriptor debe ser capaz de considerar las variaciones (incluso mínimas) en la estructura molecular.

## 2.4 Representación de la estructura molecular

Las sustancias químicas se pueden representar siguiendo diferentes reglas y criterios. Dependiendo de la representación molecular que se utilice, se encontrará diferente tipo de información reflejada en los descriptores moleculares calculados (Testa & Kier, 1991). En términos generales, las estructuras moleculares se las puede representar por esquemas gráficos que indican el tipo de complejidad estructural (Figura 2.1) o por medio de notaciones lineales de cadena (Tabla 2.1).



**Figura 2.1.** Representación de la estructura química de la sacarosa: (a) representación 2D y (b) representación conformación molecular

La fórmula química es la forma más simple de representar una molécula, en la cual se encuentra codificada la frecuencia de ocurrencia de los tipos de átomos. Evidentemente, esta representación no indica ningún tipo de información relacionada con la conectividad atómica ni conformación molecular. Por consiguiente, los descriptores moleculares calculados bajo este formato se los denomina 0D o constitucionales. Por ejemplo: número de átomos de carbono o de nitrógeno en una molécula.

**Tabla 2.1.** Diversos tipos de notación lineal para la estructura química de la sacarosa

| Tipo              | Notación   |
|-------------------|--|
| Nombre comercial  | Sacarosa   |
| Nombre IUPAC      | (2R,3R,4S,5S,6R)-2-[(2S,3S,4S,5R)-3,4-dihydroxy-2,5-bis(hydroxymethyl)oxolan-2-yl]oxy-6-(hydroxymethyl)oxane-3,4,5-triol |
| Fórmula molecular | $C_{12}H_{22}O_{11}$   |
| Número CAS        | 57-50-1  |
| SMILES canónico   | <chem>C(C1C(C(C(O1)OC2(C(C(C(O2)CO)O)O)CO)O)O)O</chem>   |
| SMILES isomérico  | <chem>C([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O)[C@]2([</chem>  |

La lista de subestructura consiste en una lista de fragmentos estructurales que se encuentran dentro de la molécula. Estos fragmentos no representan la topología completa de la estructura química, por lo que se los puede calcular e interpretar con facilidad. Los descriptores moleculares obtenidos de esta representación se denominan 1D y son comúnmente usados en los análisis de similitud/diversidad o para el barrido virtual de grandes bases de datos. Como ejemplos de estos descriptores se pueden citar los descriptores de *conteo de grupos funcionales* o *fragmentos centrados en el átomo*.

La representación en dos dimensiones de las moléculas es la más común y se la denomina representación topológica. Esta representación incluye la información de la conectividad entre los átomos que constituyen la molécula, la cual se realiza en términos de los denominados grafos moleculares. Los descriptores moleculares calculados en base a la representación topológica se denominan descriptores 2D. Aquí se encuentran índices topológicos, índices de conectividad, índices de información, autovalores de Burden, entre otros.

La representación geométrica de la disposición los átomos de una molécula en el espacio de tres dimensiones adiciona información a las conectividades atómicas. Los descriptores moleculares obtenidos a partir de esta representación se denominan 3D. Aquí se encuentran los descriptores 3D–MoRSE, WHIM, GETAWAY, CATS 3D, RDF, etc. Debido a que los compuestos químicos existen en diversas conformaciones de equilibrio, las cuales minimizan su energía, los valores de los descriptores 3D cambian en función de la conformación usada para su cálculo (García *et al.*, 2016).

### 2.4.1. Grafos moleculares

Los grafos moleculares (Polansky, 1991; Janežič *et al.*, 2015) son el tipo de representación molecular 2D comúnmente usado. Un grafo  $G = (V, E)$ , es una representación matemática de un grupo de vértices ( $V$ ) y un grupo de aristas entre los vértices ( $E$ ). En un grafo molecular los átomos corresponden a los vértices; mientras que las aristas corresponden a los enlaces químicos entre pares de átomos. Un grafo molecular es no dirigido (cuando los vértices no tienen orientación), ponderado y disperso. En la ponderación se asigna un número a cada vértice, que normalmente corresponden a los órdenes de los enlaces. Finalmente, el grafo es disperso cuando el número de aristas es mucho menor que el cuadrado del número de vértices  $|E| \ll |V|^2$ .

La aplicación de la teoría de grafos a la estructura molecular permite manejar los grafos químicos mediante algoritmos útiles y bien definidos para explorar las propiedades estructurales de las moléculas (Balaban,

1985). Particularmente, la teoría de grafos puede usarse para calcular matrices que dependen de las conexiones en el grafo y la definición de los invariantes del grafo. Los invariantes son representaciones numéricas de un grafo de aquellas propiedades que dependen únicamente de la estructura abstracta de los grafos y no de la representación de los mismos. Los invariantes de grafo pueden ser valores simples, una secuencia de números o una característica polinomial derivada del grafo. Los invariantes se pueden usar para reconocer rápidamente isomorfismo del grafo, debido a que, para cualquier invariante, dos grafos con diferentes valores no pueden ser isomorfos. Sin embargo, dos grafos con el mismo valor de invariante pueden o no ser isomorfos. La teoría de invariantes de grafos es la base fundamental para la definición de los índices topológicos (Todeschini & Consonni, 2009; Bonchev, 2015).

Generalmente un grafo se representa de dos maneras diferentes: como una colección de listas de adyacencia o como una matriz de adyacencia. Las listas de adyacencia son una colección de  $|V|$  listas, una para cada  $i$ -ésimo átomo, donde cada lista  $Adj[i]$  incluye los átomos conectados al  $i$ -ésimo átomo junto con el correspondiente orden de enlace.

### 2.4.2. Representación matricial

Varias matrices se pueden calcular a partir de un grafo molecular, las cuales se conocen como matrices grafo-teóricas y permiten el cálculo de muchos descriptores moleculares. Estos descriptores moleculares dependen del tipo de información que se incluye en dichas matrices. La mayoría de las matrices grafo-teóricas se derivan a partir de grafos que no incluyen los átomos de hidrógenos (libre de hidrógenos); sin embargo, también pueden ser calculadas considerando el grafo completo (con átomos de hidrógeno), cuando dicho átomo es útil para obtener una mejor representación de la estructura molecular (puentes de hidrógeno, donante/aceptor de enlaces de hidrógeno). La ventaja de representar una estructura molecular en forma de matriz es que se pueden aplicar varios operadores a las diferentes matrices grafo-teóricas, lo que permite calcular diversos grupos de descriptores moleculares. Así, un gran número de matrices grafo-teóricas ha sido propuesto en la literatura; entre las cuales, las más importantes son (Todeschini & Consonni, 2009; Janežič *et al.*, 2015):

1. **Matrices de vértices:** son matrices cuadradas de dimensión  $|V| \times |V|$ , cuyas filas y columnas representan a los átomos (vértices del grafo), y cada elemento de la matriz codifica una propiedad asociada a un par de átomos.

2. **Matrices de aristas:** son matrices cuadradas de dimensión  $|E| \times |E|$ , cuyos elementos codifican la información de pares de enlaces.
3. **Matrices de incidencia:** contienen información concerniente a las relaciones entre dos diferentes conjuntos de objetos; tales como átomos, enlaces, ciclos, subestructuras o trayectos moleculares; por lo que no son matrices cuadradas y cuya dimensión depende de los objetos considerados.

Las matrices de vértices son las más usadas para el cálculo de descriptores moleculares, y entre ellas las más representativas son las matrices de adyacencia, de distancias topológicas, Laplaciana y detour.

### **Matriz de adyacencia**

La matriz de adyacencia **A** es una matriz cuadrada simétrica de dimensión  $|V| \times |V|$  donde cada elemento  $a_{ij}$  es definido de la siguiente manera:

$$[\mathbf{A}]_{ij} \equiv a_{ij} = \begin{cases} 1 & \text{si } (i,j) \in E \\ 0 & \text{caso contrario} \end{cases} \quad (2.2)$$

es decir, los átomos son adyacentes si existe un enlace químico entre ellos.

### **Matriz de distancias topológicas**

La matriz de distancias topológicas **D** es una matriz cuadrada simétrica de dimensión  $|V| \times |V|$ , en la que cada elemento  $d_{ij}$  es igual a la distancia topológica entre los vértices  $i$  y  $j$ :

$$[\mathbf{D}]_{ij} \equiv d_{ij} = \begin{cases} |\min p_{ij}| & \text{si } i \diamond j \\ 0 & \text{si } i = j \end{cases} \quad (2.3)$$

donde  $|\min p_{ij}|$  es el número de aristas a lo largo de la ruta entre los átomos  $i$  y  $j$ .

### **Matriz de distancias geométricas**

La matriz de distancias geométricas **G** representa las distancias entre pares de átomos en el espacio 3D. Es una matriz simétrica de dimensión  $|V| \times |V|$ , cuyos elementos diagonales  $r_{ii} = 0$  y los elementos fuera de la

diagonal  $r_{ij}$  corresponden a las distancias Euclidianas entre los pares de átomos  $i$  y  $j$  presentes en la molécula.

### **Matriz Laplaciana**

La matriz Laplaciana  $\mathbf{L}$  o matriz de Kirchhoff es una matriz cuadrada simétrica de dimensión  $|V| \times |V|$  definida como una matriz aumentada; es decir, se la obtiene al combinar una matriz cuadrada y una matriz diagonal (cuyos elementos diagonales son las propiedades atómicas). De forma más específica, la matriz  $\mathbf{L}$  se obtiene como la diferencia entre la matriz de vértices y la matriz de adyacencia. La matriz de vértices es una matriz en la que los elementos diagonales corresponden a los grados de vértices  $\delta_i$ , que son el número de átomos (excluido el H) adyacentes al átomo  $i$ :

$$[\mathbf{L}]_{ij} = \begin{cases} \delta_i & \text{si } i = j \\ -1 & \text{si } (i, j) \in E \\ 0 & \text{si } (i, j) \notin E \end{cases} \quad (2.4)$$

donde los grados de vértices se calculados como la suma de las filas de la matriz  $\mathbf{A}$ :

$$\delta_i = \sum_{j=1}^{|V|} a_{ij} \quad (2.5)$$

### **Matriz detour**

La matriz detour  $\mathbf{\Delta}$  o matriz de trayectoria máxima, es una matriz de vértices simétrica definida como:

$$[\mathbf{\Delta}]_{ij} = \begin{cases} |\max p_{ij}| & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases} \quad (2.6)$$

donde  $|\max p_{ij}|$  es el número de bordes a través del camino más largo entre los vértices  $i$  y  $j$ ; es decir, la distancia detour.

## **2.5 Codificación de la estructura molecular**

La información de la estructura molecular se almacena en los denominados archivos de datos químicos. Sin embargo, a pesar de la gran cantidad de tipos de formatos existentes para dichos archivos químicos (O'Boyle *et al.*, 2011), todos ellos representan la estructura

molecular brindando información específica de tipo atómica o molecular, por lo que no pueden capturar todos los datos químicos posibles. Particularmente, se tienen dos tipos de archivos químicos, el primero relacionado a la representación topológica de las moléculas y el segundo a la representación geométrica de la estructura molecular.

Generalmente, el primer caso incluye todos los tipos de notación lineal, entre los cuales el más utilizado es la notación de cadena SMILES (simplified molecular-input line-entry system) (Weininger, 1988); mientras que para el segundo caso, los archivos que pueden almacenar la información de las coordenadas tridimensionales de los átomos son los formatos MDL y SYBYL. Sin embargo, se ha propuesto el formato CML (Chemical Markup language) para proporcionar un formato estándar para el almacenamiento de datos químicos basados en el dialecto XML (Murray-Rust & Rzepa, 2001).

## 2.6 Curado de las estructuras moleculares

Un aspecto importante previo al cálculo de los descriptores o huellas digitales moleculares es comprobar que las estructuras moleculares sean correctas. Desafortunadamente, las estructuras químicas disponibles en publicaciones científicas o en bases de datos (públicas o comerciales) no están exentas de errores.

En años recientes, el curado de las estructuras moleculares ha sido fuertemente enfatizado en varias publicaciones, donde se han expuesto los problemas relacionados al uso de datos incorrectos en estudios QSAR y en quimioinformática (Young *et al.*, 2008; Tropsha, 2010). El curado se aplica a la base de datos completa y se compone de 5 pasos principales (Fourches *et al.*, 2010):

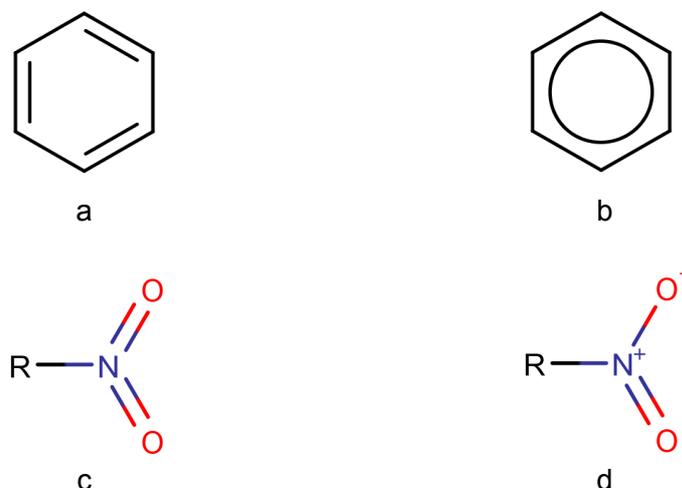
1. Eliminación de estructuras químicas que no pueden ser procesadas por algunos programas de cálculo de descriptores.
2. Conversión y limpieza de las estructuras químicas.
3. Estandarización y normalización de quimiotipos específicos.
4. Eliminación de compuestos repetidos.
5. Control manual final.

En el primer paso, los compuestos inorgánicos, organometálicos y las mezclas de sustancias químicas son los casos más comunes a ser excluidos durante el curado de las estructuras. Esta exclusión se realiza cuando el programa utilizado para el cálculo de descriptores moleculares no admite este tipo de compuestos. El segundo paso se relaciona propiamente a la preparación de la base de datos, donde las estructuras químicas deben ser codificadas en formatos computacionales específicos (SMILES, SYBYL o MDL). Esta conversión se puede realizar diseñando la estructura química en un editor molecular o directamente descargándolas a partir de una biblioteca de compuestos comerciales o de acceso público. Posteriormente, un control exhaustivo debe ser

realizado sobre las estructuras generadas con la finalidad de buscar estructuras incorrectas.

En el tercer paso se considera el hecho de que una molécula (incluso un fragmento o grupo funcional), puede ser representada de diferentes formas. Por ejemplo, los anillos aromáticos que se representan en el formato de Kekulé (Figura 2.2a) o en el formato aromático (Figura 2.2b); el grupo funcional nitro ( $R-NO_2$ ) que posee múltiples mesómeros y cuya estructura puede ser representada usando la notación neutra (Figura 2.2c) o cargada (Figura 2.2d). Debido a que las representaciones son diversas, en principio afectaría a los valores de los descriptores moleculares. Por lo tanto, se debe elegir la misma forma de representación y aún más importante es que ese tipo de normalización se aplique a todas las estructuras moleculares.

Finalmente, se debe realizar un control manual de todas las estructuras moleculares codificadas, a pesar de que constituya una tarea tediosa y demandante de tiempo (dependiendo del tamaño de la base de datos). Sin embargo, es un paso vital para identificar moléculas no diseñadas correctamente y para corregir cualquier inconveniente que haya aparecido en los pasos anteriores.



**Figura 2.2.** Representación molecular del anillo aromático en el formato de Kekulé (a) y en el formato de enlaces conjugados (b); y del grupo funcional nitro en la forma neutra (c) y cargada (d)

### 2.6.1. KNIME

Debido a que el curado de las estructuras moleculares para los estudios QSAR/QSPR requiere el control de todos los errores anteriormente mencionados, un curado manual resulta tedioso, especialmente cuando se manejan bases de datos grandes. Para tal

efecto se puede recurrir a programas quimioinformáticos, por ejemplo KNIME (Berthold *et al.*, 2008).

KNIME (Konstanz Information Miner) es un programa gratuito y de acceso abierto, que permite realizar minería de datos. Se inicia con el acceso a los datos para posteriormente realizar transformaciones, filtrados, etc., hasta llegar a realizar predicciones, visualizaciones y reportes de resultados. Para este efecto, se encuentran a disposición varios módulos (denominados nodos). El programa se encuentra organizado en diferentes secciones, de las cuales las más importantes son:

- Editor del diagrama de flujo.
- Repositorio de nodos.
- Descripción de nodos.

El repositorio de nodos contiene todos los nodos instalados y organizados en categorías y subcategorías. Para programar un diagrama de flujo, los nodos necesarios son buscados en el repositorio y arrastrados al editor del diagrama. Entonces, los nodos se conectan según el orden deseado mediante los puertos de entrada/salida y posteriormente se los configura para realizar las operaciones requeridas.

KNIME brinda acceso a los nodos instalados por defecto. Sin embargo, es posible instalar extensiones de KNIME que permiten el acceso a nodos adicionales específicos, los cuales han sido desarrollados por la comunidad de usuarios de KNIME. En el campo de la quimioinformática existen algunos nodos útiles que han sido incluidos en el repositorio, tales como ChemAxon tools, the Chemistry Development Kit CDK, DRAGON, CIR (Solucionador de identificación química), PaDEL (Pharmaceutical Data Exploratory Laboratory) y otros más, los que permiten la recopilación de datos, el curado, modelado y predicción de las actividades/propiedades de nuevos compuestos químicos.

## **2.7 Principales tipos de descriptores moleculares**

En esta sección se van a describir diversos tipos de descriptores moleculares, particularmente aquellos utilizados durante el desarrollo de la presente tesis.

### **2.7.1. Descriptores constitucionales**

Son los descriptores moleculares más simples que se pueden calcular a partir de una estructura molecular (Mauri *et al.*, 2016). En este grupo de descriptores se encuentran los que representan la estructura molecular sin tomar en cuenta ni la topología ni la geometría. Algunos ejemplos de descriptores constitucionales son:

- Peso molecular.

- Número de átomos: átomos terminales, heteroátomos, átomos ponderados.
- Número de enlaces múltiples: dobles, triples, aromáticos.
- Número de enlaces rotables.
- Frecuencia de ocurrencia absoluta o relativa de un átomo específico y tipos de enlace.

### 2.7.2. Descriptores de anillo

Los descriptores de anillo codifican la información relacionada a la presencia de anillos (estructuras cíclicas o anillos aromáticos) en una molécula. Algunos tipos de esta clase de descriptores son:

- Número de anillos.
- Número de circuitos o ciclos.
- Tamaño total del anillo.
- Perímetro del anillo.
- Conteo de puentes en el anillo.
- Grado de ciclizado molecular.
- Índice de complejidad del anillo.
- Relación aromática.

### 2.7.3. Índices topológicos

Los índices topológicos (TIs) (Todeschini & Consonni, 2009; Bonchev, 2015; Roy *et al.*, 2015a), también conocidos como descriptores moleculares topológicos, se derivan de una representación topológica de la estructura química (2D); por ejemplo, un grafo molecular. De esta manera, estos índices no consideran ninguna información acerca de la distribución espacial de los átomos. Existen diferentes categorías de TIs y serán tratadas a continuación.

#### ***Número de trayectos moleculares***

Son descriptores que cuentan el número de caminos, trayectos de ida, y trayectos de ida y vuelta en un grafo molecular libre de hidrógenos usando diferentes longitudes topológicas. Cuando se trabaja con moléculas grandes, se realiza una transformación logarítmica, de tal forma que permita obtener trayectos de tamaño apropiado para su uso (Kode srl., 2016).

### **Índices de conectividad molecular**

Los índices de conectividad molecular (MCIs) (Kier & Hall, 1986) se calculan a partir de un grafo molecular libre de hidrógenos, donde cada vértice se pondera por su grado; es decir, el número total de átomos conectados. Estos índices son utilizados muy asiduamente. Su fórmula general es:

$$X_k = \sum_{j=1}^K \left( \prod_{i=1}^n \delta_i \right)_j^{-1/2} \quad (2.7)$$

donde  $j$  recorre todos los trayectos de orden  $k$ , incluyendo los  $n$  vértices ( $n = k + 1$  para subgrafos acíclicos);  $K$  es el número total de los trayectos de orden  $k$  presentes en el grafo molecular. El producto se realiza sobre los grados de vértice  $\delta$  de todos los vértices involucrados en cada subgrafo. Entre los índices de conectividad más conocidos se tienen:

- Índice de conectividad de Randić.
- Índices de conectividad molecular de Kier–Hall.
- Índices de conectividad promedio.
- Índices de conectividad de solvatación.
- Índice modificado de Randić.
- Índice de conectividad de Kupchik.

El índice de conectividad de Randić (Randić, 1975), también conocido como índice de ramificación, fue el primer índice de conectividad propuesto para medir el grado de ramificación de hidrocarburos.

### **Índices de información**

Los índices de información (Todeschini & Consonni, 2009; Roy *et al.*, 2015b; Kode srl., 2016) son descriptores moleculares que indican el contenido de información de las moléculas; mediante el uso de diversos criterios para definir las clases de equivalencia (es decir, equivalencia de los átomos en una molécula), tal como la identidad química, formas de enlace en el espacio, topología molecular y simetría. La mayoría de estos índices se derivan de un grafo molecular y se basan en la partición de sus elementos o de los elementos de la matriz en clases de equivalencia siguiendo diversos criterios.

### **Autovalores de Burden**

Son descriptores moleculares derivados de un grafo molecular libre de hidrógenos y calculados a partir de la matriz de Burden **B** (Burden, 1989), la cual es una matriz ponderada que proporciona los denominados autovalores de Burden, y es definida de la siguiente forma:

$$[\mathbf{B}]_{ij} = \begin{cases} \pi_{ij}^* \times 10^{-1} & \text{si } (i,j) \in E \\ Z_i & \text{si } i = j \\ 0.001 & \text{si } (i,j) \notin E \end{cases} \quad (2.8)$$

Los elementos diagonales  $B_{ii}$  son los números atómicos  $Z_i$  de los átomos. Los elementos no diagonales  $B_{ij}$  representan dos átomos enlazados que son iguales a  $\pi^* \times 10^{-1}$ , donde  $\pi^*$  es el orden de enlace convencional, igual a 1, 2, 3, y 1.5 para enlaces simples, dobles, triples y aromáticos, respectivamente. Los  $B_{ij}$  pertenecientes a enlaces terminales son multiplicados por 0.01; mientras los demás elementos se los fija a un valor de 0.001.

### **Índices ETA**

Los índices del átomo topoquímico ampliado (ETA) (Roy & Ghosh, 2003; Roy & Das, 2012; Roy *et al.*, 2015b), se calculan a partir de un grafo molecular libre de hidrógenos; en el que cada vértice se asume que está compuesto por un núcleo y un entorno electrónico de valencia. La electronegatividad ETA se obtiene combinando el conteo del núcleo de un átomo con su respectivo número de electrones de valencia.

### **Índices de Adyacencia de Arista**

Los índices de adyacencia de arista (Kode srl., 2016) son índices calculados a partir de la matriz de adyacencia, la cual brinda información relacionada a la conectividad entre las aristas (enlaces) del grafo molecular.

### **Índices del estado Electrotopológico por tipo de átomo**

Los índices del estado Electrotopológico por tipo de átomo (Kier & Hall, 1999), son descriptores moleculares usados para describir a una molécula mediante la información relacionada a sus diferentes tipos de átomos. Comúnmente, estos índices se derivan a partir de algunas propiedades provenientes de átomos del mismo tipo y de su entorno estructural. Así, estos índices combinan la información estructural sobre la accesibilidad electrónica asociada a cada tipo de átomo, un indicador de la presencia/ausencia de dicho tipo de átomo y el número de ocurrencia del mismo.

Cada tipo de átomo se define inicialmente por su identidad (basada en el número atómico  $Z$ ) y su estado de valencia el cual está basado en el identificador VSI:

$$VSI = \delta^v + \delta \quad (2.9)$$

donde  $\delta^v$  es el grado de vértice de valencia y  $\delta$  es el grado de vértice del átomo.

### 2.7.4. Descriptores tipo P\_VSA

Los descriptores tipo P\_VSA (Labute, 2000) indican la cantidad de área superficial de van der Waals (VSA) con una determinada propiedad P dentro de un rango definido. Las propiedades P que se usan son: logP, refractividad molar, masa, volumen de van der Waals, electronegatividad de Sanderson, polarizabilidad, energía de ionización, estado intrínseco. De esta forma, los descriptores tipo P\_VSA corresponden a una partición del área superficial molecular condicionada por los valores atómicos de P.

**Tabla 2.2.** Potenciales farmacóforos

| PPP                                 | Tipos  |
|-------------------------------------|--|
| Donante de enlaces de hidrógeno (D) | Átomo de oxígeno de un grupo OH                                      |
| Aceptor de enlaces de hidrógeno (A) | Átomo de nitrógeno de un grupo NH o NH <sub>2</sub>                  |
|                                     | Átomo de oxígeno   |
|                                     | Átomo de nitrógeno no adyacente a un átomo de hidrógeno              |
| Positivo (P)                        | Átomo con carga positiva   |
|                                     | Átomo de nitrógeno de un grupo NH <sub>2</sub>                       |
| Negativo (N)                        | Átomo con carga negativa   |
|                                     | Átomo de carbono, azufre o fósforo de un grupo COOH, SOOH o POOH     |
| Lipofílico (L)                      | Átomo de cloro, bromo o yodo   |
|                                     | Átomo de azufre adyacente únicamente a dos átomos de carbono (C-S-C) |
|                                     | Átomo de carbono adyacente únicamente a átomos de carbono            |
| Aromático (ar)                      | Átomo en una subestructura aromática                                 |
| Conjugado (con)                     | Átomo en una subestructura conjugada                                 |
| Halógeno (hal)                      | Cualquier átomo halógeno   |
| Cíclico (cyc)                       | Átomo que pertenece a ciclos   |
| Terminal (ter)                      | Átomo con un enlace no hidrogénico                                   |

Recientemente se han propuesto los descriptores P\_VSA\_ppp\_w (Kode srl., 2016), los cuales se basan en la presencia de potenciales farmacóforos (PPP); es decir, dado un determinado PPP, los descriptores P\_VSA se calculan como la suma de las contribuciones VSA de todos los átomos asignados a aquel PPP. Los potenciales farmacóforos se detallan en la Tabla 2.2.

### 2.7.5. Propiedades moleculares

Engloban a descriptores moleculares heterogéneos que describen propiedades fisicoquímicas y biológicas, y ciertas características obtenidas por modelos teóricos (Kode srl., 2016). Algunas de estas propiedades son:

- Conteo de insaturaciones e índice de insaturación.
- Factor hidrofílico.
- Refractividad molar.
- Área de superficie polar topológica.
- Coeficientes de partición octanol–agua de Moriguchi y de Ghose–Crippen–Viswanadhan.
- Volúmenes de van der Waals calculados a partir del volumen de McGowan y de la ecuación de Zhao–Abraham–Zissimos.
- Índice de densidad de empaquetamiento.

### **2.7.6. Fragmentos centrados en el átomo y número de grupos funcionales**

Estos dos tipos de descriptores resultan del conteo de fragmentos y grupos funcionales presentes en la molécula. El algoritmo de cálculo considera la composición química y las conectividades atómicas. Los fragmentos centrados en el átomo (Viswanadhan *et al.*, 1989; Ghose *et al.*, 1998) realizan conteos de los distintos fragmentos (tipos de átomos específicos) presentes en la estructura molecular, en la que cada fragmento es un átomo en la molécula descrito por sus átomos vecinos. Por otro lado, el número de grupos funcionales indica el número de grupos funcionales químicos presentes en una molécula (Kode *srl.*, 2016).

### **2.7.7. Descriptores geométricos**

Los descriptores geométricos se derivan de una representación 3D de la estructura molécula; es decir, a partir de la posición de los átomos en el espacio cartesiano XYZ. También se pueden calcular descriptores geométricos derivados de una representación a partir de un grafo 3D (índices topográficos), donde además de las posiciones de los átomos, también se consideran las conexiones entre los mismos. En cualquiera de los casos, estos descriptores se calculan a partir de una geometría molecular optimizada por métodos computacionales o a partir de las coordenadas cristalográficas (Mauri *et al.*, 2016).

Los descriptores geométricos proporcionan información adicional a la proporcionada por los descriptores constitucionales y topológicos. Sin embargo su uso es “complejo”, debido a que los valores que toman estos descriptores van a depender de dos aspectos fundamentales:

1. Tipo de optimización geométrica utilizada, la cual puede ser computacionalmente costosa y requiere tiempo considerable cuando la base de datos a analizar es grande.

2. Flexibilidad molecular, la cual indica que un compuesto químico se encuentra en diferentes conformaciones moleculares.

### **Descriptores RDF**

Los descriptores de función de distribución radial (RDF) (Hemmer *et al.*, 1999) se basan en la distribución de las distancias en la geometría molecular, los cuales se pueden interpretar como la distribución de probabilidad de encontrar un átomo dentro de un volumen esférico de radio R.

### **Descriptores 3D-MoRSE**

Los descriptores de representación molecular 3D de estructuras basadas en la difracción de electrones (3D-MoRSE) (Schuur *et al.*, 1996) obtienen información a partir de las coordenadas XYZ mediante la transformación usada en los estudios de difracción de electrones para obtener curvas de dispersión teóricas.

### **Descriptores WHIM**

Los descriptores moleculares invariantes holísticos ponderados (WHIM) (Todeschini *et al.*, 1994; Todeschini & Gramatica, 2002) son descriptores geométricos que se basan en índices estadísticos calculados a partir de las proyecciones de los átomos a lo largo de los ejes principales. Estos descriptores capturan información importante relacionada con el tamaño, forma y simetría de la molécula, así como la distribución de átomos con respecto a los ejes principales de la molécula (marco de referencia invariante). De esta forma, los átomos se proyectan en cada uno de los ejes y se evalúa la dispersión y distribución alrededor del centro geométrico.

### **Descriptores GETAWAY**

Por otro lado, los descriptores de ensamblado de pesos de átomos, geometría y topología (GETAWAY) (Consonni *et al.*, 2002a; Consonni *et al.*, 2002b) son derivados de la matriz de influencia molecular (**H**), la cual es una matriz tipo "influencia" similar a la definida para el diagnóstico en regresión.

$$\mathbf{H} = \mathbf{M} \left( \mathbf{M}^T \mathbf{M} \right)^{-1} \mathbf{M}^T \quad (2.10)$$

La matriz de información molecular **M**, es una matriz simétrica que consiste en coordenadas cartesianas centradas con respecto a todos los átomos de la molécula. Los elementos diagonales ( $h_{ii}$ ) de la matriz **H** son denominados valores de influencia o leverages y codifican la información de la influencia de cada átomo de la molécula en la determinación de su

forma; mientras que cada elemento fuera de la diagonal ( $h_{ij}$ ) indica la capacidad de interacción entre el  $i$ -ésimo y el  $j$ -ésimo átomo.

### **Perfiles moleculares de Randić**

Los perfiles moleculares de Randić (Randić, 1995a, 1995b; Randić & Razinger, 1995) son secuencias de descriptores propuestos por Milan Randić y calculados a partir de las distancias interatómicas geométricas de la molécula. Debido a que estos descriptores caracterizan bien la forma de las moléculas, son particularmente útiles para realizar análisis de similitud/diversidad.

### **Descriptores de carga y descriptores cuánticos**

Los descriptores de carga o descriptores electrónicos (Todeschini & Consonni, 2009; Roy *et al.*, 2015a) describen la distribución de cargas en una molécula o en regiones particulares de la misma, como por ejemplo átomos, enlaces, fragmentos moleculares. Para calcular estos descriptores la molécula debe haber sido previamente optimizada mediante algún método mecanocuántico apropiado (por ejemplo los métodos semiempíricos) (García *et al.*, 2016). Las cargas eléctricas presentes en la molécula son la fuerza motriz de las interacciones electrostáticas, y es bien conocido que la densidad de electrones local o cargas tienen un rol importante en muchas propiedades fisicoquímicas y reacciones químicas (Roy *et al.*, 2015a). Entre los descriptores cuánticos usados en estudios QSAR/QSPR se tiene (Karelson *et al.*, 1996; Doucet & Panaye, 2010):

1. Dipolo o momento dipolar: es un descriptor usado para describir la polaridad de una molécula; representando el comportamiento de resistencia y orientación de la misma en presencia de un campo electrostático. Se calcula usando las cargas atómicas parciales y las coordenadas atómicas.
2. Orbital molecular de más alta energía ocupado (HOMO): es el nivel de energía más alto en la molécula que contiene electrones y sirve para medir la nucleofilicidad de la misma. Cuando la molécula actúa como una base de Lewis (es decir, un donante de pares de electrones) en la formación de enlaces, los electrones son donados por este orbital.
3. Orbital molecular no ocupado de más baja energía (LUMO): es el nivel de energía más bajo en la molécula que no contiene electrones y sirve para medir la electrofilicidad de la misma. Cuando la molécula actúa como un ácido de Lewis (aceptor de pares de electrones) en la formación de enlaces, los electrones entrantes se dirigen a este orbital.

4. Superdelocalizabilidad: es un descriptor de reactividad de los orbitales ocupados y desocupados. Indica la contribución realizada por un átomo específico de la molécula a la energía de estabilización durante la formación de complejos de transferencia de carga con otra molécula; así como a la capacidad de un compuesto de formar enlaces a través de la transferencia de carga.
5. Polarizabilidad: es un descriptor que representa la habilidad de una molécula para formar dipolos instantáneos, y que se encuentra relacionado con la hidrofobicidad. Asimismo, la polarizabilidad electrónica de una molécula comparte características comunes con la superdelocalizabilidad electrofílica.

### **2.7.8. Descriptores topo-geométricos**

En este grupo se detallan aquellos tipos de descriptores que, dada su definición, se pueden calcular ya sea a partir de una representación topológica o una representación geométrica de la molécula.

#### ***Pares de átomos***

Los pares de átomos (Carhart *et al.*, 1985) son representaciones de cadena de la estructura química que consideran pares de átomos (excepto los pares con el átomo de hidrógeno) y la separación interatómica entre ellos. Los pares de átomos 2D consideran la distancia topológica (por ejemplo de 1 a 10) y los pares de átomos 3D usan la distancia Euclidiana para medir la separación entre átomos.

#### ***Descriptores CATS***

Los descriptores CATS (búsqueda de plantillas químicamente avanzadas) (Schneider *et al.*, 1999; Renner *et al.*, 2006) son similares a los pares de átomos. La definición del tipo de átomo en estos descriptores está relacionado con la presencia de potenciales farmacóforos (PPP), donde un PPP es un tipo de átomo generalizado definido teniendo en cuenta algunos aspectos fisicoquímicos. Los PPP propuestos para el cálculo de los descriptores CATS son los cinco primeros PPP descritos en la Tabla 2.2: donante de enlaces de hidrógeno (D), aceptor de enlaces de hidrógeno (A), positivo (P), negativo (N), lipofílico (L). De esta manera, en los descriptores CATS 2D, a cualquier átomo de una molécula se le puede asignar cero, uno o dos tipos de PPP; usándose la distancia topológica (entre 0 y 9) para medir la distancia entre ellos. Por otro lado, los descriptores CATS 3D (Fechner *et al.*, 2003) no permiten la asignación de múltiples PPP y

utilizan la distancia Euclidiana como medida de distancia en el espacio tridimensional.

### ***Descriptores basados en la matriz 2D y 3D***

Los descriptores basados en la matriz 2D (Kode srl., 2016) son índices topológicos que se calculan mediante la aplicación de un conjunto de operadores algebraicos a las diferentes matrices grafo–teóricas obtenidas a partir de un grafo molecular libre de hidrógenos. Análogamente, los descriptores basados en la matriz 3D o índices topográficos se calculan a partir de la matriz de distancias geométricas, en la que se considera la distancia Euclidiana en lugar de la distancia topológica para medir la separación entre los átomos.

### ***Autocorrelaciones 2D y 3D***

Las autocorrelaciones 2D (Kode srl., 2016) describen cómo una determinada propiedad se distribuye a lo largo de la estructura molecular topológica. Son autocorrelaciones calculadas a partir de un grafo molecular libre de hidrógenos ponderado por las propiedades fisicoquímicas escaladas con respecto al valor del átomo de carbono. Aquí se incluyen las autocorrelaciones de Broto–Moreau, autocorrelaciones centradas de Broto–Moreau, autocorrelaciones de Geary, autocorrelaciones de Moran y autocorrelaciones de carga topológica. Por otro lado, las autocorrelaciones 3D son autocorrelaciones espaciales basadas en las distancias geométricas (Euclidianas) entre los átomos presentes en la superficie molecular. Un tipo particular de estos descriptores son las autocorrelaciones basadas en las distancias topológicas (TDB) (Klein *et al.*, 2004), en el cual se consideran contemporáneamente las distancias topológicas y geométricas a un desplazamiento determinado.

## **2.8 Descriptores para moléculas desconectadas**

La mayoría de los descriptores moleculares existentes han sido desarrollados para capturar la información estructural de moléculas orgánicas, pequeñas y con todos sus átomos conectados. La disponibilidad de datos para sistemas conformados por estructuras desconectadas, ha permitido que en estudios recientes se evalúe la factibilidad de aplicar los descriptores moleculares clásicos a este tipo de estructuras. Los sistemas de estructuras moleculares desconectadas comprenden las sales, mezclas de compuestos y los líquidos iónicos. Indudablemente, la caracterización de sistemas con estructuras desconectadas es mucho más compleja, debido a que se debería también considerar las interacciones entre las mezclas que conforman el sistema (Mauri *et al.*, 2016).

Uno de los principales inconvenientes cuando se trabaja con sistemas que contienen estructuras desconectadas, es saber la forma en que se

deben representar dichas estructuras mediante descriptores moleculares adecuados (Oprisiu *et al.*, 2013). Los descriptores calculados sobre estos sistemas, según el algoritmo de cálculo, se pueden dividir en dos clases:

1. Descriptores moleculares cuya definición matemática es invariante a la conexión de todos los átomos en la molécula.
2. Descriptores moleculares que pueden ser calculados únicamente a partir de grafos moleculares donde todos los átomos estén conectados.

Los descriptores moleculares del primer tipo incluyen a los constitucionales, el conteo de grupos funcionales y fragmentos centrados en el átomo. La interpretación de estos descriptores para las estructuras conectadas y estructuras desconectadas es la misma. Por otro lado, para el segundo tipo de descriptores se requiere establecer la forma en que serán calculados en cada sistema de moléculas desconectadas. Una estrategia es calcular los descriptores sobre cada molécula conectada del sistema y posteriormente obtener un único valor mediante la suma, el promedio o el valor absoluto de la diferencia de los valores del descriptor (Oprisiu *et al.*, 2013; Cassotti *et al.*, 2014).

## 2.9 Huellas digitales moleculares

Las huellas digitales moleculares (FPs) (Shemetulskis *et al.*, 1996) son vectores booleanos (cuyos elementos se denominan bits) de dimensión fija que definen un conjunto de patrones de la molécula en un índice. Los patrones pueden ser un trayecto de longitud predefinida, siendo cada trayecto caracterizado por la naturaleza de los átomos y enlaces a través del mismo, o una subestructura circular a partir de un átomo específico. Las FPs se generan de manera tal que permitan capturar las características estructurales locales presentes en las moléculas. Particularmente identifican un grupo de fragmentos que componen la estructura molecular.

Existen dos tipos principales de patrones que se pueden identificar: trayectos y átomos centrados. En consecuencia, se pueden obtener 2 tipos de FPs: huellas digitales moleculares de trayectoria (PFPs) y huellas digitales moleculares de conectividad ampliada (ECFPs). Los patrones se generan a partir de la propia molécula y debido a que estos patrones difieren de un compuesto a otro, el significado de cualquier bit no está bien definido.

El número de los distintos fragmentos producidos por cualquier estructura puede ser muy largo, y va a depender del tipo de patrón usado y el tamaño de la molécula. Con la finalidad de reducir la longitud, las huellas digitales se tratan mediante una función de hash, en el que cada patrón se usa como semilla para un generador de números pseudo-casuales, cuya salida es un conjunto de bits. El algoritmo de

hashing produce varias características estructurales que pueden colisionar; es decir, ser representadas por un mismo bit. La ausencia de un bit significa que la característica no está presente en la molécula; mientras que la presencia del bit simplemente sugiere que la característica podría estar presente.

Las huellas digitales moleculares fueron inicialmente desarrolladas para la búsqueda rápida de bases de datos, debido a que son capaces de identificar subestructuras comunes. Sin embargo, las FPs también se las aplica en estudios QSAR, particularmente en el desarrollo de modelos basados en similitudes locales (Cassotti *et al.*, 2014).

### **2.9.1. Huellas digitales moleculares de conectividad ampliada**

Las huellas digitales moleculares de conectividad ampliada (ECFPs) (Rogers & Hahn, 2010) son FPs topológicas circulares que codifican la presencia/ocurrencia de fragmentos centrados en el átomo identificados mediante una variante del algoritmo de conectividad ampliado de Morgan. Este tipo de representación molecular también se denomina huellas digitales moleculares de Morgan o huellas digitales moleculares circulares (Mauri *et al.*, 2016). Las ECFPs se las puede aplicar en campos como:

- Cribado de alto rendimiento: para identificar moléculas falsas positivas y/o falsas negativas.
- Barrido virtual de bibliotecas: para identificar los compuestos activos.
- Análisis de conglomerados para analizar similitudes moleculares.
- Reconocimiento de patrones moleculares.
- Relaciones cuantitativas estructura–actividad basadas en similitudes locales.

#### ***Generación de patrones***

Constituye el primer paso para el cálculo de las ECFPs y consiste en la identificación de todas las subestructuras posibles en una molécula. Para las ECFPs, los fragmentos centrados en el átomo constituyen subestructuras circulares que generan “árboles” de una longitud establecida, fijada en un vértice particular del grafo molecular (Kode srl., 2016) y que dependerá de la longitud del patrón. La Figura 2.3 esquematiza las etapas para la generación de huellas digitales moleculares de conectividad ampliada.

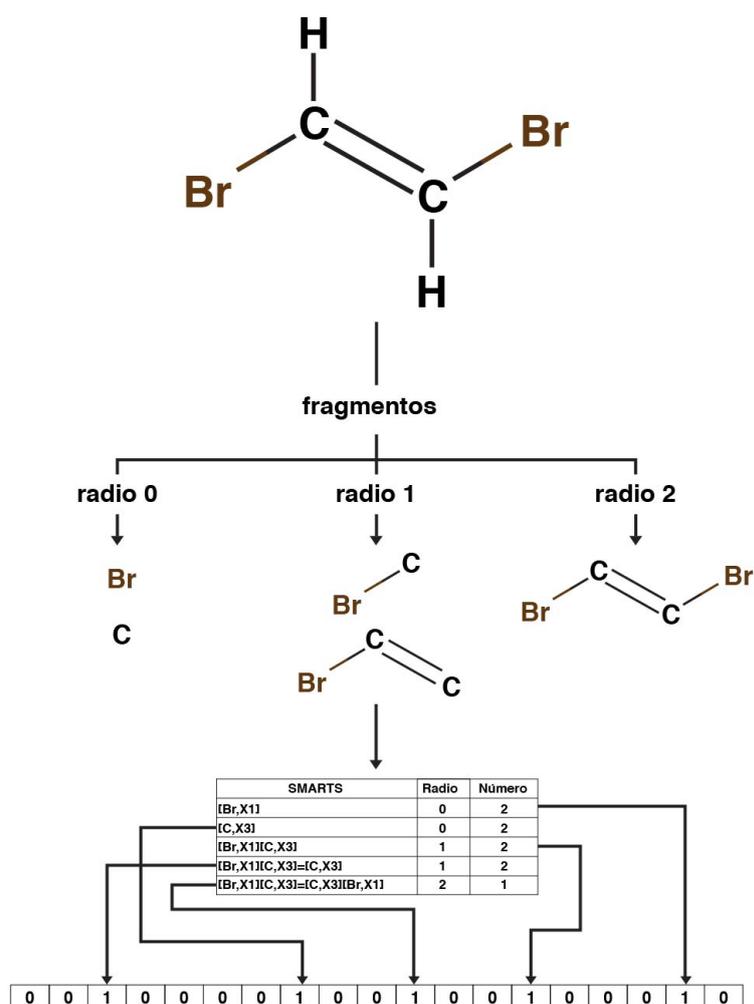
### ***Longitud del patrón***

La longitud del patrón indica el radio máximo a ser explorado desde el átomo centrado; es decir, el número de enlaces máximos a recorrer. La longitud del patrón indica la forma en que se explora el entorno circundante al átomo central. Por ejemplo, a radio 0 se obtienen subestructuras representadas únicamente por los átomos individuales, a radio 1 se obtiene la información de todos los átomos directamente enlazados al átomo centrado; mientras que a radio 2 se obtiene la información de todos los átomos dentro de un diámetro de 4 enlaces químicos. Seguidamente, cada fragmento se codifica mediante la notación lineal SMARTS (Daylight Chemical Information Systems, 2011). Finalmente se eliminan fragmentos iguales o equivalentes; es decir, aquellos fragmentos que contienen exactamente el mismo grupo de enlaces. Generar mayor cantidad de subestructuras permite obtener mayor información estructural; pero contemporáneamente aumenta el porcentaje de bits igual a 1 (opacidad).

### ***Función de Hash***

Es un algoritmo matemático que permite llevar varios vectores de entrada con dimensión variable a una dimensión fija, de tal forma que sea posible el tratamiento matricial de los datos. En el caso de las ECFPs, esta función se usa para mapear una subestructura dada dentro de un vector booleano de una longitud fija (dimensión final de la huella digital molecular) y se usa el operador lógico "o" para generar una huella digital molecular de toda la molécula. Es decir, cada subestructura es mapeada a la FP final mediante un cierto número de bits. Esta función es determinística, lo que permite que cada subestructura sea siempre mapeada por el mismo conjunto de bits.

El número de bits por patrón corresponde al número de bits que se usan para codificar con la función de hash una misma subestructura. La probabilidad de colisión se reduce al aumentar el número de bits por patrón, pero al mismo tiempo incrementa la opacidad.



**Figura 2.3.** Esquema de generación de huellas digitales moleculares de conectividad ampliada para el 1,2-dibromoetileno

## 2.10 Programas para el cálculo de descriptores moleculares y huellas digitales moleculares

El campo de los descriptores moleculares y huellas digitales está ganando mayor popularidad e interés por parte de los investigadores, con el desarrollo de programas y plataformas en línea orientadas a calcular dichas características químicas de las moléculas y para posteriormente usarlas para la predicción de actividades/propiedades moleculares. Una lista de los principales programas y plataformas para el cálculo de descriptores y/o huellas digitales se muestra en la Tabla 2.3. Existen programas de acceso libre (CORAL, EPI Suite™, PaDEL-Descriptor, etc.) y programas con licencia comercial (CODESSA PRO, DRAGON, etc.).

**Tabla 2.3.** Lista de programas y plataformas en línea disponibles para el cálculo de descriptores moleculares y huellas digitales moleculares

| Programa o Plataforma | Enlace WEB  |
|-----------------------|---|
| ADAPT                 | <a href="http://research.chem.psu.edu/pcjgroup/adapt.html">http://research.chem.psu.edu/pcjgroup/adapt.html</a>     |
| ADMET Predictor       | <a href="http://www.simulations-plus.com/">http://www.simulations-plus.com/</a>                                     |
| ADRIANA.Code          | <a href="https://www.mn-am.com/products/adriana-code">https://www.mn-am.com/products/adriana-code</a>               |
| Cerius <sup>2</sup>   | <a href="http://lms.chem.tamu.edu/cerius2.html">http://lms.chem.tamu.edu/cerius2.html</a>                           |
| CODESSA PRO           | <a href="http://www.codessa-pro.com/">http://www.codessa-pro.com/</a>   |
| CORAL                 | <a href="http://www.insilico.eu/coral/CORALSEA.html">http://www.insilico.eu/coral/CORALSEA.html</a>                 |
| CORINA Symphony       | <a href="https://www.mn-am.com/products/corinasymphony">https://www.mn-am.com/products/corinasymphony</a>           |
| DRAGON                | <a href="https://chm.kode-solutions.net/products_DRAGON.php">https://chm.kode-solutions.net/products_DRAGON.php</a> |
| E-DRAGON              | <a href="http://www.vcclab.org/lab/eDRAGON/">http://www.vcclab.org/lab/eDRAGON/</a>                                 |
| EPI Suite™            | <a href="https://www.epa.gov/tsca-screening-tools">https://www.epa.gov/tsca-screening-tools</a>                     |
| GRID                  | <a href="http://www.moldiscovery.com/software/grid/">http://www.moldiscovery.com/software/grid/</a>                 |
| ISIDA Fragmentor      | <a href="http://infochim.u-strasbg.fr/">http://infochim.u-strasbg.fr/</a>   |
| JOELib                | <a href="http://www.ra.cs.uni-tuebingen.de/">http://www.ra.cs.uni-tuebingen.de/</a>                                 |
| MARVIN Beans          | <a href="https://www.chemaxon.com/">https://www.chemaxon.com/</a>   |
| MOE                   | <a href="http://www.chemcomp.com/">http://www.chemcomp.com/</a>   |
| Molcode               | <a href="http://www.molcode.com/">http://www.molcode.com/</a>   |
| Molconn-Z             | <a href="http://www.edusoft-lc.com/molconn/">http://www.edusoft-lc.com/molconn/</a>                                 |
| MOLE-db               | <a href="http://michem.disat.unimib.it/mole_db/">http://michem.disat.unimib.it/mole_db/</a>                         |
| MOLGEN                | <a href="http://www.molgen.de/">http://www.molgen.de/</a>   |
| Open Babel            | <a href="http://openbabel.org/">http://openbabel.org/</a>   |
| PaDEL-Descriptor      | <a href="http://www.yapcwsoft.com/dd/padeldescriptor/">http://www.yapcwsoft.com/dd/padeldescriptor/</a>             |
| PCLIENT               | <a href="http://www.vcclab.org/lab/pclient/">http://www.vcclab.org/lab/pclient/</a>                                 |
| Pentacle              | <a href="http://www.moldiscovery.com/software/pentacle/">http://www.moldiscovery.com/software/pentacle/</a>         |
| PowerMV               | <a href="http://www.niss.org/research/software/powermv">http://www.niss.org/research/software/powermv</a>           |
| PreADMET              | <a href="https://preadmet.bmdrc.kr/">https://preadmet.bmdrc.kr/</a>   |
| QikProp               | <a href="https://www.schrodinger.com/qikprop">https://www.schrodinger.com/qikprop</a>                               |
| RDKit                 | <a href="http://www.rdkit.org/">http://www.rdkit.org/</a>   |
| ToMoCoMD-CARDD        | <a href="http://tomocomd.com/">http://tomocomd.com/</a>   |
| VolSurf+              | <a href="http://www.moldiscovery.com/software/vsplus/">http://www.moldiscovery.com/software/vsplus/</a>             |

Durante el desarrollo de la presente tesis se ha usado el programa DRAGON (versión 6 y versión 7) para el cálculo de los descriptores moleculares y de las huellas digitales de conectividad ampliada.

### 2.10.1. DRAGON

El programa DRAGON como tal fue lanzado en el año 1997; sin embargo, este programa nació en el año 1994 bajo el nombre de WHIM/3D QSAR con la finalidad de calcular los descriptores WHIM. DRAGON calcula descriptores moleculares a partir de diferentes representaciones y teorías, los cuales posteriormente se pueden usar en estudios QSAR/QSPR. Desde el año 1997, el programa se actualiza con frecuencia mediante la inclusión de nuevos descriptores y bloques de descriptores, algoritmos que permitan optimizar su precisión y tiempo de cómputo, capacidad de leer diversos tipos de formatos de representación molecular (Mauri *et al.*, 2006). El nombre del programa se debe al dragón mitológico que se encuentra en las puertas de Istar de Babilonia en el museo de Pérgamo en Berlín, cuyo cuerpo está formado por otros diversos animales (Figura 2.4). De esta forma, un descriptor

molecular puede ser pensado como este dragón, es decir, tener diversos significados los cuales dependen del punto de vista (Todeschini & Consonni, 2009).



**Figura 2.4.** Mušhuššu: Dragón mitológico de las puertas de Ištar de Babilonia (museo de Pérgamo)

DRAGON es uno de los programas más usados para el cálculo de los descriptores moleculares, debido al gran número de descriptores disponibles y su capacidad ilimitada para trabajar con complejos de iones metálicos. La versión 6 (Talete srl, 2015), lanzada en el año 2010, permite el cálculo de 4885 descriptores moleculares; mientras que la versión 7 (Kode srl., 2016), disponible desde el año 2016, permite el cálculo de 5270 descriptores moleculares y huellas digitales moleculares, tanto para estructuras moleculares conectadas y desconectadas. Por otra parte, 62 propiedades moleculares se encuentran también disponibles en la aplicación dProperties (Talete srl, 2012).

Para el cálculo de los descriptores RDF, tipo P\_VSA, autocorrelaciones 3D, autovalores de Burden, 3D-MoRSE, WHIM y GETAWAY, se ha implementado el uso de seis propiedades atómicas como criterio de ponderación:

- Masa atómica (m).
- Volumen de van der Waals (v).
- Electronegatividad del átomo (e).
- Polarizabilidad del átomo (p).
- Energía de ionización (i).
- Estado intrínseco (s).

**Tabla 2.4.** Bloques lógicos de descriptores presentes en DRAGON

| Bloque                        | No. | Bloque                         | No. |
|-------------------------------|-----|--------------------------------|-----|
| Índices constitucionales      | 47  | Descriptores WHIM              | 114 |
| Descriptores de anillo        | 32  | Descriptores GETAWAY           | 273 |
| Índices topológicos           | 75  | Perfiles moleculares de Randić | 41  |
| Cuenta de caminos y trayectos | 46  | Conteo de grupos funcionales   | 154 |

|                                      |     |  |      |
|--------------------------------------|-----|--|------|
| Índices de conectividad              | 37  | Fragmentos centrados en el átomo                       | 115  |
| Índices de información               | 50  | Índices del estado Electrotopológico por tipo de átomo | 172  |
| Descriptores basados en la matriz 2D | 607 | Descriptores basados en la matriz 3D                   | 99   |
| Autocorrelaciones 2D                 | 213 | Autocorrelaciones 3D                                   | 80   |
| Autovalores de Burden                | 96  | CATS 2D  | 150  |
| Descriptores tipo P_VSA              | 55  | Pares de átomos 2D                                     | 1596 |
| Índices ETA                          | 23  | Pares de átomos 3D                                     | 36   |
| Índices de la matriz de adyacencia   | 324 | Descriptores de carga                                  | 15   |
| Descriptores geométricos             | 38  | Propiedades moleculares                                | 20   |
| Descriptores RDF                     | 210 | Índices tipo fármaco                                   | 28   |
| Descriptores 3D-MoRSE                | 224 | CATS 3D  | 300  |

Adicional al cálculo de descriptores, este programa permite realizar un análisis preliminar de los descriptores calculados:

- Correlación entre pares de descriptores.
- Análisis de componentes principales.
- Análisis gráfico de los descriptores.
- Importación y exportación de variables externas.
- Eliminación de descriptores constantes y casi constantes.

DRAGON también puede ser usado como nodo para el desarrollo de diagramas de flujo de trabajo QSAR/QSPR en el programa KNIME. Los 5270 descriptores moleculares calculados por la versión más reciente de DRAGON se encuentran agrupados en 30 bloques lógicos (Tabla 2.4).

## 2.11 Referencias

- Balaban, A. T. (1985). Applications of Graph Theory in Chemistry. *Journal of chemical information and computer sciences*, 25(3), 334-343.
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz Information Miner. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications*, (pp. 319-326): Springer.
- Bonchev, D. (2015). On the Concept for Overall Topological Representation of Molecular Structure. In S. C. Basak, G. Restrepo & J. L. Villaveces (Eds.), *Advances in Mathematical Chemistry and Applications*, (pp. 42-75): Elsevier.
- Burden, F. R. (1989). Molecular Identification Number for Substructure Searches. *Journal of chemical information and computer sciences*, 29(3), 225-227.
- Carhart, R. E., Smith, D. H., & Venkataraghavan, R. (1985). Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *Journal of chemical information and computer sciences*, 25(2), 64-73.
- Cassotti, M., Consonni, V., Mauri, A., & Ballabio, D. (2014). Validation and Extension of a Similarity-Based Approach for Prediction of Acute Aquatic Toxicity Towards *Daphnia magna*. *SAR and QSAR in Environmental Research*, 25(12), 1013-1036.
- Consonni, V., Todeschini, R., & Pavan, M. (2002a). Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *Journal of chemical information and computer sciences*, 42(3), 682-692.
- Consonni, V., Todeschini, R., Pavan, M., & Gramatica, P. (2002b). Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. *Journal of chemical information and computer sciences*, 42(3), 693-705.
- Consonni, V., & Todeschini, R. (2009). Molecular Descriptors. In T. Puzyn, J. Leszczynski & C. M. T. (Eds.), *Recent Advances in QSAR Studies: Methods and Applications*, (pp. 29-102): Springer.
- Consonni, V., & Todeschini, R. (2010). Structure-Activity Relationships by Autocorrelation Descriptors and Genetic Algorithms. In H. Lodhi & Y. Yamanishi (Eds.), *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, (pp. 60-93): IGI Global.
- da Silva, V. B., de Almeida, J. R., & da Silva, C. H. T. d. P. (2010). General Aspects of Molecular Interaction Fields in Drug Design. In

- C. A. Taft & C. H. T. d. P. da Silva (Eds.), *New Developments in Medicinal Chemistry*, vol. 1 (pp. 70-78).
- Daylight Chemical Information Systems, I. (2011). *Daylight Theory Manual*.
- Doucet, J. P., & Panaye, A. (2010). *Three Dimensional QSAR: Applications in Pharmacology and Toxicology*: CRC Press.
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687-698.
- Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but Verify: on the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *Journal of chemical information and modeling*, 50(7), 1189-1204.
- Franke, R., & Gruska, A. (2003). General Introduction to QSAR. In R. Benigni (Ed.), *Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens*, (pp. 1-40): CRC press.
- Garcia, J., Duchowicz, P. R., & Castro, E. A. (2016). Considering the Molecular Conformational Flexibility in QSAR Studies. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research: QSAR in Medicinal Chemistry*, (pp. 129-158): CRC Press.
- Gasteiger, J., & Engel, T. (2006). *Cheminformatics: A Textbook*: Wiley-VCH.
- Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1998). Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *The Journal of Physical Chemistry A*, 102(21), 3762-3772.
- Guha, R., & Willighagen, E. (2012). A Survey of Quantitative Descriptions of Molecular Structure. *Current topics in medicinal chemistry*, 12(18), 1946-1956.
- Hansch, C., Leo, A., & Hoekman, D. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*: American Chemical Society.
- Hemmer, M. C., Steinhauer, V., & Gasteiger, J. (1999). Deriving the 3D Structure of Organic Molecules from Their Infrared Spectra. *Vibrational spectroscopy*, 19(1), 151-164.
- Janežič, D., Miličević, A., Nikolić, S., & Trinajstić, N. (2015). *Graph-Theoretical Matrices in Chemistry*: CRC Press.
- Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical Descriptors in QSAR/QSPR Studies. *Chemical reviews*, 96(3), 1027-1044.
- Kier, L. B., & Hall, L. H. (1986). *Molecular Connectivity in Structure-Activity Analysis*: Research Studies.

- Kier, L. B., & Hall, L. H. (1999). *Molecular Structure Description. The Electrotopological State*: Academic Press.
- Klein, C. T., Kaiser, D., & Ecker, G. (2004). Topological Distance Based 3D Descriptors for Use in QSAR and Diversity Analysis. *Journal of chemical information and computer sciences*, 44(1), 200-209.
- Kode srl. (2016). Dragon version 7. Software for Molecular Descriptor Calculation. <http://chm.kode-solutions.net/>.
- Kubinyi, H. (1998). Similarity and Dissimilarity: A Medicinal Chemist's View. In *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*, vol. 2 (pp. 225-252): Kluwer Academic Publishers.
- Kubinyi, H. (2008). *QSAR: Hansch Analysis and Related Approaches* (Vol. 1): John Wiley & Sons.
- Labute, P. (2000). A Widely Applicable Set of Descriptors. *Journal of Molecular Graphics and Modelling*, 18(4), 464-477.
- Martin, Y. C. (2010). *Quantitative Drug Design. A Critical Introduction* (Second ed.): CRC Press.
- Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006). Dragon Software: An Easy Approach to Molecular Descriptor Calculations. *MATCH Communications in Mathematical and in Computer Chemistry*, 56(2), 237-248.
- Mauri, A., Consonni, V., & Todeschini, R. (2016). Molecular Descriptors. In J. Leszczynski (Ed.), *Handbook of Computational Chemistry*, (pp. 1-29). Dordrecht: Springer.
- Murray-Rust, P., & Rzepa, H. S. (2001). Chemical Markup, XML and the World-Wide Web. 2. Information Objects and the CMLDOM. *Journal of chemical information and computer sciences*, 41(5), 1113-1123.
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An Open Chemical Toolbox. *Journal of cheminformatics*, 3(33), 1-14.
- Oprisiu, I., Novotarskyi, S., & Tetko, I. V. (2013). Modeling of Non-Additive Mixture Properties Using the Online CHEmical Database and Modeling Environment (OCHEM). *Journal of cheminformatics*, 5(1), 1-7.
- Polansky, O. E. (1991). Elements of Graph Theory for Chemists. In D. Bonchev & D. Rouvray (Eds.), *Chemical Graph Theory: Introduction and Fundamentals*, vol. 1 (pp. 41-96).
- Randić, M. (1975). Characterization of Molecular Branching. *Journal of the American Chemical Society*, 97(23), 6609-6615.
- Randić, M. (1995a). Molecular Profiles Novel Geometry-Dependent Molecular Descriptors. *New journal of chemistry*, 19(7), 781-791.
- Randić, M. (1995b). Molecular Shape Profiles. *Journal of chemical information and computer sciences*, 35(3), 373-382.

- Randić, M., & Razinger, M. (1995). On Characterization of Molecular Shapes. *Journal of chemical information and computer sciences*, 35(3), 594-606.
- Randić, M. (1996). Molecular Bonding Profiles. *Journal of Mathematical Chemistry*, 19(3), 375-392.
- Renner, S., Fechner, U., & Schneider, G. (2006). Alignment-Free Pharmacophore Patterns—A Correlation-Vector Approach. In T. Langer & R. D. Hoffmann (Eds.), *Pharmacophores and pharmacophore searches*, (pp. 49-79): Wiley-VCH.
- Rogers, D., & Hahn, M. (2010). Extended-Connectivity Fingerprints. *Journal of chemical information and modeling*, 50(5), 742-754.
- Roy, K., & Ghosh, G. (2003). Introduction of Extended Topochemical Atom (ETA) Indices in the Valence Electron Mobile (VEM) Environment as Tools for QSAR/QSPR Studies. *Internet Electronic Journal of Molecular Design*, 2(9), 599-620.
- Roy, K., & Das, R. N. (2012). On Extended Topochemical Atom (ETA) Indices for QSPR Studies. In E. A. Castro & A. K. Haghi (Eds.), *Advanced Methods and Applications in Chemoinformatics: Research Progress and New Applications*, (pp. 380-412): IGI Global.
- Roy, K., Kar, S., & Das, R. N. (2015a). *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*: Springer.
- Roy, K., Kar, S., & Das, R. N. (2015b). *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*: Academic Press.
- Schneider, G., Neidhart, W., Giller, T., & Schmid, G. (1999). "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angewandte Chemie International Edition*, 38(19), 2894-2896.
- Schuur, J. H., Selzer, P., & Gasteiger, J. (1996). The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. *Journal of chemical information and computer sciences*, 36(2), 334-344.
- Shemetulskis, N. E., Weininger, D., Blankley, C. J., Yang, J., & Humblet, C. (1996). Stigmata: An Algorithm to Determine Structural Commonalities in Diverse Datasets. *Journal of chemical information and computer sciences*, 36(4), 862-871.
- Talete srl. (2012). dProperties. Software for Molecular Property Calculation. <http://www.talete.mi.it/>.
- Talete srl. (2015). Dragon version 6. Software for Molecular Descriptor Calculation. <http://www.talete.mi.it/>.
- Testa, B., & Kier, L. B. (1991). The Concept of Molecular Structure in Structure-Activity Relationship Studies and Drug Design. *Medicinal research reviews*, 11(1), 35-48.

- Todeschini, R., Lasagni, M., & Marengo, E. (1994). New Molecular Descriptors for 2D and 3D Structures. Theory. *Journal of chemometrics*, 8(4), 263-272.
- Todeschini, R., & Gramatica, P. (2002). New 3D Molecular Descriptors: the WHIM Theory and QSAR Applications. In H. Kubinyi, G. Folkers & Y. C. Martin (Eds.), *3D QSAR in drug design*, (pp. 355-380): Springer.
- Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*: Wiley-VCH.
- Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, 29(6-7), 476-488.
- Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989). Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *Journal of chemical information and computer sciences*, 29(3), 163-172.
- Weininger, D. (1988). SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *Journal of chemical information and computer sciences*, 28(1), 31-36.
- Young, D., Martin, T., Venkatapathy, R., & Harten, P. (2008). Are the Chemical Structures in Your QSAR Correct? *QSAR & Combinatorial Science*, 27(11) -12451337



# Capítulo 3

---

## MÉTODOS QUIMIOMÉTRICOS EN QSAR/QSPR

---

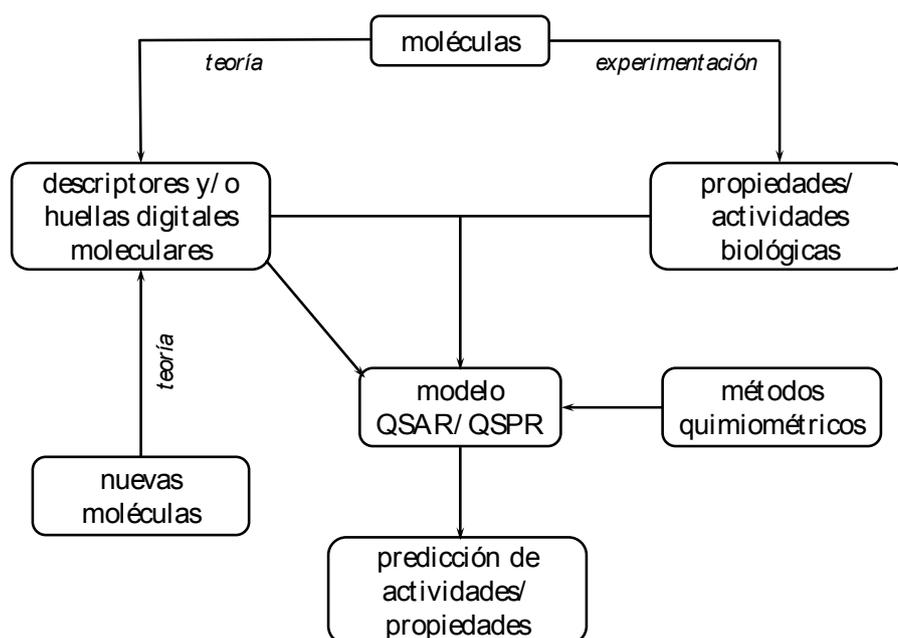
### 3.1 Introducción

La Quimiometría ha sido definida en términos generales por la Sociedad Internacional de Quimiometría (2002) como la ciencia de extraer información de las medidas realizadas en un sistema o proceso químico, mediante la aplicación de métodos matemáticos o estadísticos. Sin embargo, fue Svante Wold, en el año 1972, quien inventó el término quimiometría, con la finalidad de describir la disciplina de extraer información relevante de experimentos químicos. Posteriormente, Wold redefinió el concepto para indicar la forma de obtener y representar información relevante de datos medidos en sistemas químicos (Wold, 1995).

Una definición más precisa, y ceñida a la proposición de Svante Wold y Bruce Kowalsky cuando fundaron la primera Sociedad de Quimiometría en 1974 (Wold, 2015), es la siguiente: quimiometría es la disciplina química que usa la matemática, estadística y la lógica formal para: 1) diseñar o seleccionar procedimientos experimentales óptimos, 2) maximizar la información química relevante de un análisis de datos químicos, y 3) obtener conocimiento de los sistemas químicos en estudio (Massart *et al.*, 1997).

Uno de los fundamentos de la quimiometría es el uso de un enfoque multivariado para explorar sistemas complejos químicos y diseñar racionalmente los experimentos. Los sistemas complejos por naturaleza necesitan varias variables para ser descritos e indudablemente, la quimiometría ofrece los métodos adecuados para recabar la mayor información posible de dichos sistemas.

Las técnicas quimiométricas se caracterizan por analizar todas las variables simultáneamente, permitiendo obtener una visión holística del sistema complejo. En la Figura 3.1. se observa la importancia de la quimiometría en el desarrollo de QSAR/QSPR.



**Figura 3.1.** Importancia de la quimiometría en el desarrollo de modelos QSAR/QSPR

### 3.2 Estructura multivariada de los datos

El análisis multivariado involucra la observación y análisis de varias variables simultáneamente, tal como sucede en los sistemas complejos, donde existen muchas variables que describen una muestra o un sistema (conjunto de muestras). Para el análisis multivariado los datos son arreglados en forma de matriz (o matrices), la cual se denota como **X**, donde el número de filas ( $n$ ) representa las muestras y el número de columnas ( $p$ ) las variables. De esta forma, cada objeto de esta matriz,  $x_{ij}$ , representa el valor de la  $j$ -ésima variable para la  $i$ -ésima muestra. Cuando se trabaja con técnicas de modelado (regresión o clasificación), se presenta información adicional en forma de un vector respuesta, denotado con **Y**. En este vector, cada entrada  $y_i$ , contiene la respuesta cuantitativa o cualitativa para la  $i$ -ésima muestra.

En todas las aplicaciones de esta tesis, cada fila de la matriz **X** representa un compuesto químico para el cual se han calculado descriptores moleculares y/o huellas digitales moleculares, los cuales son representados en las columnas.

### 3.3 Pretratamiento de los datos

Previo a cualquier análisis multivariado, es siempre necesario un pretratamiento de los datos. El primer paso consiste en verificar que no existan errores de transcripción, así como la presencia de datos faltantes (representados como celdas vacías o con códigos numéricos propios de cada programa quimiinformático). Asimismo, es útil identificar si alguna

de las variables que describen el sistema son constantes (todos los valores iguales) o casi constantes (un valor diferente de todos los demás), dado que no aportan información y es siempre útil reducir la dimensionalidad del sistema.

Un control más exhaustivo consiste en verificar el tipo de variables que describen el sistema, pudiendo clasificarlas en:

- Variables binarias: indican la presencia o ausencia de la característica de dicha variable en la muestra. Se las identifica con 0 (ausencia) y 1 (presencia).
- Variables discretas ordinales: indican la presencia de una propiedad ordenada de forma ascendente (1, 2, 3, 4, etc.).
- Variables continuas: son las que pueden tomar cualquier valor dentro de un intervalo dado.

### 3.3.1. Escalado de los datos

El escalado de los datos (Vandeginste *et al.*, 1998) permite poner en una misma escala todas las variables en estudio independientemente de sus unidades de medida, es decir, permite confrontar las variables que provienen de diversas fuentes de medida dado que en el análisis multivariado es muy común el buscar relaciones entre tales variables; tal es el caso de todos los métodos que se basan en el cálculo de distancias: análisis de conglomerados, escalado multidimensional, clasificación, etc. Los tipos de escalado de variables comúnmente utilizados en quimiometría son los que se describen a continuación:

- *Centrado*: permite centrar los datos respecto al valor medio de cada variable, es decir, el valor central será igual a cero. El centrado no modifica la varianza de los datos.
- *Escalado con respecto al valor máximo*: en este tipo de escalado el valor máximo para cada variable es igual a 1.
- *Escalado de intervalo*: este escalado coloca un doble vínculo a cada variable, es decir, el valor mínimo es igual a cero y el valor máximo es igual a uno.
- *Autoescalado*: es el método de escalado de datos usado con mayor frecuencia, el cual consiste en un centrado, seguido de una normalización a varianza unitaria. Las variables autoescaladas tienen media cero y desviación estándar igual a uno.
- *Escalado a varianza unitaria*: es un tipo de escalado en el que se coloca como restricción que la varianza de la variable transformada sea igual a 1.

## **3.4 Técnicas de exploración de la estructura de los datos**

### **3.4.1. Análisis de componentes principales**

El análisis de componentes principales (PCA) (Wold *et al.*, 1987; Jolliffe, 2002; Bro & Smilde, 2014) es el método quimiométrico mayormente usado para mostrar la estructura multivariable de los datos. PCA fue descrito por primera vez por Karl Pearson en el año 1901 y desarrollado formalmente por Harold Hotelling en el año de 1933. En esta técnica se proyectan los datos en un hiperespacio ortogonal reducido y definido por las componentes principales significativas. Las componentes son combinaciones lineales de las variables originales, donde la primera componente tiene la máxima varianza, la segunda componente tiene la segunda máxima varianza y así sucesivamente. De esta forma es posible retener un número de componentes reducido en comparación al número de variables originales (reduce la dimensión de los datos). El número de componentes significativas se puede elegir en base de la varianza explicada por cada componente. Una de las ventajas que presenta PCA es una resolución matemática que permite de observar de manera gráfica la información, en el gráfico de puntuaciones (scores) o proyección de las muestras y el gráfico de cargas (loadings) o proyección de las variables. También se puede obtener un gráfico combinado (biplot) mediante la superposición del gráfico de puntuaciones y cargas. De esta manera, PCA ayuda a evaluar la correlación que existe entre las variables y su relevancia, identificar grupos, tendencias, datos atípicos (outliers), etc., mediante la visualización únicamente de la información útil.

### **3.4.2. Escalado multidimensional**

Un objetivo dentro de la visualización de los datos es el buscar cómo y qué tan cerca se encuentran las muestras entre sí. En algunos casos, esto se puede lograr realizando un diagrama de dispersión; sin embargo, cuando existen varias variables describiendo el sistema, se tendría que buscar una alternativa para visualizar los datos en un espacio multidimensional.

El escalado multidimensional (MDS) es una técnica que permite reconstruir las similitudes/disimilitudes entre pares de objetos mediante el uso de distancias y proyectarlas en un número pequeño de dimensiones (Kruskal, 1964; Winsberg & Carroll, 1989). Este significa que el MDS busca acomodar los objetos en un espacio definido por un cierto número de dimensiones de forma tal que se reproduzcan las distancias observadas. Así, el MDS permite obtener un gráfico de dispersión de los objetos en un espacio reducido, en el cual se observa

una reproducción de las distancias originales y que permite un análisis rápido y fácil de las relaciones entre los objetos.

Al igual que el PCA, se debe considerar la selección del número apropiado de dimensiones. Con MDS, el número de dimensiones se puede lograr en función de los residuos entre las distancias originales y las distancias representadas en las dimensiones seleccionadas; es decir, mientras más bajo es el residuo, mejor es la aproximación de las distancias en las proyecciones.

### **3.4.3. Análisis de conglomerados**

El objetivo del análisis de conglomerados es detectar las similitudes entre las muestras, con las cuales se busca identificar grupos consistentes en los datos (Massart & Kaufman, 1983). Es una técnica alternativa al análisis de componentes principales para describir la estructura de los datos. Las similitudes entre las muestras se construyen mediante distancias, es decir, muestras similares estarán a una distancia pequeña, mientras que muestras disímiles estarán a una distancia grande.

#### ***Métodos jerárquicos***

Los métodos jerárquicos usan las medidas de distancia (indicadores de vinculación), para cuantificar las similitudes entre grupos de muestras o conglomerados. Los métodos jerárquicos aglomerativos (Vandeginste *et al.*, 1998) requieren que se defina una medida de uso para el cálculo de la matriz de distancias y posterior definición de la correspondiente matriz de similitudes. El algoritmo inicia con un número de conglomerados igual al número de observaciones. Posteriormente, se identifican los dos objetos más similares y son unidos en un único nuevo conglomerado a un cierto nivel de similitud. Seguidamente, se recalculan las similitudes de este nuevo conglomerado con respecto a los objetos restantes y se une al objeto con mayor similitud. El proceso continúa hasta que se genera un único conglomerado que agrupa a todas las observaciones. El resultado de este proceso iterativo se muestra de forma gráfica en el dendrograma.

#### ***Métodos no jerárquicos***

Los métodos no jerárquicos se basan en técnicas denominadas de reposición. Aquí, luego de una partición inicial de los datos, los mismos se mueven entre los conglomerados hasta que se cumpla un cierto criterio. De estos métodos el más representativo es el *k*-medias (Kaufman & Rousseeuw, 2005), el cual reposiciona los datos de acuerdo a la comparación de las distancias de cada objeto con respecto al centroide de cada conglomerado. El número de conglomerados y la medida de distancia son establecidos por el usuario a priori.

El algoritmo inicia con una partición casual de los objetos en un número preestablecido de conglomerados, para los cuales se determina su centroide y se calcula la distancia de cada objeto con respecto a cada centroide. Los objetos se asignan al conglomerado más cercano y, si al menos un objeto se mueve a otro conglomerado, se recalculan los centroides y las asignaciones.

### 3.5 Técnicas de reducción de variables

Los métodos de reducción no supervisados de variables son técnicas quimiométricas útiles cuando se trabaja con bases de datos que contienen grandes cantidades de variables y en las que se puede encontrar redundancia, multicolinealidad, ruido y correlación casual. La presencia de estas variables irrelevantes puede cambiar el patrón de los datos e influenciar los modelos finales. La reducción no supervisada de variables se lleva a cabo sin considerar la respuesta experimental (Whitley *et al.*, 2000; Questier *et al.*, 2005; Consonni *et al.*, 2009). Es indudable que estas patologías estén presentes en las bases de datos QSAR/QSPR, donde se analizan las relaciones entre las actividades/propiedades y los descriptores moleculares. Por consiguiente, es importante que se retengan únicamente los descriptores relevantes, a partir de los cuales se puedan obtener modelos de regresión o clasificación parsimoniosos (es decir, con pocas variables) y que proporcionen predicciones confiables (Bagheri *et al.*, 2013).

Estas técnicas son comúnmente usadas como un filtro preliminar para remover variables según un criterio y de esta manera proporcionar bases de datos depuradas (variables relevantes), lo que se traduce en menor costo computacional durante la selección supervisada de variables, robustez del modelo final y fiabilidad de las predicciones realizadas con dicho modelo.

#### 3.5.1. Correlación entre pares de variables

Una práctica común en la reducción de las variables es la que se desarrolla por análisis de los pares de variables que se encuentran correlacionadas más allá de un cierto umbral de corte (*thr*), evitando así la presencia de redundancia y variables correlacionadas en los modelos QSAR/QSPR (Todeschini *et al.*, 2009). Este aspecto es particularmente importante cuando se desarrollan modelos lineales multivariados. El análisis de correlación entre pares de variables se realiza por comparación del coeficiente absoluto de correlación de Pearson ( $R_{ij}$ ) con el valor de umbral de corte preestablecido (normalmente 0.95). Si se cumple  $R_{ij} \geq thr$ , se eliminará el descriptor que presenta la mayor correlación promedio con las demás variables.

### 3.5.2. Método V–WSP

El método de reducción de variables V–WSP selecciona un subgrupo representativo de variables de un conjunto de datos, de tal forma que se encuentre una mínima correlación entre ellas en un espacio multidimensional definido (Ballabio *et al.*, 2014). Este método es una modificación del algoritmo propuesto por Wootton, Sergent y Phan–Tan–Luu (WSP) para diseño de experimentos, en el cual se selecciona un subconjunto de puntos a partir de un conjunto de candidatos de tal forma que se encuentren a una mínima distancia (por ejemplo la Euclidiana) en un espacio multidimensional definido.

Dada una matriz de datos de  $n \times p$ , el método de reducción V–WSP sigue el siguiente algoritmo:

1. Elegir una variable inicial  $j$  como semilla y un valor umbral de correlación ( $thr$ ).
2. Calcular el coeficiente lineal de correlación de Pearson ( $R$ ) entre la variable  $j$  y todas las demás
3. Eliminar las variables  $d$  cuyo valor absoluto  $R_{dj} \geq thr$
4. Se fija la variable  $j$  y se selecciona entre las restantes variables aquella que tenga la correlación absoluta más alta con  $j$ .
5. Repetir los pasos 2, 3 y 4 hasta que no existan variables para ser seleccionadas.

### 3.5.3. Otros métodos basados en técnicas multivariable

La reducción de variables se puede también realizar mediante un análisis de componentes principales, es decir, se retienen únicamente las variables que tienen valor alto de carga (loading) en las componentes significativas, o se eliminan las variables con valores altos de carga en las últimas componentes (Jolliffe, 1972, 1973). Otras formas de reducir variables se logran mediante: 1) la aplicación del análisis de conglomerados ( $k$ –medias, métodos jerárquicos, Jarvis–Patrick, etc.) a la matriz transpuesta de los datos; 2) la utilización de mapas de Kohonen (dada su simplicidad de uso y eficiencia); 3) el uso del diseño D–optimal (diseño experimental) ó 4) el uso del índice de correlación multivariado  $K$ , para ordenar las variables según su correlación global con todas las demás (Todeschini *et al.*, 2009).

## 3.6 Técnicas de selección de variables

### 3.6.1. Búsqueda exacta

La búsqueda exacta es el método de selección de variables más simple pero computacionalmente demandante. Este método garantiza que todos los posibles modelos de las  $p$  variables sea explorado, partiendo desde modelos con 1 variable hasta los de  $p$  variables. Esta característica lo convierte en un método demandante computacionalmente. El número de combinaciones posibles de las  $p$  variables está dada por  $2^p - 1$ . Por consiguiente, el método se torna engorroso cuando se tienen bases de datos con grandes números de variables.

Se puede reducir el tiempo de cálculo y la interpretabilidad de los modelos cuando se limita a desarrollar modelos con un cierto número  $k$  de variables. En este caso se deben calcular todas las posibles combinaciones de las  $p$  variables tomadas desde 1 hasta  $k$ . Por ejemplo, dado un número  $p$  de variables, el número total de modelos  $t$ , desde 1 hasta  $k$  está dado por la siguiente expresión:

$$t = \sum_k \left( \frac{p!}{k!(p-k)!} \right) \leq 2^p - 1 \quad (3.1)$$

Sin embargo, se debe tener en cuenta que el número de modelos es proporcional al número de variables ( $p$ ).

### 3.6.2. Métodos de regresión a pasos

La regresión a pasos (SWR) es un método muy conocido y ampliamente utilizado para la selección de variables. SWR consiste en dos estrategias: selección hacia adelante (FS) y eliminación hacia atrás (BE) (Hastie *et al.*, 2011). La selección hacia adelante comienza con un modelo de tamaño 0 y comienza a adicionar variables que cumplen con un criterio preestablecido hasta un modelo de tamaño deseado. La eliminación hacia atrás se realiza de forma inversa a FS, es decir, inicia con un modelo de tamaño igual al número  $p$  total de variables y va eliminando variables irrelevantes en cada paso. El criterio de inclusión o exclusión es la suma de los cuadrados de los residuos (RSS), de tal forma que a cada paso, la variable a ser incluida o eliminada es aquella que permite el mayor decrecimiento de RSS.

Posteriormente, el algoritmo SWR se mejoró al combinar FS y BE, es decir, se inicia con un FS y luego de que cada variable (diferente a la primera) ha sido adicionada al modelo, se evalúa con un BE si alguna de las variables se puede eliminar sin un incremento significativo del RSS (Efroymsen, 1960).

### 3.6.3. Algoritmos Genéticos

Los algoritmos genéticos (GAs) (Leardi *et al.*, 1992) se aplican a la selección de aquellas variables mediante la búsqueda de las variables que generen un máximo o mínimo de una respuesta; por ejemplo, el coeficiente de determinación ( $R_{cv}^2$ ) en regresión o la tasa de aciertos ( $NER_{cv}$ ) en clasificación. Debido a que es un método que se basa en la teoría de la evolución de Darwin, se parte de una población de “cromosomas” generada de manera casual. Un cromosoma es un vector binario de  $p$  bits (número total de variables), que se asocia a un modelo con ciertas variables de la siguiente manera: bits con valor 1 indican que los descriptores correspondientes están presentes en el modelo y viceversa para los bits nulos.

Durante una corrida, el tamaño de la población se mantiene constante; de esta manera, los mejores cromosomas entran en la misma y los peores son eliminados. Existen dos operaciones principales mediante las cuales se generan nuevas cromosomas:

1. Reproducción o crossover: se seleccionan dos padres a partir de la población y se generan cromosomas hijos. Esta selección puede ser casual o sesgada hacia los mejores cromosomas. Los cromosomas hijos comparten el patrimonio genético de los padres; es decir, los bits 0 o 1 en ambos padres se mantienen, mientras que los bits con valores diferentes serán fijados en 0 o 1 de acuerdo a una regla de probabilidad.
2. Mutación: en esta operación los cromosomas pueden invertirse de forma tal que se generan mutantes. El proceso de mutación se utiliza para limitar la posibilidad de que la población se quede atrapada en un mínimo local. La probabilidad de ocurrencia de mutación es menor a la probabilidad de reproducción, debido a que se debe evitar que la población se aleje de la probable región óptima.

Cuando los GAs se aplican a bases de datos que contienen grandes cantidades de variables o pocas observaciones, se corre el riesgo de sobreajuste en los modelos. Para evitar este problema, se ha propuesto una variante de los GAs (Leardi & Gonzalez, 1998; Leardi, 2009), en la cual en lugar de realizar una simple corrida y muchas interacciones, se realiza un número pequeño de corridas independientes a partir de varias poblaciones iniciales. Posteriormente, se registra la frecuencia de selección de las variables en dichas corridas. Finalmente, el modelo se construye adicionando la variable más frecuente (modelo de 1 variable) y posteriormente se adicionan las demás variables en función de su frecuencia de selección.

#### **3.6.4. Método de Reemplazo**

El método de reemplazo (RM) fue propuesto por Duchowicz et al. (Duchowicz *et al.*, 2005; Duchowicz *et al.*, 2006), como alternativa para buscar los mejores subconjuntos de descriptores en grandes bases de datos. RM requiere la generación de un número de regresiones inferior que la búsqueda exacta y produce resultados muy cercanos al mismo. Los pasos involucrados en la técnica RM son los siguientes:

1. Elegir  $d$  descriptores ( $X_1, X_2, \dots, X_d$ ) de forma aleatoria a partir de la base de datos con  $D$  descriptores y realizar una regresión lineal.
2. Seleccionar uno de los descriptores de este subconjunto ( $X_i$ ) y reemplazarlo por cada uno de los  $D$  descriptores manteniendo el mejor resultado.
3. Debido a que se puede reemplazar cualquiera de los  $d$  descriptores del modelo inicial, se tendrán  $d$  posibles rutas para alcanzar el resultado final. Por ejemplo, la elección anterior desarrollará la  $i$ -ésima ruta.
4. Elegir el descriptor que contiene el mayor error relativo en su coeficiente (excepto el reemplazado en el paso 2) y reemplazarlo con los restantes manteniendo el mejor subconjunto.
5. Reemplazar todos los descriptores restantes de la misma manera, excepto aquellos reemplazados en los pasos 2 y 4.
6. Al finalizar, iniciar nuevamente con el descriptor que presente el mayor error relativo y repetir todo el proceso. Este procedimiento se repite tantas veces como sea necesario hasta que el subgrupo de descriptores se mantenga invariante. En este punto se tiene el mejor modelo para la ruta  $i$ .
7. Proceder exactamente de la misma forma para todas las restantes rutas.
8. Comparar los modelos resultantes y seleccionar el mejor como el modelo QSAR/QSPR para validaciones posteriores.

### **3.7 Técnicas de modelado de datos**

#### **3.7.1. Métodos de regresión**

La regresión lineal múltiple (MLR) es un enfoque matemático que permite establecer relaciones lineales entre un grupo de variables independientes y una variable dependiente o respuesta cuantitativa (Rencher & Schaalje, 2008). Esta relación toma la siguiente forma:

$$y = \mathbf{X}\beta + e \quad (3.2)$$

donde  $y$  es el vector respuesta,  $\mathbf{X}$  es la matriz del modelo,  $\beta$  es el vector de los coeficientes de regresión verdaderos y  $e$  es el vector de errores. Un modelo matemático particular se expresa de la siguiente manera:

$$\hat{y} = \mathbf{X}b \quad (3.3)$$

donde  $b$  es el vector de las estimaciones de los coeficientes verdaderos  $\beta$  e  $\hat{y}$  es el vector de las respuestas calculadas.

### **Mínimos cuadrados ordinarios**

El método de los mínimos cuadrados ordinarios (OLS) (Rencher & Schaalje, 2008; Varmuza & Filzmoser, 2009), es el método más simple y el que se aplica con mayor frecuencia cuando se busca realizar una regresión. Sea  $\mathbf{X}$  la matriz de datos de dimensión  $n \times p$ ,  $\beta$  el vector de los coeficientes del modelo, de dimensión  $p$  ( $d+1$ ), el modelo OLS se define de la siguiente manera:

$$y_i = b_0 + \sum_{j=1}^p b_j x_{ij} \quad (3.4)$$

El método OLS brinda una estimación de los coeficientes de un modelo lineal por minimización de la suma de cuadrados de los residuos (RSS) entre el vector de respuestas calculadas y el vector de respuestas experimentales. RSS se define como:

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.5)$$

donde  $y_i$  es la respuesta experimental e  $\hat{y}_i$  es la respuesta calculada para la  $i$ -ésima observación. La sumatoria se realiza sobre las  $n$  observaciones.

Por otro lado, los coeficientes del modelo OLS,  $b_j$ , se calculan de la siguiente manera:

$$b = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y \quad (3.6)$$

Donde  $\mathbf{X}$  es la matriz del modelo e  $y$  es el vector de respuestas de dicha matriz.

La estimación de los coeficientes es influenciada por la presencia de multicolinealidad entre las variables del modelo, debido a problemas para la inversión de  $\mathbf{X}^T\mathbf{X}$ . Por tal motivo, se debe mantener una correlación baja entre las variables del modelo. También se presentan inconvenientes en la estimación de los coeficientes cuando el número de parámetros es mayor al número de observaciones ( $p > n$ ).

Una vez obtenidos los coeficientes de regresión, se pueden también obtener los coeficientes de regresión estandarizados de la siguiente forma:

$$b'_j = b_j \frac{s_j}{s_y} \quad (3.7)$$

donde  $s_j$  es la desviación estándar de la  $j$ -ésima variable y  $s_y$  la desviación estándar de la respuesta. Los coeficientes estandarizados son independientes de la escala de medida de las variables, por lo que permite evaluar la contribución de cada variable en el cálculo de la respuesta.

### **Mínimos cuadrados parciales**

El método de mínimos cuadrados parciales (PLS) (Wold *et al.*, 2001) es un método de regresión lineal que tiene algunas similitudes con el PCA. Este método trabaja con proyecciones de los objetos en un nuevo espacio definido por combinaciones lineales de las variables originales, denominadas variables latentes (LVs). Por este motivo, PLS es un método más apropiado que OLS cuando la relación observaciones/variables es menor a uno y cuando existe multicolinealidad en la matriz de diseño.

PLS es un enfoque que busca la dirección de máxima varianza en la matriz de diseño (similar a PCA), al mismo tiempo que la máxima correlación con el vector respuesta (Hastie *et al.*, 2011). Las variables deben ser escaladas apropiadamente (por ejemplo, el autoescalado) y cuando la actividad/propiedad lo amerita, se transforma a una escala logarítmica. Cuando el número de variables latentes es igual al número de descriptores, PLS produce el mismo modelo que se obtendría mediante OLS.

La proyección de los objetos en el nuevo espacio se desarrolla mediante la transformación:

$$\mathbf{T} = \mathbf{X}\mathbf{V} \quad (3.8)$$

Donde  $\mathbf{T}$  es la matriz de puntuaciones, es decir, las coordenadas de las observaciones en el nuevo espacio;  $\mathbf{X}$  es la matriz del modelo y  $\mathbf{V}$  es la matriz de cargas (loadings). Entonces, el modelo se calibra en el espacio de las LVs:

$$y = \mathbf{T}q' + \mathbf{e} \quad (3.9)$$

Donde  $q$  es el vector de los coeficientes de regresión PLS. Al reemplazar  $T$  de la Ec. 3.8 en la Ec. 3.9, se obtiene:

$$y = \mathbf{XV}q' + \mathbf{e} \quad (3.10)$$

La expresión anterior se tiene que resolver para  $q$  y se observa que de acuerdo a la Ec. 3.3 los coeficientes de las variables resulta ser  $b = \mathbf{V}q'$ .

PLS permite reducir la dimensionalidad de los datos mediante la selección de un número apropiado de LVs menor al número de variables. Normalmente, la selección del número óptimo de LVs se la realiza en validación cruzada.

### ***Parámetros de evaluación de los modelos de regresión***

Los parámetros de evaluación en regresión permiten evaluar qué tan bueno es un modelo, tanto en ajuste como en predicción (Rencher & Schaalje, 2008; Varmuza & Filzmoser, 2009). El punto de partida para los modelos de regresión es la suma total de los cuadrados (TSS), la cual se define como:

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.11)$$

donde  $\bar{y}$  es el valor medio de la respuesta.

Un modelo de regresión es mejor cuando más pequeña es la suma de los residuos (Ec. 3.5) o cuanto más grande es la suma de los cuadrados del modelo (MSS):

$$MSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.12)$$

donde  $\hat{y}_i$  es el valor predicho de la  $i$ -ésima muestra.

La suma total de los cuadrados es igual a la suma de MSS y RSS. Con estos tres parámetros se puede calcular el coeficiente de determinación, que indica la varianza explicada por el modelo de regresión:

$$R^2 = \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS} \quad (3.13)$$

El coeficiente de determinación,  $R^2 \in [0,1]$ , mide la capacidad de ajuste del modelo de regresión, es decir, la capacidad para describir la respuesta de los  $n$  objetos, sin considerar la capacidad predictiva. Para medir la predictividad de un modelo se usa la suma de los cuadrados de predicción (PRESS), que es análoga a RSS:

$$PRESS = \sum_{i=1}^n \left( y_i - \hat{y}_{i|i} \right)^2 \quad (3.14)$$

donde  $\hat{y}_{i|i}$  es el valor predicho para la  $i$ -ésima observación, en un modelo en el cual dicha observación no fue considerada para calibrar el modelo. Este parámetro define la varianza explicada por el modelo en predicción:

$$R_{pred}^2 \equiv Q^2 = 1 - \frac{PRESS}{TSS} \quad (3.15)$$

El parámetro de la Ec. 3.15 tiene un máximo para la complejidad óptima del modelo y disminuye cada vez que se adicionan variables no predictivas o se eliminan variables explicativas al modelo. Durante la validación del modelo de regresión, al excluir una observación a la vez (dejar-uno-fuera), el parámetro se define como  $R_{1oo}^2$ ; mientras que al excluir un cierto porcentaje de observaciones, el parámetro de toma el nombre  $R_{lmo}^2$ .

Otro parámetro importante en MLR es la desviación estándar, definida en el caso multivariable como:

$$S = \sqrt{\frac{\sum_{i=1}^n \left( y_i - \hat{y}_i \right)^2}{n - p}} \quad (3.16)$$

donde  $p$  es el número de parámetros del modelo, es decir,  $p = d + 1$ ; y la diferencia  $n-p$  corresponde a los grados de libertad del modelo de regresión.

Finalmente, durante el desarrollo de esta Tesis, se evaluaron otros criterios de validación propuestos por Golbraikh y Tropsha (Golbraikh & Tropsha, 2002), los cuales ayudan a evitar la selección de modelos sobreajustados y erróneamente predictivos.

$$k = \frac{\sum_{i=1}^{N_{pred}} (y_i \times \hat{y}_i)^2}{\sum_{i=1}^{N_{pred}} (\hat{y}_i)^2}$$

$$y_{0i} = k \times \hat{y}_i$$

$$k' = \frac{\sum_{i=1}^{N_{pred}} (y_i \times \hat{y}_i)^2}{\sum_{i=1}^{N_{pred}} (y_i)^2}$$

$$\hat{y}_{0i} = k' \times y_i$$

$$R_0^2 = 1 - \frac{\sum_{i=1}^{N_{pred}} (y_i - y_{0i})^2}{\sum_{i=1}^{N_{pred}} (y_i - \bar{y})^2}$$

$$R_0'^2 = 1 - \frac{\sum_{i=1}^{N_{pred}} (\hat{y}_i - \hat{y}_{0i})^2}{\sum_{i=1}^{N_{pred}} (\hat{y}_i - \hat{y}_{prom})^2}$$

$$R_m^2 = R^2 \times \left(1 - \sqrt{|R^2 - R_0^2|}\right)$$

donde  $y_{0i}$  son los valores de la respuesta en regresión a través del origen entre las respuestas experimentales ( $y_i$ ) y las respuestas predichas por el modelo ( $\hat{y}_i$ );  $\hat{y}_{0i}$  son los valores de la respuesta en regresión a través del origen entre las respuestas predichas ( $\hat{y}_i$ ) y las respuestas experimentales ( $y_i$ );  $\hat{y}_{prom}$  es el valor promedio de las respuestas predichas del grupo de predicción;  $R_0^2$  es el coeficiente de correlación de regresión entre  $y_i$  e  $\hat{y}_i$  a través del origen;  $R_0'^2$  es el coeficiente de regresión entre  $\hat{y}_i$  e  $y_i$  a través del origen;  $R^2$  es el coeficiente de determinación entre los valores observados ( $y_i$ ) y predichos ( $\hat{y}_i$ ) para el grupo de predicción con ordenada al origen b.

### 3.7.2. Métodos de clasificación

#### **Análisis discriminante**

El análisis discriminante (DA) (McLachlan, 2004) es uno de los métodos de clasificación más usados. DA es una técnica de clasificación paramétrica probabilística, la cual maximiza la varianza entre las clases y minimiza la varianza dentro de las clases, mediante una proyección de los datos en un espacio de pocas dimensiones. De esta manera, se obtiene un número de funciones discriminantes lineales y ortogonales, iguales al número de las clases menos uno.

Cuando se trabaja con DA, se tiene la posibilidad de usar el análisis discriminante lineal (LDA) y el análisis discriminante cuadrático (QDA). Ambos casos se basan en el teorema de Bayes, por lo que requieren cumplir el supuesto de multinormalidad. La elección entre LDA y QDA dependerá de la separación lineal o no lineal de las clases, y de la confiabilidad de las matrices de covarianza de las clases. Para el LDA se

calcula únicamente la matriz de covarianza global; mientras que para el QDA la matriz de covarianza se calcula separadamente para cada clase. El cálculo de la matriz de covarianza de cada clase en QDA se realiza únicamente si el número de objetos de cada clase es mayor que el número de variables. Por otro lado, LDA considera únicamente que el número de muestras totales sea mayor que el número de variables.

### **Análisis discriminante de mínimos cuadrados parciales**

El análisis discriminante de mínimos cuadrados parciales (PLSDA) (Wold *et al.*, 2001; Barker & Rayens, 2003; Ballabio & Consonni, 2013) es un método de clasificación ampliamente usado, que combina las propiedades de los mínimos cuadrados parciales en regresión (método basado en el PLS2) con la capacidad discriminante lineal en clasificación. PLSDA busca relaciones entre la matriz de variables y el vector de clases mediante el cálculo de variables latentes, las cuales son combinaciones lineales ortogonales de las variables originales. Cuando se trabaja con PLSDA, los datos deben ser autoescalados.

### ***k*NN**

El método de los *k*-vecinos más cercanos (*k*NN) (Cover & Hart, 1967) es un método de clasificación no lineal y no paramétrico, es decir, que no considera el conocimiento a priori de las distribuciones estadísticas que siguen las variables. *k*NN clasifica en función de analogías, siguiendo una estrategia conceptualmente simple: un objeto es clasificado según las clases a las que pertenecen la mayoría de los *k* objetos más cercanos en el espacio de los datos. Computacionalmente *k*NN calcula y analiza la matriz de distancias entre los objetos, para la cual normalmente se usa la distancia Euclidiana se selecciona un número entero de *k* vecinos cercanos al objeto a clasificar (los valores normales que toma *k* es de 1 a 10). Entonces, los objetos son ordenados según sus distancias y clasificado según la clase mayoritaria a la cual pertenecen la mayoría de los *k* vecinos. El algoritmo en que se basa el algoritmo *k*NN es el siguiente:

1. Escalado de los datos.
2. Selección de la distancia a usar.
3. Optimización del número de entornos *k*.
4. Cálculo de la matriz de distancias.
5. Cada objeto se clasifica según la clase más representativa de los *k* vecinos más cercanos.

El valor óptimo de *k* se puede obtener mediante validación cruzada, es decir, se evaluarán diversos valores de *k* y se elegirá el valor que genere el menor error en clasificación (*NER*). Este método no proporciona un modelo matemático (función analítica) para efectos de predicción. El modelo como tal está constituido por los objetos del grupo de

entrenamiento, el valor  $k$  y la matriz de distancias. De esta manera, para predecir la clase de un nuevo objeto, éste debe ser introducido en la matriz de datos y se corre nuevamente el algoritmo para evaluar la clase mayoritaria de los  $k$  vecinos del grupo de entrenamiento.

Al estar  $k$ NN basado en el análisis de analogías entre muestras, es eficaz y proporciona buenos resultados cuando las superficies de separación de las clases no es lineal o son particularmente complejas (una clase contenida en otra). Es por estas razones y por su característica no paramétrica que  $k$ NN ha sido sugerido como un método estándar de comparación cuando se desarrollan y proponen nuevos métodos de clasificación.

### N3

El método de los  $N$ -vecinos más cercanos (N3) (Todeschini *et al.*, 2015) es una técnica de clasificación derivada del método  $k$ NN. A diferencia de  $k$ NN donde se busca el valor óptimo de  $k$ , N3 usa todos los  $n-1$  vecinos (objetos) para clasificar el  $i$ -ésimo objeto. Los  $n-1$  objetos son ordenados desde el más similar hasta el menos similar (ranking), para obtener el correspondiente vector de similitud  $r$ , el cual mide la contribución de los vecinos a la asignación de las clases, cuyo rol es modulado por un parámetro  $\alpha$ . Para cada  $i$ -ésimo objeto a ser clasificado, la contribución de la  $g$ -ésima clase se calcula de la siguiente manera:

$$w_{ig} = \frac{1}{n_g} \times \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{s_{ij}}{r_{ij}^\alpha} \times \delta_j \quad (3.17)$$

donde  $s_{ij}$  es la similitud entre la  $i$ -ésima y la  $j$ -ésima observación;  $r_{ij}$  es el valor de ranking de la similitud del  $j$ -ésimo objeto con respecto al  $i$ -ésimo objeto;  $\alpha$  es un parámetro de valor real a ser optimizado dentro del intervalo  $[0.1, 2.5]$ ;  $\delta_{ij}$  es la delta de Dirac que es igual a 1 cuando el  $j$ -ésimo objeto pertenece a la  $g$ -ésima clase y su contribución al peso de la clase es mayor al valor  $\varepsilon$ :

$$\delta_j = \begin{cases} 1 & \text{si } c_j = g \wedge \frac{s_{ij}}{r_{ij}^\alpha} > \varepsilon \\ 0 & \text{caso contrario} \end{cases}$$

donde  $c_j$  es la clase del  $j$ -ésimo objeto. Finalmente,  $\hat{n}_g$  es el número de vecinos que contribuyen al peso de la clase:

$$\hat{n}_g = \sum_j \delta_j$$

El exponente  $\alpha$  de la Ec. 3.17 se optimiza de tal forma que genere la mayor NER en validación cruzada.

**Parámetros de evaluación de los modelos de clasificación**

Para el cálculo de los parámetros de evaluación de un modelo de clasificación se parte de la denominada “*matriz de confusión*” (Ballabio & Consonni, 2013). Esta matriz se construye a partir de las clases verdaderas y las clases predichas por el modelo. Es una matriz simétrica de dimensión  $G \times G$ , donde  $G$  es el número de clases. Cada elemento de la matriz, denotado con  $n_{gk}$ , representa el número de muestras que pertenecen a la clase  $g$  y que son asignados a la clase  $k$ . La última columna representa el número de objetos que no son clasificados por el modelo.

**Tabla 3.1.** Matriz de confusión de un modelo de clasificación

|                   |     | Clases asignadas |          |     |          |
|-------------------|-----|------------------|----------|-----|----------|
|                   |     | 1                | 2        | ... | G        |
| Clases verdaderas | 1   | $n_{11}$         | $n_{12}$ | ... | $n_{1G}$ |
|                   | 2   | $n_{21}$         | $n_{22}$ | ... | $n_{2G}$ |
|                   | ... | ...              | ...      | ... | ...      |
|                   | G   | $n_{G1}$         | $n_{G2}$ |     | $n_{GG}$ |

Los elementos diagonales representan el número de elementos clasificados correctamente, mientras que los elementos fuera de la diagonal indican el número de errores de clasificación (elementos incorrectamente clasificados). Los parámetros que se derivan de la matriz de confusión son los siguientes (Varmuza & Filzmoser, 2009; Ballabio & Consonni, 2013):

**Precisión:** indica la capacidad del modelo de no incluir objetos de otras clases en la clase considerada:

$$Pr_g = \frac{n_{gg}}{n'_g} \tag{3.18}$$

donde  $n_{gg}$  es el número de elementos de la  $g$ -ésima clase correctamente clasificados y  $n'_g$  es número total de muestras asignadas a la  $g$ -ésima clase.

**Sensibilidad:** describe la capacidad del modelo de reconocer correctamente elementos que pertenecen a la  $g$ -ésima clase:

$$Sn_g = \frac{n_{gg}}{n_g} \quad (3.19)$$

donde  $n_g$  es el número total de muestras perteneciente a la  $g$ -ésima clase.

**Especificidad:** representa la capacidad de la  $g$ -ésima clase del modelo para rechazar muestras de todas las demás clases:

$$Sp_g = \frac{\sum_{k=1}^G (n'_k - n_{gk})}{n - n_g} \quad \text{para } k \neq g \quad (3.20)$$

donde  $n$  es el número total de muestras y  $n'_k$  es el número de muestras asignadas a la  $k$ -ésima clase, el cual se calcula de la siguiente manera:

$$n'_k = \sum_{g=1}^G n_{gk}$$

Cuando se tiene un problema que involucre únicamente dos clases, la sensibilidad de la clase 1 corresponde a la especificidad de la clase 2 y viceversa.

**Exactitud:** es la proporción de muestras correctamente clasificadas:

$$AC = \frac{\sum_{g=1}^G n_{gg}}{n} \quad (3.21)$$

**Tasa de aciertos:** se calcula como el promedio de las sensibilidades de las clases. Este parámetro estima mejor la calidad de los modelos clasificación con respecto a la Exactitud, particularmente cuando las clases son desbalanceadas.

$$NER = \frac{\sum_{g=1}^G Sn_g}{G} \quad (3.22)$$

### 3.7.3. Análisis de consenso

Debido a que es posible obtener modelos por diversas vías para una misma actividad/propiedad, se puede recurrir al análisis de consenso entre los mismos (van Rhee, 2003; Todeschini *et al.*, 2009); es decir, se basa en la combinación de las predicciones realizadas por dos o más modelos QSAR/QSPR. Esta metodología parte del fundamento de que cada modelo tiene implícitamente algunos defectos que determinan sus errores de predicción. Por tal motivo, el objetivo del análisis de consenso es mejorar las predicciones de tales modelos, donde la debilidad de un modelo es contrabalanceado por las fortalezas del otro (u otros) y vice versa. Esta estrategia se puede utilizar entre modelos de regresión o entre modelos de clasificación. Comúnmente se usan dos tipos de consenso: 'estricto' y 'no estricto'.

En *regresión*, el consenso estricto considera únicamente las predicciones de las moléculas que se encuentran dentro del dominio de aplicabilidad (AD) de todos los modelos, proveyendo de esta forma una predicción media de los modelos individuales. Contrariamente, el consenso no estricto considera únicamente las predicciones para las moléculas que caen dentro del AD de al menos un modelo; es decir, si un compuesto se encuentra dentro del AD de solamente un modelo, solo dicha predicción será considerada.

En *clasificación*, el consenso estricto considera únicamente las predicciones correctas de todos los modelos de clasificación; mientras que cuando las predicciones son dispares, el consenso no estricto usa únicamente el modelo para el cual la predicción es confiable. En algunos casos se puede directamente etiquetar tal compuesto como una molécula no asignada (clase 0). Evidentemente, el consenso no estricto permite obtener predicciones para un porcentaje más alto de moléculas.

## 3.8 Técnicas de validación

La validación es la fase del desarrollo del modelo y del control de su capacidad predictiva. Para evaluar y mejorar la predictividad, se somete al modelo a pequeñas perturbaciones. En otras palabras, es la búsqueda de la estructura del modelo que maximiza la capacidad predictiva. Al mismo tiempo debe tener estabilidad independientemente de los datos utilizados para su desarrollo. En efecto, al aumentar la complejidad del modelo, aumenta la calidad descriptiva del mismo; sin embargo, un incremento descontrolado de la complejidad desemboca en una ineficaz predicción. Por tal motivo, se buscan aquellas técnicas que permiten evaluar la presencia de sobreajuste, debida al cambio de correlación (*change correlation*) o al método utilizado (Hawkins, 2004).

### 3.8.1. Validación cruzada o interna

#### ***Dejar–uno–fuera***

Dejar–uno–fuera (LOO) es una técnica de validación cruzada en la que se calculan  $n$  modelos, en cada uno de los cuales se excluye un elemento a la vez. Así, los modelos construidos con los  $n-1$  elementos se usan para predecir la respuesta del objeto excluido (Arlot & Celisse, 2010). Este método de validación cruzada aporta una mínima perturbación al modelo, debido a que se eliminan de a uno, y en forma secuencial, cada objeto del grupo de entrenamiento. Adicionalmente, cuando el número de objetos presentes aumenta, el parámetro de calidad en validación cruzada tiende al valor del parámetro en regresión, indicando por tanto que la capacidad predictiva puede ser optimista

El diagnóstico de la calidad del modelo se realiza evaluando la diferencia entre la respuesta experimental y la predicha luego de la exclusión para todos los objetos del conjunto del grupo de entrenamiento. El método LOO es el único que permite una comparación inequívoca entre diversos modelos. Esto es particularmente útil cuando para una misma actividad/propiedad se tienen modelos que provienen de diversos tipos de descriptores.

#### ***Dejar–varios–fuera***

La técnica de dejar–varios–fuera (LMO) es una generalización del método LOO, el cual introduce una perturbación mayor en el subgrupo de entrenamiento para tener una mejor estimación de la capacidad predictiva de un modelo. En LMO un porcentaje de objetos del grupo de entrenamiento se seleccionan de forma casual para conformar el subgrupo de validación. Cada subgrupo se excluye del grupo de entrenamiento y el modelo recalculado con los restantes se usa para predecir la respuesta de los objetos pertenecientes al subgrupo excluido (Burden *et al.*, 1997; Baumann & Stiefl, 2004). En LMO el grupo de entrenamiento se divide en  $G$  subgrupos de exclusión de forma que a cada subgrupo se asignan  $k$  objetos (es decir,  $n/G$ ).

#### ***k–grupos de validación cruzada***

Debido a que LMO es una técnica que se basa en una selección casual de las muestras para ser predichas y que dicha selección se repite tantas veces como sea posible, produce diferentes resultados cada vez que se aplica. Por esta razón, se ha propuesto la técnica LMO llamada  $k$ –grupos de validación cruzada ( $k$ –Fold Cross Validation), la cual permite realizar la partición del grupo de entrenamiento siguiendo una lógica (Hastie *et al.*, 2011). Esta metodología consiste en dividir el grupo de entrenamiento en  $k$  grupos de validación. Cada grupo  $k$  de validación se excluye una sola vez. La Tabla 3.2 muestra los valores de  $k$  generalmente usados y el porcentaje de muestras respectivo que se

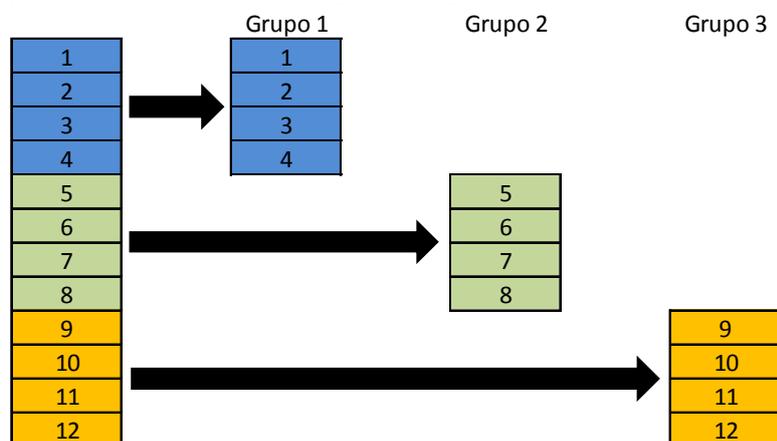
incluyen en el subgrupo de validación. Cuando el número de grupos  $k$  es igual a  $n$ , se está trabajando en la modalidad LOO.

**Tabla 3.2.** Valores de  $k$  típicos y porcentaje de muestras que son colocadas en el subgrupo de validación en la técnica de  $k$ -grupos de validación cruzada

| $k$ | % de muestras |
|-----|---------------|
| 2   | 50            |
| 3   | 33.3          |
| 4   | 25            |
| 5   | 20            |
| 10  | 10            |
| $n$ | 1             |

Las muestras se dividen en los  $k$  grupos de validación según dos procedimientos: ventanas venecianas (venetian blinds) o bloques continuos (contiguous blocks) (Ballabio & Consonni, 2013), en el cual cada  $k$  grupo contiene  $n/k$  elementos.

En bloques continuos, cada subgrupo  $k$  de validación se selecciona a partir de los primeros  $n/k$  elementos del grupo de entrenamiento ordenados en bloques continuos. En ventanas venecianas, cada objeto del grupo de validación es seleccionado a partir del primer objeto en el grupo de entrenamiento y los siguientes cada  $k$ -ésimo objeto. Las Figura 3.2 y 3.3 esquematizan la forma de dividir un conjunto de entrenamiento de 12 elementos en tres grupos de validación ( $k = 3$ ), para el método de bloques continuos y ventanas venecianas, respectivamente.

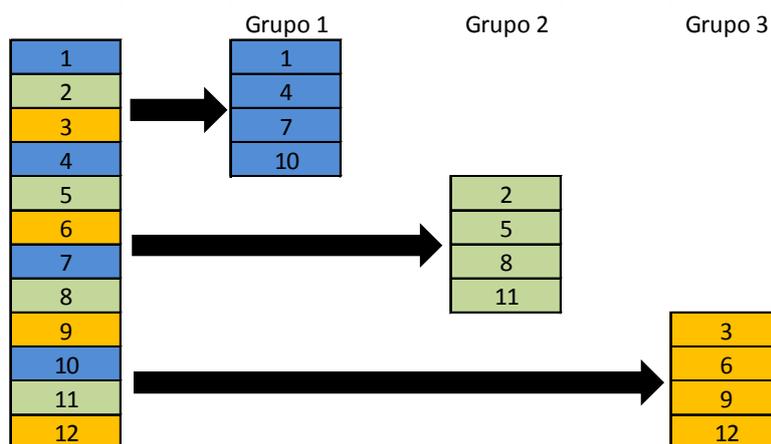


**Figura 3.2.** Ejemplo de una partición en bloques continuos usando 12 observaciones y 3 grupos de validación cruzada

Para garantizar un buen porcentaje de elementos en el subgrupo de entrenamiento, se recomienda usar un valor alto de  $k$  cuando se trabaja con bases de datos con pocos elementos y un valor bajo de  $k$  para grandes bases de datos. En el primer caso se evita introducir una perturbación grande durante la validación; mientras que en el segundo

caso se evita obtener una sobreestimación de la capacidad predictiva del modelo.

Cuando se trabaja en clasificación, la elección entre estos dos tipos de validación cruzada depende de la forma en que las clases están distribuidas en el vector respuesta. Bloques continuos es útil cuando las clases se encuentran distribuidas aleatoriamente. Por el contrario, cuando las clases siguen un orden lógico, es conveniente trabajar con ventanas venecianas.



**Figura 3.3.** Ejemplo de una partición en ventanas venecianas usando 12 observaciones y 3 grupos de validación cruzada

### **Monte Carlo**

El método Monte Carlo es un método de validación LMO bastante versátil, en el cual se definen en varias oportunidades distintos subgrupos de forma aleatoria. Cada una de estas etapas se denomina interacción. En cada interacción los elementos son divididos aleatoriamente en un subgrupo de entrenamiento (por ejemplo el 80%) y un grupo de evaluación (20%). En cada interacción se calibra el modelo con las moléculas del subgrupo de entrenamiento, el cual es luego utilizado para predecir las respuestas de las moléculas de evaluación. La calidad de la validación Monte Carlo se efectúa por comparación de las predicciones acumuladas versus las clases del subgrupo de evaluación (Xu *et al.*, 2004; Krakowska *et al.*, 2016). Dado que la partición se realiza de forma independiente para cada interacción, los objetos aparecerán varias veces en el subgrupo de evaluación.

### **Aleatorización-Y**

La aleatorización-Y (Lindgren *et al.*, 1996; Eriksson *et al.*, 1997) es una metodología que busca medir la presencia de correlación casual (*change correlation*) en el modelo, es decir, si la variable respuesta se encuentra correlacionada casualmente con las variables independientes. En esta metodología se modifica de forma casual los elementos del vector respuesta, de tal forma que no correspondan a las asignaciones

originales. Si el modelo calculado no presenta correlación casual, existirá una diferencia significativa entre la calidad del modelo original y el modelo obtenido con la aleatorización–Y. El proceso de aleatorización se repite cientos de veces a fin de obtener una estimación promedio del modelo aleatorizado.

### 3.8.2. Validación externa

Para la validación externa del modelo, el conjunto de datos se divide en dos subgrupos:

1. Grupo de entrenamiento (training set): se utiliza para construir el modelo que será utilizado posteriormente para la predicción de los objetos que forman parte del grupo de validación.
2. Grupo de validación (validation o test set): se usa para predecir la respuesta de sus elementos con el modelo desarrollado con el grupo de entrenamiento, es decir, no se utiliza durante la calibración del modelo.

Normalmente, en el subgrupo de validación se coloca entre el 10% y 50% de los elementos. Este procedimiento se puede realizar una sola vez (*Single Evaluation Set*) o varias veces (*Repeated Evaluation Set*). La partición de varias veces permite obtener un valor más estable del poder predictivo.

#### **Método de subconjuntos balanceados**

El método de subconjuntos balanceados (BSM) (Rojas *et al.*, 2015a, 2015b) es un método desarrollado en la presente tesis. Este enfoque permite dividir una base de datos de tal forma que se alcancen relaciones estructura–propiedad similares en los subconjuntos de calibración, de validación y/o de predicción. De esta forma, las moléculas de calibración son representativas de las moléculas presentes en los grupos de validación y predicción (Martin *et al.*, 2012). BSM considera la propiedad experimental y únicamente los descriptores moleculares independientes de la conformación, para así evitar sesgos debidos a los valores de descriptores conformacionales. BSM crea *k*–conglomerados de compuestos en términos de mínima distancia Euclidiana entre ellos, de tal forma que compuestos similares se ubicarán en el mismo conglomerado. Los pasos involucrados en este método son:

1. Crear una matriz ( $\mathbf{C}_1$ ) que incluya la propiedad experimental y los descriptores moleculares independientes de la conformación. Así, la dimensión de  $\mathbf{C}_1$  es  $N \times d$ .

2. Eliminar los descriptores linealmente dependientes. De esta forma, se crea una matriz  $\mathbf{C}_2$  de dimensión  $N \times d'$ .
3. Estandarizar la matriz  $\mathbf{C}_2$  para centrar y escalar sus elementos de matriz.
4. Crear  $N_{cal}^0$  conglomerados a través del método  $k$ -medias, para lo cual se usa la matriz  $\mathbf{C}_2$ , la distancia Euclidiana y varias corridas para optimizar el algoritmo y obtener la mejor solución. Este proceso calcula  $N_{cal}^0$  ubicaciones de los centroides, cada uno de dimensión  $1 \times d'$ .  $N_{cal}^0$  se define como  $N_{cal} - N_{\min \max}$ , donde  $N_{cal}$  es el número de moléculas en el subgrupo de calibración y  $N_{\min \max}$  es el número de compuestos que poseen el valor mínimo y máximo de la propiedad experimental.
5. El grupo de calibración ( $N_{cal}$ ) se construye incluyendo un compuesto de cada conglomerado, aquel más cercano al centroide en dicho conglomerado. Aquí también se incluyen las  $N_{\min \max}$  moléculas.
6. Crear  $N_{val}$  conglomerados con las restantes  $N - N_{cal}$  moléculas mediante el algoritmo  $k$ -medias en las mismas condiciones descritas previamente. Aquí se calculan  $N_{val}$  ubicaciones de los centroides.
7. Construir el grupo de validación ( $N_{val}$ ) mediante la inclusión de un compuesto por conglomerado, es decir, aquella molécula que se encuentra más cerca al centroide de cada conglomerado.
8. Finalmente, el grupo de predicción ( $N_{pred}$ ) es constituido por las restantes  $N - N_{cal} - N_{val}$  moléculas.

BSM tiene la capacidad de realizar la partición de una base de datos en: 1) grupos de calibración y predicción o, 2) grupos de calibración, validación y predicción.

### 3.9 Dominio de Aplicabilidad

El dominio de aplicabilidad (AD) (Jaworska *et al.*, 2005) de un modelo QSAR se define como el espacio en el cual el grupo de calibración del modelo ha sido desarrollado y es aplicable con la finalidad de realizar la predicción de nuevos compuestos. Para definir este espacio se necesita conocer la información fisicoquímica, estructural o biológica de las moléculas. Debido a que el AD es definido por las moléculas de grupo

de entrenamiento (estructuralmente limitadas), las predicciones se limitarán a aquellas moléculas que son estructuralmente similares a los compuestos en el grupo de calibración que se usaron para construir el modelo (Dimitrov *et al.*, 2005).

Por otra parte, la definición del AD de un modelo QSAR es el tercer requisito solicitado en los principios OECD de modelado QSAR (OECD, 2007) presentados en el Capítulo 1. Por lo tanto, cada modelo QSAR debe indicar la confianza de sus predicciones. Normalmente el AD de un modelo QSAR se define *a posteriori* en base a los descriptores moleculares del grupo de calibración. Entre los diversos enfoques *a posteriori* se describirán los dos utilizados en el desarrollo de la presente tesis:

### 3.9.1. AD basado en el valor de influencia

Esta estrategia se basa en el cálculo del valor de influencia ( $h$ ) o leverage (Rencher & Schaalje, 2008), el cual mide la distancia de cada objeto del grupo de predicción con respecto al centro del modelo. La matriz de influencia (matriz de leverage o matriz hat) se calcula a partir de la matriz del modelo  $\mathbf{X}$ :

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (3.23)$$

$\mathbf{H}$  es una matriz cuyos elementos diagonales ( $h_{ii}$ ) son los valores de influencia de cada compuesto. El valor  $h_{ii}$  es la contribución del  $i$ -ésimo elemento en la estimación de su respuesta. Para el grupo de entrenamiento estos valores están acotados entre los valores de 0 y 1, mientras que los valores calculados para el grupo de predicción solo tienen cota inferior 0. Para el análisis del AD se define un valor de umbral superior ( $h^*$ ) definido como tres veces el valor promedio de los valores de influencia del grupo de calibración:

$$h^* = 3 \frac{p}{n} \quad (3.24)$$

donde  $p$  es el número de parámetros del modelo y  $n$  es el número de compuestos en el grupo de calibración. El valor de influencia para una molécula del grupo de predicción se calcula de la siguiente manera:

$$h_{ii} = x_i (\mathbf{X}^T \mathbf{X})^{-1} x_i^T \quad (3.25)$$

Si el valor de influencia de una observación es grande, entonces la observación determina su respuesta sin considerar demasiado el resto de las observaciones. Por el contrario, valores de influencia pequeños y

parecidos indican que las observaciones contribuyen equitativamente en la construcción de las respuestas. De esta forma, moléculas que poseen grandes valores de leverage son las que se encuentran distantes del centro del modelo y, por lo tanto, se las considera predicciones poco confiables o extrapolaciones del modelo.

### **3.9.2. AD basado en la similitud *k*NN**

En este enfoque se define el AD del modelo mediante la evaluación de la similitud entre los compuestos del grupo de calibración y de predicción (Sheridan *et al.*, 2004; Sahigara *et al.*, 2013). Se parte del cálculo de la distancia promedio de cada molécula del grupo de predicción con respecto a sus *k* vecinos más cercanos del grupo de calibración, y se compara dicha distancia promedio con un valor de umbral pre-definido. Si la distancia promedio de una molécula del grupo de predicción es menor al umbral, dicha molécula estará dentro del AD del modelo y su predicción será confiable debido a que existe suficiente similitud con los compuestos del grupo de entrenamiento. Caso contrario, si la distancia promedio de la molécula es mayor al umbral, caerá fuera del AD y su predicción será considerada una extrapolación del modelo. De esta forma, el dominio de aplicabilidad se define de forma implícita con la similitud. Este método basado en similitudes caracteriza mejor la distribución de las moléculas y es útil cuando se usa con métodos de modelado basado en similitudes, por ejemplo el método *k*NN.

### 3.10 Referencias

- Arlot, S., & Celisse, A. (2010). A Survey of Cross-Validation Procedures for Model Selection. *Statistics surveys*, 4, 40-79.
- Bagheri, S., Omidikia, N., & Kompany-Zareh, M. (2013). Unsupervised Selection of Informative Descriptors in QSAR Study of Anti-HIV Activities of HEPT Derivatives. *Chemometrics and Intelligent Laboratory Systems*, 128, 135-143.
- Ballabio, D., & Consonni, V. (2013). Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA. *Analytical Methods*, 5(16), 3790-3798.
- Ballabio, D., Consonni, V., Mauri, A., Claeys-Bruno, M., Sergent, M., & Todeschini, R. (2014). A Novel Variable Reduction Method Adapted from Space-Filling Designs. *Chemometrics and Intelligent Laboratory Systems*, 136, 147-154.
- Barker, M., & Rayens, W. (2003). Partial Least Squares for Discrimination. *Journal of chemometrics*, 17(3), 166-173.
- Baumann, K., & Stiefl, N. (2004). Validation Tools for Variable Subset Regression. *Journal of Computer-Aided Molecular Design*, 18(7), 549-562.
- Bro, R., & Smilde, A. K. (2014). Principal Component Analysis. *Analytical Methods*, 6(9), 2812-2831.
- Burden, F. R., Brereton, R. G., & Walsh, P. T. (1997). Cross-Validatory Selection of Test and Validation Sets in Multivariate Calibration and Neural Networks as Applied to Spectroscopy. *Analyst*, 122(10), 1015-1022.
- Consonni, V., Ballabio, D., Manganaro, A., Mauri, A., & Todeschini, R. (2009). Canonical Measure of Correlation (CMC) and Canonical Measure of Distance (CMD) between Sets of Data: Part 2. Variable Reduction. *Analytica Chimica Acta*, 648(1), 52-59.
- Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. (2005). A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. *Journal of chemical information and modeling*, 45(4), 839-849.
- Duchowicz, P. R., Castro, E. A., Fernández, F. M., & Gonzalez, M. P. (2005). A New Search Algorithm for QSPR/QSAR Theories: Normal Boiling Points of Some Organic Molecules. *Chemical Physics Letters*, 412(4), 376-380.
- Duchowicz, P. R., Castro, E. A., & Fernández, F. M. (2006). Alternative Algorithm for the Search of an Optimal Set of Descriptors in QSAR-QSPR Studies. *MATCH Communications in Mathematical and in Computer Chemistry*, 55, 179-192.

- Efroymson, M. (1960). Multiple Regression Analysis. In A. Ralston & H. S. Wilf (Eds.), *Mathematical methods for digital computers*, (pp. 191-203): Wiley.
- Eriksson, L., Johansson, E., & Wold, S. (1997). Quantitative Structure-Activity Relationship Model Validation. In F. Chen & G. Schüürmann (Eds.), *Quantitative Structure-Activity Relationships in Environmental Sciences-VII*, (pp. 381-397): SETAC Press.
- Golbraikh, A., & Tropsha, A. (2002). Beware of  $q^2$ ! *Journal of Molecular Graphics and Modelling*, 20(4), 269-276.
- Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Second ed.): Springer.
- Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.
- Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR Applicability Domain Estimation by Projection of the Training Set Descriptor Space: A Review. *ATLA*, 33(5), 445-459.
- Jolliffe, I. T. (1972). Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Applied statistics*, 21(2), 160-173.
- Jolliffe, I. T. (1973). Discarding Variables in a Principal Component Analysis. II: Real Data. *Applied statistics*, 22(1), 21-31.
- Jolliffe, I. T. (2002). *Principal Component Analysis* (Second ed.): Springer.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.
- Krakowska, B., Custers, D., Deconinck, E., & Daszykowski, M. (2016). The Monte Carlo Validation Framework for the Discriminant Partial Least Squares Model Extended with Variable Selection Methods Applied to Authenticity Studies of Viagra[registered sign] Based on Chromatographic Impurity Profiles. *Analyst*, 141(3), 1060-1070.
- Kruskal, J. B. (1964). Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, 29(1), 1-27.
- Leardi, R., Boggia, R., & Terrile, M. (1992). Genetic Algorithms as a Strategy for Feature Selection. *Journal of chemometrics*, 6(5), 267-281.
- Leardi, R., & Gonzalez, A. L. (1998). Genetic Algorithms Applied to Feature Selection in PLS Regression: How and When to Use Them. *Chemometrics and Intelligent Laboratory Systems*, 41(2), 195-207.
- Leardi, R. (2009). Genetic Algorithms. In R. Tauler, B. Walczak & S. D. Brown (Eds.), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, (pp. 631-653): Elsevier.
- Lindgren, F., Hansen, B., Karcher, W., Sjöström, M., & Eriksson, L. (1996). Model Validation by Permutation Tests: Applications to Variable Selection. *Journal of chemometrics*, 10(5-6), 521-532.

- Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., & Tropsha, A. (2012). Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *Journal of chemical information and modeling*, 52, 2570-2578.
- Massart, D. L., & Kaufman, L. (1983). *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*: Wiley.
- Massart, D. L., Vandeginste, B. G., Buydens, L., Lewi, P., & Smeyers-Verbeke, J. (1997). *Handbook of Chemometrics and Qualimetrics: Part A*: Elsevier Science Inc.
- McLachlan, G. (2004). *Discriminant Analysis and Statistical Pattern Recognition*: John Wiley & Sons.
- OECD. (2007). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. In: OECD Publishing.
- Questier, F., Put, R., Coomans, D., Walczak, B., & Vander Heyden, Y. (2005). The Use of CART and Multivariate Regression Trees for Supervised and Unsupervised Feature Selection. *Chemometrics and Intelligent Laboratory Systems*, 76(1), 45-54.
- Rencher, A. C., & Schaalje, G. B. (2008). *Linear Models in Statistics*: John Wiley & Sons.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015a). QSPR Analysis for the Retention Index of Flavors and Fragrances on a OV-101 Column. *Chemometrics and Intelligent Laboratory Systems*, 140, 126-132.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015b). Quantitative Structure-Property Relationship Analysis for the Retention Index of Fragrance-Like Compounds on a Polar Stationary Phase. *Journal of Chromatography A*, 1422, 277-288.
- Sahigara, F., Ballabio, D., Todeschini, R., & Consonni, V. (2013). Defining a Novel k-Nearest Neighbours Approach to Assess the Applicability Domain of a QSAR Model for Reliable Predictions. *Journal of cheminformatics*, 5(1), 27.
- Sheridan, R. P., Feuston, B. P., Maiorov, V. N., & Kearsley, S. K. (2004). Similarity to Molecules in the Training Set is a Good Discriminator for Prediction Accuracy in QSAR. *Journal of chemical information and computer sciences*, 44(6), 1912-1928.
- Todeschini, R., Consonni, V., & Gramatica, P. (2009). Chemometrics in QSAR. In R. Tauler, B. Walczak & S. D. Brown (Eds.), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, (pp. 129-170): Elsevier.
- Todeschini, R., Ballabio, D., Cassotti, M., & Consonni, V. (2015). N3 and BNN: Two New Similarity Based Classification Methods in Comparison with Other Classifiers. *Journal of chemical information and modeling*, 55(11), 2365-2374.
- van Rhee, A. M. (2003). Use of Recursion Forests in the Sequential Screening Process: Consensus Selection by Multiple Recursion

- Trees. *Journal of chemical information and computer sciences*, 43(3), 941-948.
- Vandeginste, B. G., Massart, D. L., Buydens, L., De Jon, S., Lewi, P., & Smeyers-Verbeke, J. (1998). *Handbook of Chemometrics and Qualimetrics: Part B*: Elsevier Science Inc.
- Varmuza, K., & Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics*: CRC press.
- Whitley, D. C., Ford, M. G., & Livingstone, D. J. (2000). Unsupervised Forward Selection: A Method for Eliminating Redundant Variables. *Journal of chemical information and computer sciences*, 40(5), 1160-1168.
- Winsberg, S., & Carroll, J. D. (1989). A Quasi-Nonmetric Method for Multidimensional Scaling Via an Extended Euclidean Model. *Psychometrika*, 54(2), 217-229.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52.
- Wold, S. (1995). Chemometrics; What do We Mean with It, and What do We Want from It? *Chemometrics and Intelligent Laboratory Systems*, 30(1), 109-115.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130.
- Wold, S. (2015). Chemometrics and Bruce: Some Fond Memories. In B. K. Lavine, S. D. Brown & K. S. Booksh (Eds.), *40 Years of Chemometrics – From Bruce Kowalski to the Future*, vol. 1199 (pp. 1-13): American Chemical Society.
- Xu, Q.-S., Liang, Y.-Z., & Du, Y.-P. (2004). Monte Carlo Cross-Validation for Selecting a Model and Estimating the Prediction Error in Multivariate Calibration. *Journal of chemometrics*, 18(2), 112-120.



# Capítulo 4

---

## APLICACIONES EN QUÍMICA ANALÍTICA

---

En este capítulo se presentarán las diversas aplicaciones desarrolladas en el campo de la Química Analítica, particularmente aquellas relacionadas con el estudio QSPR de los índices de retención medidos en cromatografía de gases (GC) y cromatografía de gases–espectrometría de masa (GC–MS) usando fases estacionarias de diversa polaridad.

Los aromas son sustancias orgánicas volátiles cuya característica principal es la de poseer olor fuerte placentero. Estos compuestos químicos son detectados por receptores (células olfativas) ubicados en la nariz. Por esta razón, son compuestos que tienen un rol importante en distintas industrias: perfumería, alimentos, medicina y tabaco (Surburg & Panten, 2016). Debido a que los aromas son compuestos químicos volátiles, su control de calidad basado en el análisis del perfil aromático se fundamenta en la cromatografía de gases o cromatografía de gases–espectrometría de masa. Este último permite alta calidad en la separación de picos y buena capacidad de identificación de los compuestos orgánicos volátiles (Cserháti, 2010).

El índice de retención cromatográfico ( $I$ ) es un parámetro útil para la identificación de los compuestos orgánicos volátiles, por lo que permite investigar los mecanismos por los cuales se desarrolla la retención en la fase móvil, particularmente mediante las relaciones cuantitativas estructura–propiedad. En el año 1977 se publican los primeros trabajos de las aplicaciones QSPR en índices de retención por parte de Kaliskan y Foks (Kaliszan, 1977; Kaliszan & Foks, 1977), al igual que Michotte y Massart (Michotte & Massart, 1977). El creciente interés en este campo de investigación ha permitido evitar procedimientos experimentales costosos y prolongados para la predicción de los índices de retención de nuevas moléculas, así como diseñar nuevas fases móviles y comprender los mecanismos de retención en las mismas (Héberger, 2007; Kaliszan, 2007). En este capítulo, al igual que en el siguiente, los modelos QSAR/QSPR han sido desarrollados siguiendo los cinco principios sugeridos por la Organización para la Cooperación Económica y el Desarrollo (OECD).

## 4.1 Modelo QSPR para índices de retención medidos en la columna apolar OV-101

### 4.1.1. Introducción

Varios estudios QSPR han sido publicados con la finalidad de predecir el  $I$  de compuestos volátiles en la columna apolar OV-101. En 1989, Staton y Jurs (Stanton & Jurs, 1989) usaron el programa ADAPT para estudiar una serie de 107 compuestos sustituidos de la pirazina tomados de la literatura, obteniendo buenos resultados en calibración ( $R_{cal}^2 = 0.99$ ,  $S_{cal} = 21.4$ ) y validación ( $R_{val}^2 = 0.99$  y  $S_{val} = 32.4$ ). Un año más tarde, Anker et al. (Anker *et al.*, 1990) midieron los índices de retención de 115 compuestos aromáticos (38 alcoholes, 11 aldehídos, 19 cetonas y 47 ésteres), para posteriormente calibrar un modelo QSPR con un coeficiente de determinación  $R_{cal}^2 = 0.99$ . Ese mismo año, Gerasimenko y Nabivach (Gerasimenko & Nabivach, 1990) usaron los  $I$  de alquilbencenos y alquilnaftalenos, obteniendo buenos resultados, tanto en calibración ( $R_{cal}^2 = 1.00$ ), como en validación ( $R_{val}^2 = 0.99$ ). En el año 2008, Goodner (Goodner, 2008) desarrolló un modelo de regresión, utilizando el punto de ebullición y el logaritmo del coeficiente de partición octanol-agua de 91 moléculas. Adicionalmente, este autor usó una combinación de varios test para detectar valores atípicos, tales como el test de Grubb y el Q-test de Dixon. En un estudio reciente, Yan et al. (Yan *et al.*, 2012) desarrollaron un modelo QSPR para los índices de retención de 297 compuestos, usando 4 descriptores moleculares: índices de conectividad molecular de orden cero ( ${}^0\chi$ ) y primer orden ( ${}^1\chi$ ), donante de enlaces de hidrógeno ( $ndonr$ ) y la masa molecular ( $MW$ ).

El objetivo del presente estudio fue el desarrollo de una relación cuantitativa estructura-propiedad para los índices de retención de 1208 compuestos volátiles medidos en la columna apolar OV-101 en cromatografía de gases. La base de datos usada constituye la más grande considerada hasta el momento, permitiendo una mayor generalización del modelo desarrollado a diversos tipos de compuestos. Los resultados de este trabajo han sido previamente publicados (Rojas *et al.*, 2015a).

### 4.1.2. Materiales y métodos

En este estudio se utilizaron 1208 compuestos orgánicos volátiles (sabores y aromas) reportados por Jennings y Shibamoto (Jennings & Shibamoto, 1980), para los cuales han sido reportados el índice de retención en la fase estacionaria metil silicona OV-101. Esta base de

datos no ha sido previamente usada en estudios QSPR y constituye la más extensa usada con dicho propósito. La propiedad experimental reportada por estos autores es el índice de retención en una columna capilar apolar (0.28 mm×50 m), recubierta con metil silicona OV-101 y programada desde 80 hasta 200°C a 2°C×min<sup>-1</sup>. El dominio experimental se encuentra en el rango de 350 a 2180 unidades. Para moléculas replicadas, se utilizó el índice de retención promedio. Los nombres químicos y los índices de retención experimentales se presentan en la Tabla 1A, que se encuentra en el CD que acompaña a esta tesis.

Los compuestos fueron representados en el programa HyperChem (Hypercube Inc.), donde fueron optimizados mediante los campos de fuerza de la mecánica molecular (MM+) y posteriormente refinadas sus geometrías mediante el método semiempírico PM3. En ambos casos, se usó el algoritmo de gradiente conjugado en la versión Polak-Ribiere y las geometrías se consideraron optimizadas cuando la desviación estándar del vector gradiente es menor a 0.01kcal×(Å×mol)<sup>-1</sup>. Posteriormente se calcularon 4885 descriptores moleculares usando el programa DRAGON versión 6 (Talete srl, 2015). Con la finalidad de analizar si el uso de descriptores conformacionales tiene relevancia significativa para modelar esta propiedad, se han construido tres conjuntos de descriptores: el primer conjunto considera todos los bloques de descriptores obtenidos con DRAGON, el segundo conjunto está formado por descriptores independientes de la conformación y el último engloba únicamente a los descriptores dependientes de la conformación (3D). En todos los casos, se han excluido descriptores con valores constantes o casi constantes y aquellos descriptores con al menos un valor faltante.

Un aspecto importante en los estudios QSPR, es la selección de los descriptores moleculares que permiten obtener los mejores modelos. Con esta finalidad, se utilizó el método de reemplazo para explorar los tres conjuntos de descriptores. RM es un algoritmo inequívoco (segundo principio del modelado QSPR) que se encuentra disponible en el lenguaje de programación MatLab (The MathWorks Inc).

Para validación del modelo QSPR, el conjunto de datos se dividió en conjuntos de calibración, validación y predicción. El grupo de calibración se usó para la selección de descriptores mediante RM y la calibración de los modelos; mientras que el grupo de validación se usó para la validación cruzada de los modelos con la finalidad de evitar la presencia de sobreajuste. El grupo de predicción se utilizó para medir la capacidad predictiva del modelo QSPR seleccionado. También se han usado las técnicas de validación cruzada de dejar-uno-fuera (loo) y dejar-varios-fuera (lmo) mediante la exclusión aleatoria del 20% de las moléculas, con 50000 repeticiones. La ausencia de correlación casual se estableció mediante la aleatorización-Y con 10000 repeticiones. Finalmente, se

evaluaron los criterios de validación propuestos por Golbraikh y Tropsha (Golbraikh & Tropsha, 2002).

Posteriormente, y según el tercer principio del modelado QSPR, se analizó el dominio de aplicabilidad siguiendo el enfoque del valor de influencia crítico ( $h^*$ ). Asimismo, se estableció la contribución de cada descriptor molecular para, de acuerdo al cuarto principio, brindar una interpretación del mecanismo de acción de los descriptores involucrados en el modelo.

### 4.1.3. Resultados y discusión

Como primer paso, se usó el BSM para dividir la base de datos de 1208 compuestos en los grupos de calibración ( $N_{cal} = 400$ ), validación ( $N_{val} = 405$ ) y predicción ( $N_{pred} = 403$ ). Las asignaciones a estos grupos se encuentran en la Tabla 1A. Seguidamente, el RM permitió explorar los tres grupos de descriptores: a) 2895 descriptores independientes y dependientes de la conformación, b) 1815 descriptores independientes de la conformación y c) 1080 descriptores dependientes de la conformación. El uso de estos tres grupos de descriptores se realizó con la finalidad de identificar si los descriptores dependientes de la conformación aportan información relevante al modelado del  $I$ .

Las Tablas 4.1, 4.2 y 4.3 resumen los mejores modelos de 1 a 7 descriptores moleculares para los tres bloques de descriptores. El mejor modelo se muestra en negrita. Se observa la ausencia de sobreajuste, debido a que para cualquier modelo de dimensión  $d$ , los parámetros de calibración y validación no presentan variaciones significativas.

**Tabla 4.1.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes y dependientes de la conformación.

| d        | $R_{cal}^2$ | $S_{cal}$   | $R_{val}^2$ | $S_{val}$   | $R_{ij\ max}^2$ | descriptores   |
|----------|-------------|-------------|-------------|-------------|-----------------|--|
| 1        | 0.87        | 124.4       | 0.89        | 107.1       | 0.00            | <i>X1sol</i>   |
| 2        | 0.89        | 112.5       | 0.92        | 91.9        | 0.00            | <i>Chi0_EA, GATS1p</i>                                     |
| 3        | 0.91        | 101.7       | 0.93        | 83.6        | 0.88            | <i>nHDon, RDF010e, Sp</i>                                  |
| 4        | 0.92        | 97.8        | 0.94        | 79.7        | 0.88            | <i>nHDon, DP02, RDF010e, Sp</i>                            |
| <b>5</b> | <b>0.92</b> | <b>95.4</b> | <b>0.94</b> | <b>78.3</b> | <b>0.32</b>     | <b><i>PDI, Hy, ATSC2s, EE_B(s), X1sol</i></b>              |
| 6        | 0.93        | 91.5        | 0.95        | 76.2        | 0.76            | <i>H-050, R1s+, Mor05p, RDF010s, ATSC2s, X1sol</i>         |
| 7        | 0.93        | 89.1        | 0.94        | 76.9        | 0.76            | <i>H-050, nCconj, R2s+, Mor05p, RDF010s, ATSC2s, X1sol</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\ max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

**Tabla 4.2.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d | $R_{cal}^2$ | $S_{cal}$ | $R_{val}^2$ | $S_{val}$ | $R_{ij\ max}^2$ | descriptores           |
|---|-------------|-----------|-------------|-----------|-----------------|------------------------|
| 1 | 0.87        | 124.4     | 0.89        | 107.1     | 0.00            | <i>X1sol</i>           |
| 2 | 0.89        | 112.5     | 0.92        | 91.9      | 0.00            | <i>Chi0_EA, GATS1p</i> |

|          |             |              |             |             |             |  |
|----------|-------------|--------------|-------------|-------------|-------------|--|
| 3        | 0.91        | 105.1        | 0.93        | 83.80       | 0.19        | <i>PDI, Hy, X1sol</i>                                |
| <b>4</b> | <b>0.91</b> | <b>100.9</b> | <b>0.93</b> | <b>83.0</b> | <b>0.17</b> | <b><i>PDI, H-050, SpMax1_Bh(s), X1sol</i></b>        |
| 5        | 0.92        | 96.4         | 0.94        | 82.7        | 0.38        | <i>PDI, O-058, H-050, ATSC4s, X1sol</i>              |
| 6        | 0.93        | 94.0         | 0.93        | 85.2        | 0.76        | <i>O-058, H-050, C-044, ATSC4e, H_Dt, X1sol</i>      |
| 7        | 0.93        | 91.1         | 0.94        | 80.1        | 0.32        | <i>PDI, Hy, C-044, C-033, ATSC2s, EE_B(s), X1sol</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

La selección del modelo óptimo se realizó considerando los parámetros de calidad del grupo de validación, pues son los indicadores de que no se sobreajusta un modelo. Debido a que RM es un método de selección de descriptores que se basa en la optimización mediante la minimización de la desviación estándar de calibración, se selecciona el modelo que brinde la menor desviación estándar tanto en calibración y validación. Asimismo, se busca que el coeficiente de correlación máximo entre descriptores ( $R^2_{ij\max}$ ) sea el más bajo posible, al igual que la dimensión del modelo siga el principio de parsimonia de Ockham (navaja de Ockham) (Hoffmann *et al.*, 1996), el cual sugiere que en igualdad de calidad de diversos modelos, se seleccione el más sencillo. Este criterio de selección de modelos se aplicará en todos los estudios en el que use RM como método de selección de descriptores.

**Tabla 4.3.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores dependientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$    | $R^2_{val}$ | $S_{val}$   | $R^2_{ij\max}$ | descriptores  |
|----------|-------------|--------------|-------------|-------------|----------------|---|
| 1        | 0.84        | 138.1        | 0.88        | 114.0       | 0.00           | <i>G2</i>   |
| 2        | 0.86        | 130.0        | 0.90        | 101.5       | 0.06           | <i>Mor11p, G2</i>                                       |
| 3        | 0.88        | 120.0        | 0.91        | 99.1        | 0.37           | <i>Mor03p, Mor18v, G2</i>                               |
| 4        | 0.89        | 112.3        | 0.91        | 96.2        | 0.57           | <i>DP08, TDB03p, SpMAD_RG, G2</i>                       |
| <b>5</b> | <b>0.91</b> | <b>103.1</b> | <b>0.94</b> | <b>88.6</b> | <b>0.81</b>    | <b><i>R1e+, H2v, RDF025p, RDF010p, Chi_RG</i></b>       |
| 6        | 0.91        | 102.0        | 0.93        | 87.4        | 0.81           | <i>R1m+, H2v, RDF010s, RDF025p, RDF010p, Chi_RG</i>     |
| 7        | 0.92        | 100.3        | 0.92        | 89.8        | 0.81           | <i>Q2, R2s+, L2m, RDF010s, RDF025p, RDF010p, Chi_RG</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

Se observa que los descriptores conformacionales no aportan información relevante para modelar el índice de retención en la columna OV-101. En efecto, los parámetros de calidad entre los modelos obtenidos con todos los descriptores (Tabla 4.1) y únicamente los descriptores independientes de la conformación (Tabla 4.2) son muy similares. Adicionalmente, los modelos que consideran únicamente los descriptores conformacionales (Tabla 4.3) no muestran una mejora significativa en la predicción del *I*. Este es un resultado importante

debido a que un modelo que no incluye descriptores conformacionales involucra menos dificultades para su desarrollo. Por ejemplo, se evitan los costos computacionales elevados que se requieren para la optimización de las geometrías y también se evitan problemas asociados a ambigüedades que resultan de una incorrecta optimización, debido a que una molécula puede existir en diversos estados conformacionales, generando diversos valores para un mismo descriptor molecular 3D (Garcia *et al.*, 2016). Así, se ha elegido el modelo independiente de la conformación de 4 descriptores para un análisis más detallado:

$$I = -998.1 + 169.8X1sol + 24.4SpMax1\_ Bh(s) + 124.9H - 050 + 1254.9PDI \quad (4.1)$$

$$N_{cal} = 400, R_{cal}^2 = 0.91, S_{cal} = 100.9$$

$$N_{val} = 405, R_{val}^2 = 0.94, S_{val} = 83.0$$

$$N_{pred} = 403, R_{pred}^2 = 0.93, S_{pred} = 78.0$$

$$R_{loo}^2 = 0.92, S_{loo} = 92.5, R_{lmo}^2 = 0.96, S_{lmo} = 94.0$$

$$S_{rand} = 326.2, \alpha(3S) = 10, R_{ij\ max}^2 = 0.17$$

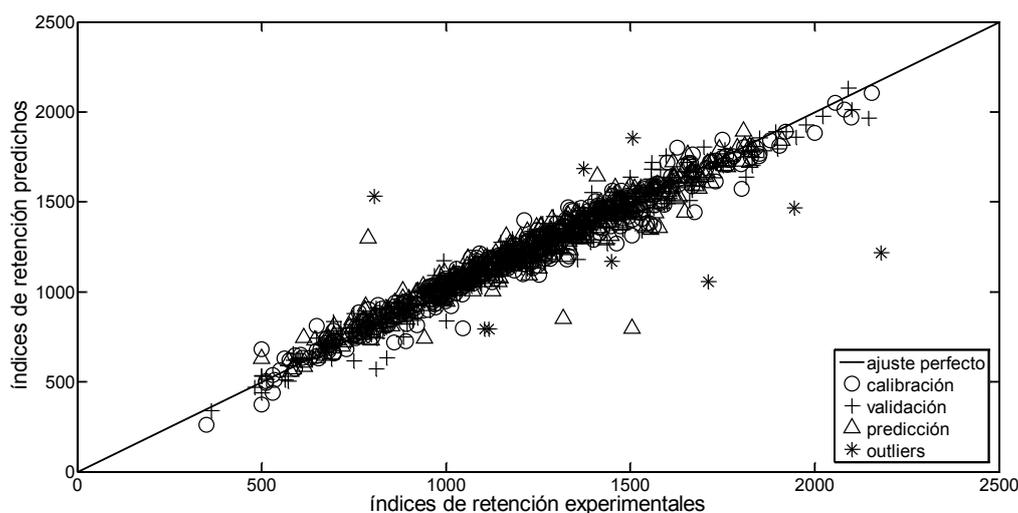
La validación cruzada de dejar–uno–fuera (loo) y dejar–varios–fuera (lmo), mediante la exclusión de una muestra a la vez y del 20% de moléculas, respectivamente, indica que el modelo es estable. Asimismo, la aleatorización–Y muestra que el  $S_{cal} < S_{rand}$ , indicando la ausencia de correlación casual en el modelo de la Ec. 4.1. Finalmente, los parámetros adicionales de validación que se presentan a continuación, indican que una relación cuantitativa estructura–propiedad predictiva se ha alcanzado:

$$R_{loo}^2 > 0.5 \text{ (0.92) y } R_{pred}^2 > 0.6 \text{ (0.93)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.000) o } 1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.009)}$$

$$0.85 < k(0.994) \leq 1.15 \text{ y } 0.85 < k'(1.002) \leq 1.15$$

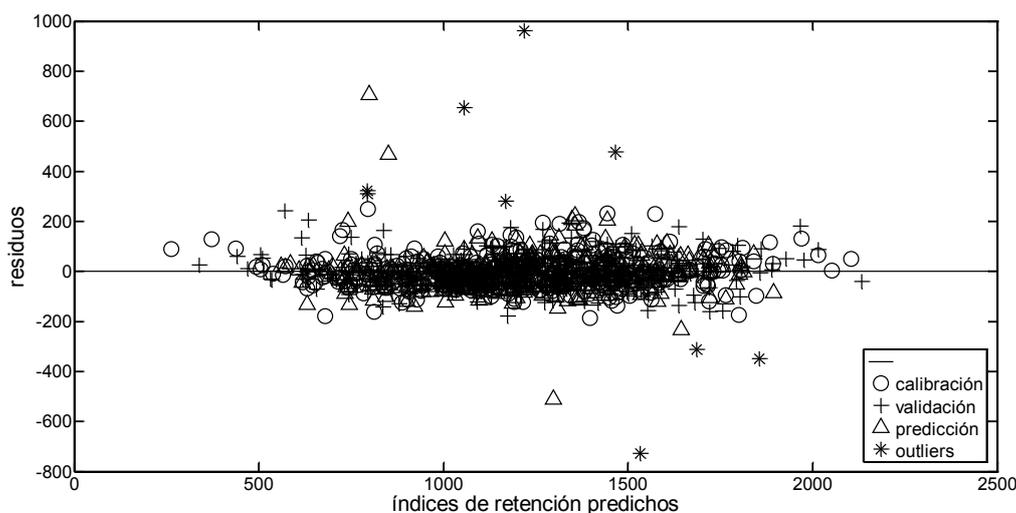
$$R_m^2 > 0.5 \text{ (0.91)}$$



**Figura 4.1.** Índices de retención experimentales y predichos para aromas medidos en la columna OV-101

La Figura 4.1 muestra la gráfica de los valores de índices de retención experimentales en función de los predichos por el modelo. Se observa la tendencia lineal alrededor de la recta de ajuste perfecto. Por otra parte, en la Figura 4.2 se muestra el gráfico de los valores de los residuos, el cual revela que los mismos tienden a seguir un patrón aleatorio alrededor de la línea cero, sugiriendo que los índices de retención se modelan mediante una regresión lineal múltiple.

Los diez valores atípicos identificados en el modelo QSPR son los compuestos: *Almizcle xilol* (523), *Elemol* (740), *2-Buteno-1,4-diol* (854), *2-hidroxiisobutirato de metilo* (861), *n-Pentadecanol* (961), *2-Decenal* (985), *Isoamil piruvato* (1044), *n-hexanoato de n-butilo* (1057), *n-octanoato de etilo* (1058), *2-Mercaptobenzotiazol* (1202). Posterior a un control en la fuente para descartar errores de transcripción en los índices de retención, así como la verificación de las estructuras químicas, se puede asumir que este comportamiento irregular se puede deber a la amplia heterogeneidad de compuestos considerados en la presente base de datos.



**Figura 4.2.** Gráfico de dispersión para el modelo QSPR de la Ec. 4.1

La máxima correlación entre los descriptores  $X1sol$  y  $PDI$  ( $R_{ij\max}^2 = 0.17$ ) refleja una baja correlación entre los descriptores del modelo QSPR y ausencia de colinealidad. Así, cada descriptor molecular explica diferentes aspectos de la estructura molecular y su relación con la propiedad en estudio (Duchowicz *et al.*, 2010). El grado de contribución relativo de cada descriptor ( $b_j^s$ ), indica la importancia de cada descriptor en modelar el  $l$ :  $X1sol$  (0.87) >  $PDI$  (0.19) >  $H-050$  (0.16) >  $SpMax1\_Bh(s)$  (0.08). Todos los descriptores tienen un efecto sinérgico sobre el índice de retención, lo que indica que al incrementar el valor de dichos descriptores, incrementa el valor de dicha propiedad.

El índice de conectividad de solvatación de primer orden ( $X1sol$ ) (Zefirov & Palyulin, 2001) ha sido propuesto para modelar la entropía de solvatación y para describir las interacciones de dispersión en solución. Este índice de conectividad se encuentra correlacionado con los puntos de ebullición de las sustancias. En efecto, el punto de ebullición es la propiedad que gobierna la retención en fases apolares en cromatografía de gases. Buydens y Massart (Buydens & Massart, 1981) también han demostrado que los índices topológicos (TIs) y parámetros estructurales son útiles para describir interacciones entre compuestos volátiles de una misma familia con la fase estacionaria.

El índice de empaquetamiento ( $PDI$ ) (Ciubotariu *et al.*, 2004) es una propiedad molecular definida como el cociente entre el volumen de McGowan y la superficie molecular de van der Waals total, representando características estéricas de una molécula. Por otra parte, el número de hidrógenos enlazados a un heteroátomo ( $H-050$ ) (Ghose *et al.*, 1998) mide la presencia de heteroátomos en un compuesto. Finalmente, el autovalor más alto de la matriz de Burden ponderada por el estado intrínseco ( $SpMax1\_Bh(s)$ ) (Gutnman *et al.*, 2003; Consonni &

Todeschini, 2008) se calcula a partir de un grafo molecular completo ponderado con el estado intrínseco atómico.

**Tabla 4.4.** Comparación de varios modelos QSPR para los índices de retención medidos en la columna OV-101

| modelo                         | compuestos                              | N    | d | $R^2_{cal}$ | $S_{cal}$ | $R^2_{pred}$   | $S_{pred}$ |
|--------------------------------|---|------|---|-------------|-----------|----------------|------------|
| (Stanton & Jurs, 1989)         | Pirazinas sustituidas                   | 107  | 6 | 0.99        | 21.4      | 0.99           | 32.4       |
| (Anker <i>et al.</i> , 1990)   | Alcoholes, aldehidos, cetonas y ésteres | 115  | 7 | 1.00        | 11.1      | — <sup>a</sup> | —          |
| (Gerasimenko & Nabivach, 1990) | Alquilbencenos                          | 39   | 7 | 1.00        | 3.4       | 1.00           | —          |
|                                | Alquilnaftalenos                        | 15   | 5 | 1.00        | 1.7       | 1.00           | —          |
|                                | Alquil aril carbamatos (Grupo 1)        | 27   | 3 | 1.00        | 1.4       | —              | —          |
|                                | Alquil aril carbamatos (Grupo 2)        | 27   | 4 | 1.00        | 0.6       | —              | —          |
|                                | Alquil aril carbamatos (Grupo 3)        | 27   | 4 | 0.99        | 7.0       | —              | —          |
| (Goodner, 2008)                | Fragancias                              | 91   | 5 | 0.99        | —         | —              | —          |
| (Yan <i>et al.</i> , 2012)     | Sabores                                 | 297  | 4 | 0.96        | 59.6      | 0.96           | 58.0       |
| Presente <sup>b</sup>          | Sabores y fragancias                    | 1208 | 4 | 0.91        | 100.9     | 0.93           | 78.0       |

<sup>a</sup> no disponible; <sup>b</sup> (Rojas *et al.*, 2015a)

La Tabla 4.4 presenta una comparación de los modelos QSPR desarrollados para índices de retención medidos en la columna apolar OV-101. Se observa que cuando se construye un modelo considerando únicamente un grupo de compuestos que tienen estructura similar, se obtienen buenos resultados, sea en calibración o predicción. Adicionalmente, tres modelos QSPR no presentan validación externa (Anker *et al.*, 1990; Gerasimenko & Nabivach, 1990; Goodner, 2008). Por el contrario, al considerar grupos heterogéneos de moléculas se obtienen modelos con calidad inferior, aunque es siempre posible dividir la base de datos para realizar validación externa (Stanton & Jurs, 1989; Yan *et al.*, 2012). La base de datos de 1208 compuestos es la más grande utilizada hasta el momento. En efecto, es 4.1 veces mayor a la más grande previamente utilizada (Yan *et al.*, 2012).

Debido a la amplia diversidad de moléculas consideradas en el modelo QSPR de la Ec. 4.1, las predicciones se restringen a compuestos volátiles para los cuales el  $l$  esté dentro del rango de 350 a 2180 y cuyo valor de influencia sea menor al valor de influencia crítico ( $h^* = 0.019$ ). El análisis del dominio de aplicabilidad (AD) indica que 5 compuestos del grupo de predicción caen fuera del AD del modelo. Estos compuestos son: *cis-3-hexenil antranilato* ( $h_{ij} = 0.026$ ), *Isoamil antranilato* ( $h_{ij} = 0.025$ ), *1,3-Butanodiol* ( $h_{ij} = 0.024$ ), *Alil antranilato* ( $h_{ij} = 0.023$ ), *o-Aminoacetofenona* ( $h_{ij} = 0.023$ ). Luego de verificar que los valores de  $l$  y las estructuras químicas para estos compuestos son correctos en la fuente, se asume que este comportamiento se debe a la

heterogeneidad de las moléculas consideradas en esta base de datos. El principio de AD en QSPR obliga a definir las limitaciones del modelo con respecto a los descriptores moleculares y el espacio definido por la propiedad. Es decir, ningún modelo robusto, validado y predictivo es capaz de producir predicciones confiables para todo tipo de moléculas (Dimitrov *et al.*, 2005; Gramatica, 2007), pues constituyen modelos locales. Por otro lado, se había indicado previamente (Tabla 4.4) que varios modelos QSPR se han desarrollado considerando únicamente familias de compuestos estructuralmente homogéneos.

#### **4.1.4. Conclusiones**

El modelo QSPR desarrollado para los índices de retención medidos en la columna apolar OV-101 tiene buena capacidad para predecir la propiedad de compuestos no evaluados y no sintetizados hasta el momento. Así, este estudio complementa modelos previamente desarrollados; sin embargo, tiene mayor capacidad de extensión a compuestos diversamente estructurales, debido a que la base de datos utilizada incluye compuestos heterogéneos. Por otro lado, el uso de descriptores independientes de la conformación es suficiente para modelar y predecir el índice de retención, mostrando claramente que estos modelos constituyen una alternativa válida para la predicción de propiedades cromatográficas.

### **4.2 Modelo QSPR para índices de retención medidos en la columna polar Carbowax 20M**

#### **4.2.1. Introducción**

Diversos estudios QSPR se han publicado para modelar y predecir el índice de retención en la fase estacionaria polar Carbowax 20M. Particularmente, grandes esfuerzos se han realizado para estudiar compuestos derivados del alquilbenceno (Woloszyn & Jurs, 1993; Dimov *et al.*, 1994; Sutter *et al.*, 1997; Yan *et al.*, 2000; Yan & Hu, 2001; Zhou & Nie, 2007; Porto *et al.*, 2008), derivados de la pirazina (Stanton & Jurs, 1989; Anker *et al.*, 1990; Edwards *et al.*, 1991; Touhami *et al.*, 2012), así como terpenos monocíclicos y acíclicos (Jalali-Heravi & Fatemi, 2001).

También se han desarrollado algunos trabajos concernientes a aromas y sabores. En 1990, Anker *et al.* (Anker *et al.*, 1990) midieron los índices de retención de 115 compuestos aromáticos y desarrollaron un modelo QSPR después de eliminar 4 valores atípicos ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 17.1$ ). Más adelante, en el año 2012, Yan *et al.* (Yan *et al.*, 2012) desarrollaron un modelo QSPR para 330 compuestos saborizantes y seis descriptores moleculares seleccionado a partir de un conjunto de

115 descriptores mediante regresión a pasos ( $R_{cal}^2 = 0.92$  y  $S_{cal} = 104.2$ ). Posteriormente, el modelo fue usado para predecir el  $I$  del grupo de validación de 107 moléculas ( $R_{val}^2 = 0.93$  y  $S_{val} = 105.5$ ).

El objetivo de este trabajo fue el desarrollo de un modelo QSPR para los índices de retención 1184 compuestos volátiles medidos en la columna polar Carbowax 20M en GC. Adicionalmente se ha usado el modelo para la predicción del  $I$  de 22 compuestos volátiles para los cuales no fue reportada dicha propiedad. Al igual que para el estudio de la sección 4.1, esta base de datos constituye la más grande considerada hasta el momento. Los resultados de este trabajo han sido previamente publicados (Rojas *et al.*, 2015b).

#### **4.2.2. Materiales y métodos**

En este estudio se utilizó la base de datos de 1208 compuestos publicada por Jennings y Shibamoto (Jennings & Shibamoto, 1980). La propiedad experimental es el índice de retención medido en la columna capilar (0.2 mm×80 m) cubierta con polietilenglicol Carbowax 20M y programada desde 70 hasta 170°C at 2°C×min<sup>-1</sup>. El dominio experimental se encuentra en el rango de 500 a 2640 unidades. Para esta columna, 22 compuestos no presentan el valor de  $I$ , por lo cual se han considerado como un grupo de predicción externo. Los nombres químicos y los índices de retención de los 1184 compuestos se presentan en la Tabla 2A, que se encuentra en el CD que acompaña a esta tesis. La metodología seguida para este estudio fue la misma usada para los índices de retención de la columna OV-101.

#### **4.2.3. Resultados y discusión**

El BSM permitió dividir la base de datos de 1184 compuestos en grupos de calibración ( $N_{cal} = 395$ ), validación ( $N_{val} = 396$ ) y predicción ( $N_{pred} = 393$ ). Las asignaciones a estos grupos se encuentran en la Tabla 2A. Posteriormente, con el grupo de calibración y predicción, se usó el RM para explorar tres grupos de descriptores: a) 2895 descriptores independientes y dependientes de la conformación, b) 1815 descriptores independientes de la conformación y c) 1080 descriptores dependientes de la conformación. Al igual que para la columna OV-101, se buscó identificar si los descriptores dependientes de la conformación aportan información relevante al modelado del índice de retención.

Las Tablas 4.5, 4.6 y 4.7 muestran los mejores modelos de 1 a 9 descriptores moleculares para cada uno de los tres grupos. En cada caso, el mejor modelo se indica en negrita y se observa la ausencia de sobreajuste en los mismos, debido a que los parámetros de calibración y validación son muy similares.

**Tabla 4.5.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes y dependientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$    | $R^2_{val}$ | $S_{val}$    | $R^2_{ij\ max}$ | Descriptores   |
|----------|-------------|--------------|-------------|--------------|-----------------|--|
| 1        | 0.72        | 232.3        | 0.69        | 222.0        | 0.00            | SCBO   |
| 2        | 0.81        | 189.9        | 0.80        | 175.7        | 0.11            | SCBO, Hy   |
| 3        | 0.85        | 170.0        | 0.85        | 155.4        | 0.11            | SCBO, C-001, Hy  |
| 4        | 0.88        | 153.2        | 0.88        | 135.6        | 0.87            | Sv, RDF010p, CATS2D_02_AA, Hy  |
| 5        | 0.89        | 146.6        | 0.90        | 127.8        | 0.88            | Sp, SM6_B(s), MATS2e, RDF010e, nHDon                                   |
| <b>6</b> | <b>0.91</b> | <b>134.9</b> | <b>0.91</b> | <b>119.4</b> | <b>0.88</b>     | <b>Sp, SM5_B(s), RDF010e, nHDon, C-005, CATS2D_04_DA</b>               |
| 7        | 0.90        | 138.0        | 0.91        | 119.9        | 0.91            | Sp, SM5_B(s), MATS2e, RDF010p, Mor18s, Hy, PDI                         |
| 8        | 0.92        | 127.9        | 0.92        | 116.1        | 0.91            | Sp, EE_B(s), MATS2e, RDF010p, C-005, CATS2D_04_DA, Hy, PDI             |
| 9        | 0.91        | 134.0        | 0.92        | 112.2        | 0.35            | SCBO, SM1_Dz(m), ATSC2s, E1s, nR=Cp, nHDon, C-001, NdssC, CATS2D_04_DA |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\ max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

**Tabla 4.6.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$    | $R^2_{val}$ | $S_{val}$    | $R^2_{ij\ max}$ | Descriptores   |
|----------|-------------|--------------|-------------|--------------|-----------------|--|
| 1        | 0.72        | 232.3        | 0.69        | 222.0        | 0.00            | SCBO   |
| 2        | 0.81        | 189.9        | 0.80        | 175.7        | 0.11            | SCBO, Hy   |
| 3        | 0.85        | 170.0        | 0.85        | 155.4        | 0.11            | SCBO, C-001, Hy  |
| 4        | 0.86        | 164.3        | 0.87        | 146.1        | 0.84            | Sv, P_VSA_m_1, C-002, Hy   |
| 5        | 0.87        | 157.8        | 0.89        | 133.2        | 0.94            | Se, P_VSA_m_1, NssCH2, CATS2D_04_DA, CATS2D_02_AA                              |
| 6        | 0.89        | 143.7        | 0.89        | 131.2        | 0.86            | H_Dz(v), SM3_B(p), SM6_B(s), C-001, SdCH2, Hy                                  |
| <b>7</b> | <b>0.90</b> | <b>137.6</b> | <b>0.91</b> | <b>119.8</b> | <b>0.85</b>     | <b>Se, Chi_H2, SM1_Dz(e), WiA_Dz(p), nHDon, SaaO, CATS2D_04_DA</b>             |
| 8        | 0.91        | 132.0        | 0.91        | 117.4        | 0.94            | pilD, XMOD, SpAD_B(m), SpDiam_B(s), C-001, CATS2D_04_DA, B02[O-O], Hy          |
| 9        | 0.91        | 134.2        | 0.92        | 115.0        | 0.71            | SCBO, Xindex, J_Dz(p), P_VSA_m_2, SpMAD_EA(dm), C-001, SdCH2, CATS2D_04_DA, Hy |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\ max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

Al igual que para la columna apolar OV-101, se observa que los descriptores conformacionales no aportan información significativa para modelar el índice de retención en la columna polar Carbowax 20M. En

efecto, cuando se usan únicamente los descriptores dependientes de la conformación (Tabla 4.7), el mejor modelo muestra menor calidad que los modelos considerando todos los descriptores (Tabla 4.5) y los modelos independientes de la conformación (Tabla 4.6).

**Tabla 4.7.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores dependientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$    | $R^2_{val}$ | $S_{val}$    | $R^2_{ij\ max}$ | descriptores   |
|----------|-------------|--------------|-------------|--------------|-----------------|--|
| 1        | 0.62        | 270.1        | 0.57        | 259.1        | 0.00            | G2   |
| 2        | 0.73        | 227.3        | 0.70        | 216.0        | 0.03            | G2, R2i  |
| 3        | 0.78        | 203.3        | 0.74        | 201.6        | 0.31            | G2, R1e+, R2i  |
| 4        | 0.82        | 186.1        | 0.78        | 186.2        | 0.35            | G2, Mor13p, H2i, R1e+  |
| 5        | 0.84        | 174.0        | 0.85        | 152.8        | 0.98            | G2, SpPos_G/D, RDF010e, RDF010p, Mor03u                            |
| <b>6</b> | <b>0.87</b> | <b>157.6</b> | <b>0.88</b> | <b>136.7</b> | <b>0.99</b>     | <b>Ho_G, SpMAD_G/D, RDF010p, RDF010i, Mor05p, Mor15i</b>           |
| 7        | 0.85        | 168.9        | 0.85        | 154.8        | 0.45            | G2, SpMAD_RG, SpMAD_G/D, Mor29v, Mor14s, Mor28s, R1s+              |
| 8        | 0.89        | 147.1        | 0.89        | 130.4        | 0.99            | Ho_G, SpMAD_G/D, RDF025v, RDF010p, RDF010i, Mor06u, Mor06i, Mor14i |
| 9        | 0.90        | 141.4        | 0.90        | 127.7        | 1.00            | Chi_RG, SpPos_G/D, RDF010u, RDF010p, Mor05p, Mor12s, Te, Tp, R1m+  |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\ max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

La selección de un modelo independiente de la conformación ayuda a evitar ambigüedades debidas a la optimización molecular y los costos computacionales involucrados (Garcia *et al.*, 2016). De esta manera, el modelo QSPR constituye la ecuación de 7 descriptores independientes de la conformación:

$$I = -157.2 - 24.7Se + 835Chi\_H2 + 1067.8SM1\_Dz(e) - 315.5WiA\_Dz(p) + 317.1nHDon - 36.3SaaO - 176CATS2D\_04\_DA \quad (4.2)$$

$$N_{cal} = 395, R^2_{cal} = 0.90, S_{cal} = 137.6$$

$$N_{val} = 396, R^2_{val} = 0.91, S_{val} = 119.8$$

$$N_{pred} = 393, R^2_{pred} = 0.90, S_{pred} = 121.5$$

$$R^2_{loo} = 0.90, S_{loo} = 130.1, R^2_{lmo} = 0.94, S_{lmo} = 134.1$$

$$S_{rand} = 408.7, \alpha(3S) = 14, R^2_{ij\ max} = 0.84$$

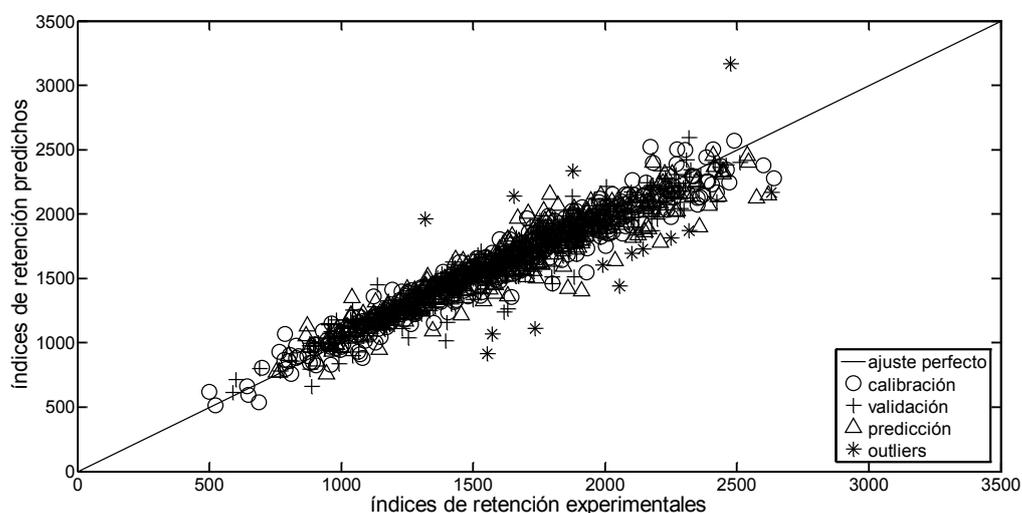
En la validación cruzada de dejar-varios-fuera (lmo) se excluyeron el 20% de moléculas. La aleatorización-Y refleja que  $S_{cal} < S_{rand}$ , indicando la ausencia de correlación casual en el modelo de la Ec. 4.2. Finalmente, los parámetros adicionales de validación que se presentan a continuación indican que una relación cuantitativa estructura-propiedad predictiva se ha alcanzado:

$$R_{loo}^2 > 0.5 \text{ (0.90)} \text{ y } R_{pred}^2 > 0.6 \text{ (0.90)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.002)} \text{ o } 1 - R_0'^2 / R_{pred}^2 < 0.1 \text{ (0.025)}$$

$$0.85 < k(1.003) \leq 1.15 \text{ y } 0.85 < k'(0.991) \leq 1.15$$

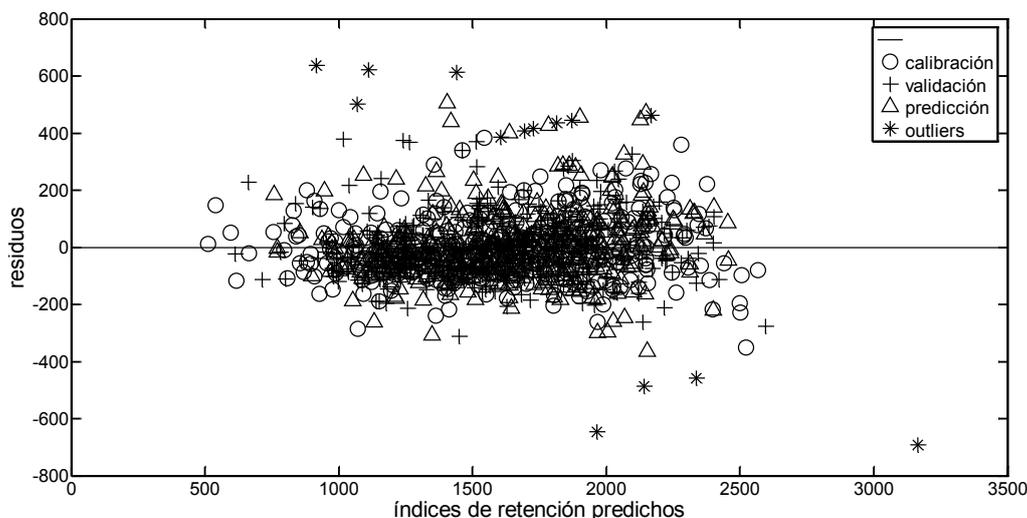
$$R_m^2 > 0.5 \text{ (0.86)}$$



**Figura 4.3.** Índices de retención experimentales y predichos para aromas medidos en la columna Carbowax 20M

La Figura 4.3 muestra la gráfica de los valores de índices de retención experimentales en función de los predichos por el modelo y la Figura 4.4 indica el gráfico de los valores de los residuos. Ambos gráficos revelan que los índices de retención son bien modelados por una regresión lineal múltiple.

Los catorce valores atípicos identificados en el modelo QSPR son las moléculas:  $\gamma$ -Decalactona (236),  $\delta$ -Decalactona (237), Éter dietílico de dietilenglicol (256),  $\gamma$ -Dodecalactona (301), Almizcle xilol (523),  $\gamma$ -Nonalactona (534),  $\delta$ -Undecalactona (671), Elemol (740), Dimetilsulfóxido (748), 2-hidroxiisobutirato de metilo (861), Isovalerato de metilo (1006), 6-Metilcumarina (1017), Bencil *n*-propilcarbinol (1136) y Jasmal (1147). Se controlaron en la fuente los valores de índices de retención y las fórmulas químicas, las cuales son correctas, por lo que se asume que estas moléculas se comporten como valores atípicos debido a la gran diversidad de compuestos presentes en la base de datos.



**Figura 4.4.** Gráfico de dispersión para el modelo QSPR de la Ec. 4.2

La correlación máxima entre los descriptores  $Se$  y  $Chi\_H2$  es  $R_{ij\ max}^2 = 0.84$ , indicando una correlación moderada entre los mismos. Los descriptores  $Chi\_H2$ ,  $SM1\_Dz(e)$  y  $nHDon$  tienen un efecto sinérgico sobre el índice de retención; mientras que  $Se$ ,  $WiA\_Dz(p)$ ,  $SaaO$  y  $CATS2D\_04\_DA$  tienen efecto antagónico sobre el  $I$ . El grado de contribución relativo de cada descriptor ( $b_j^s$ ), indica la importancia de cada descriptor en modelar el  $I$ :  $Chi\_H2$  (2.01) >  $WiA\_Dz(p)$  (0.99) >  $Se$  (0.54) >  $SM1\_Dz(e)$  (0,39) >  $nHDon$  (0.32) >  $CATS2D\_04\_DA$  (0.14) >  $SaaO$  (0.10).

El índice tipo Randić de la matriz de distancias recíprocas al cuadrado ( $Chi\_H2$ ) (Consonni & Todeschini, 2012) se calcula a partir de la matriz inversa de distancias al cuadrado ( $H2$ ) entre cualquier par de átomos (excepto H), considerando una generalización de la fórmula clásica del índice de conectividad de Randić basada en los grados de vértice. Este descriptor está estrechamente relacionado con el índice de Zagreb propuesto como un índice de Randić general. Se había indicado para la columna OV-101 que el índice de conectividad de Randić mide el grado de ramificación y compactación de las moléculas y se encuentra correlacionado con el tiempo de retención en GC y punto de ebullición. De esta manera, compuestos volátiles que tienen alto grado de ramificación tienen altos valores de la entropía de mezcla con la fase estacionaria. En consecuencia, debido a la estabilidad de la mezcla, se necesita incrementar la temperatura para lograr que el compuesto regrese a la fase móvil, lo cual lógicamente incrementa el  $I$ .

El índice tipo Wiener promedio de la matriz Barysz ponderado por la polarizabilidad ( $WiA\_Dz(p)$ ) y el momento espectral de orden 1 de la matriz de Barysz ponderado por la electronegatividad de Sanderson ( $SM1\_Dz(e)$ ) se calculan a partir de la matriz de Barysz ponderada por

las polarizabilidades atómicas y electronegatividades de Sanderson, respectivamente. La matriz de Barysz toma en cuenta contemporáneamente la presencia de heteroátomos y enlaces múltiples en la molécula (Ivanciuc, 2000).

Por otro lado, la suma de electronegatividades atómicas de Sanderson (escaladas sobre el átomo de C) ( $S_e$ ) (Kode srl., 2016) es una propiedad atómica calculada como la suma de las relaciones entre los valores de electronegatividad de cada átomo con respecto a la del carbono, indicando de esta manera cómo los electrones de enlace se distribuyen entre estos dos átomos cuando existe un enlace químico entre ellos.

$nHDon$  es un grupo funcional que indica la presencia de átomos nitrógeno y oxígeno donantes de enlaces hidrógeno. Este descriptor indica la capacidad de la molécula de interactuar con los constituyentes polares de la fase estacionaria y por consiguiente aumentar la retención de la molécula.

El descriptor CATS2D donante–aceptor a desplazamiento 4 ( $CATS2D_{04\_DA}$ ) indica la presencia de pares de átomos donante–aceptor de enlaces de hidrógeno separados por una distancia topológica de 4 unidades (Fechner *et al.*, 2003). Finalmente, La suma de los estados electrotopológicos aaO ( $S_{aaO}$ ) (Hall & Kier, 1995) cuantifica la presencia de átomos de oxígeno unido a dos enlaces aromáticos.

La Tabla 4.8 presenta una comparación de los modelos QSPR desarrollados para índices de retención medidos en la columna polar Carbowax 20M.

**Tabla 4.8.** Comparación de varios modelos QSPR para los índices de retención medidos en la columna Carbowax–20M

| modelo                         | compuestos                      | N   | d | $R^2_{cal}$ | $S_{cal}$ | $R^2_{pred}$   | $S_{pred}$ |
|--------------------------------|---------------------------------|-----|---|-------------|-----------|----------------|------------|
| (Stanton & Jurs, 1989)         | Pirazinas sustituidas           | 107 | 6 | 0.99        | 33.7      | 0.99           | 51.6       |
| (Anker <i>et al.</i> , 1990)   | Compuestos aromáticos           | 111 | 7 | 0.99        | 17.1      | — <sup>a</sup> | —          |
| (Edwards <i>et al.</i> , 1991) | Pirazinas mono y di sustituidas | 74  | 4 | 0.99        | 40.3      | —              | —          |
| (Woloszyn & Jurs, 1993)        | Mezcla de hidrocarburos         | 81  | 5 | 0.97        | 23.3      | —              | —          |
| (Dimov <i>et al.</i> , 1994)   | Alquilbencenos                  | 40  | 3 | 0.99        | 13.0      | —              | —          |
| (Sutter <i>et al.</i> , 1997)  | Alquilbencenos (MLR)            | 150 | 6 | 0.98        | 18.0      | —              | 21.8       |
| (Yan <i>et al.</i> , 2000)     | Alquilbencenos (ANN)            | 165 | 7 | —           | 11.7      | —              | 18.9       |
|                                | Alquilbencenos (MLR)            |     |   | 0.88        | 44.8      | —              | —          |
|                                | Alquilbencenos (NMLR)           |     |   | 0.97        | 23.1      | —              | —          |
|                                | Alquilbencenos (ANN1)           |     |   | 0.97        | 21.8      | —              | —          |
|                                | Alquilbencenos (ANN2)           |     |   | 0.98        | 19.5      | —              | —          |
| (Yan & Hu, 2001)               | Alquilbencenos (MLR)            | 170 | 8 | 0.97        | 26.1      | 0.99           | 18.2       |
|                                | Alquilbencenos                  |     |   | 0.99        | 15.5      | 0.99           | 16.6       |

|                                |   |      |   |      |       |      |       |
|--------------------------------|---|------|---|------|-------|------|-------|
|                                | (NMLR)                                  |      |   |      |       |      |       |
|                                | Alquilbencenos (ANN)                    |      |   | 0.99 | 12.6  | 0.99 | 13.2  |
| (Jalali-Heravi & Fatemi, 2001) | Terpenos acíclicos y monocíclicos (MLR) | 53   | 6 | 0.77 | —     | 0.50 | —     |
|                                | Terpenos acíclicos y monocíclicos (ANN) |      |   | 0.94 | —     | 0.88 | —     |
| (Zhou & Nie, 2007)             | Alquilbencenos                          | —    | 3 | 0.99 | 17.6  | —    | —     |
| (Porto <i>et al.</i> , 2008)   | Alquilbencenos                          | 34   | 1 | 0.95 | 25.1  | —    | —     |
| (Touhami <i>et al.</i> , 2012) | Pirazinas                               | 35   | 3 | 0.99 | —     | 0.98 | —     |
| (Yan <i>et al.</i> , 2012)     | Compuestos de sabor                     | 434  | 6 | 0.92 | 104.2 | 0.93 | 105.5 |
| Presente <sup>b</sup>          | Sabores y fragancias                    | 1184 | 7 | 0.90 | 137.6 | 0.90 | 121.5 |

<sup>a</sup> no disponible; <sup>b</sup> (Rojas *et al.*, 2015b)

La base de datos considerada en este estudio constituye la más grande usada hasta el momento. En efecto, es 2.7 veces mayor a la más grande previamente utilizada (Yan *et al.*, 2012). Adicionalmente, seis modelos QSPR no han sido sometidos a validación externa (Anker *et al.*, 1990; Edwards *et al.*, 1991; Dimov *et al.*, 1994; Yan *et al.*, 2000; Zhou & Nie, 2007; Porto *et al.*, 2008) y uno no presenta los resultados para el grupo de predicción de 81 hidrocarburos (Woloszyn & Jurs, 1993).

Las predicciones del modelo QSPR de la Ec. 4.2, se restringen a compuestos volátiles para los cuales el  $l$  esté dentro del rango de 500 a 2640 y cuyo valor de influencia sea menor al valor de influencia crítico ( $h^* = 0.030$ ). El análisis del dominio de aplicabilidad (AD) indica que 13 compuestos del grupo de predicción caen fuera del AD del modelo: *Isoamil antranilato* ( $h_{ij} = 0.051$ ), *Antranilato de isobutilo* ( $h_{ij} = 0.050$ ), *Antranilato de etilo* ( $h_{ij} = 0.050$ ), *3-Hidroxibutirato de metilo* ( $h_{ij} = 0.048$ ), *Salicilato de cis-3-hexenilo* ( $h_{ij} = 0.046$ ), *Salicilato de alilo* ( $h_{ij} = 0.045$ ), *Metil n-metil antranilato* ( $h_{ij} = 0.045$ ), *Metil N-propil antranilato* ( $h_{ij} = 0.045$ ), *n-Butil N-metil antranilato* ( $h_{ij} = 0.045$ ), *Salicilato de isopropilo* ( $h_{ij} = 0.045$ ), *Isopropil n-metil antranilato* ( $h_{ij} = 0.045$ ), *m-Aminoacetofenona* ( $h_{ij} = 0.034$ ) y *2-Isopropil-4-etil-5-metiloxazol* ( $h_{ij} = 0.031$ ). Se observa que la mayoría de compuestos fuera el AD del modelo corresponden a derivados del antranilato y salicilato. Posteriormente, se verificó que las respuestas experimentales y las estructuras químicas son las correctas en la fuente. Se indicó previamente (columna OV-101) que este tipo de comportamiento se puede deber a la heterogeneidad de la base de datos, así como a las limitaciones predictivas intrínsecas del modelo definidas por el AD (Dimitrov *et al.*, 2005; Gramatica, 2007).

Por otro lado, varios modelos QSPR presentados en la Tabla 4.8 se han desarrollado únicamente usando familias homogéneas de compuestos, por ejemplo alquilbencenos o pirazinas, lo cual reduce la habilidad de dichos modelos a generalizar el AD para otros compuestos volátiles (Stanton & Jurs, 1989; Edwards *et al.*, 1991; Woloszyn & Jurs,

1993; Dimov *et al.*, 1994; Sutter *et al.*, 1997; Yan *et al.*, 2000; Jalali-Heravi & Fatemi, 2001; Yan & Hu, 2001; Zhou & Nie, 2007; Porto *et al.*, 2008; Touhami *et al.*, 2012).

Usando el modelo QSPR de la Ec. 4.2, se han predicho los índices de retención de los 22 compuestos volátiles. En la Tabla 4.9 se presentan los valores predichos y el valor de influencia para estos compuestos. Existen 6 moléculas que caen fuera el AD del modelo, es decir, sus predicciones son extrapolaciones. Estos compuestos son: *pinacol*, *salicilato de fenilo*, *cis-3-hexenil antranilato*, *salicilato de p-cresilo*, *cinamil cinamato*, *antranilato de feniletilo*.

**Tabla 4.9.** Valores de influencia e índice de retención predicho para 22 compuestos aromáticos en la columna Carbowax 20M

| nombre                           | $h_{ij}$           | $I_{pred}$ | $I_{literatura}$  |
|----------------------------------|--------------------|------------|---|
| <i>Pinacol</i>                   | 0.039 <sup>a</sup> | 1499.7     | 1338 <sup>b</sup> (Shiratsuchi <i>et al.</i> , 1994; Linstrom & Mallard, 2001)  |
| <i>sec-Butilbenceno</i>          | 0.009              | 1415.3     | 1227, 1248 (Linstrom & Mallard, 2001)   |
| <i>Hidrato de trans-Sabineno</i> | 0.010              | 1664.3     | 1458 <sup>c</sup> (Acree & Arn, 2004)<br>1548.9 (Babushok <i>et al.</i> , 2011)   |
| $\gamma$ - <i>Muurolene</i>      | 0.013              | 1691.2     | 1675 (Bicchi <i>et al.</i> , 2003)<br>1681 (Acree & Arn, 2004)<br>1689.8 <sup>d</sup> (Babushok <i>et al.</i> , 2011)           |
| <i>Anisil isobutirato</i>        | 0.006              | 2024.1     | —*  |
| <i>Zingerona</i>                 | 0.019              | 2557.2     | —*  |
| <i>Anisil n-valerato</i>         | 0.008              | 2198.0     | 2400 (Linstrom & Mallard, 2001)   |
| <i>Benzoato de bencilo</i>       | 0.020              | 2485.8     | 2071 <sup>c</sup> (Acree & Arn, 2004)<br>2612.7 <sup>d</sup> (Babushok <i>et al.</i> , 2011)<br>2655 (Linstrom & Mallard, 2001) |
| <i>Salicilato de fenilo</i>      | 0.051 <sup>a</sup> | 2578.2     | —*  |
| <i>Benzoato de bornilo</i>       | 0.007              | 2294.4     | 2114 <sup>c</sup> (Acree & Arn, 2004)   |
| <i>Benzoato de p-cresilo</i>     | 0.016              | 2434.2     | —*  |
| <i>cis-3-hexenil antranilato</i> | 0.055 <sup>a</sup> | 2560.7     | —*  |
| <i>Fenilacetato de p-cresilo</i> | 0.017              | 2514.7     | —*  |
| <i>Benzoato de feniletilo</i>    | 0.022              | 2551.4     | 2189 <sup>c</sup> (Acree & Arn, 2004)<br>2654 (Linstrom & Mallard, 2001)  |
| <i>Salicilato de p-cresilo</i>   | 0.049 <sup>a</sup> | 2589.1     | —*  |
| <i>2-Mercaptobenzotiazol</i>     | 0.014              | 2031.7     | —*  |
| <i>Benzoato de geranilo</i>      | 0.011              | 2436.9     | 2617 <sup>e</sup> (Linstrom & Mallard, 2001)  |
| <i>Cinamil cinamato</i>          | 0.039 <sup>a</sup> | 3001.9     | —*  |
| <i>Antranilato de feniletilo</i> | 0.065 <sup>a</sup> | 2952.3     | —*  |
| <i>Cinamato de feniletilo</i>    | 0.030              | 2822.3     | 3100 (Linstrom & Mallard, 2001)   |
| <i>n-heptanoato de anisilo</i>   | 0.010              | 2364.2     | 2600 (Linstrom & Mallard, 2001)   |
| <i>n-hexanoato de anisilo</i>    | 0.009              | 2280.2     | —*  |

<sup>a</sup> moléculas con influencia superior a  $h^* = 0.030$ . <sup>b</sup> medido en la columna polar DB-Wax. <sup>c</sup> Interpolado. <sup>d</sup> medido en la columna Polietilenglicol (PEG). <sup>e</sup> medido en la columna polar HP-Innowax FSC. \* no reportado en la literatura.

Según los resultados presentados en la Tabla 4.9, el *Cinamato de feniletilo* presenta la mayor desviación (291 unidades) entre el  $I$  predicho

y reportado en la literatura. También otros compuestos muestran diferencias importantes entre estos dos valores: *n*-heptanoato de anisilo (245 unidades), Anisil *n*-valerato (211 unidades), *sec*-Butilbenceno (186 unidades), Benzoato de geranilo (185 unidades), pinacol (172 unidades), Hidrato de *trans*-Sabineno (127 unidades) y Benzoato de feniletilo (114 unidades). Sin embargo, el modelo es confiable para predecir el índice de retención del  $\gamma$ -muurolene (28 unidades) y Benzoato de feniletilo (95 unidades). Un aspecto interesante es que los compuestos benzoato de bencilo y benzoato de feniletilo presentan *I* predichos que están en mejor concordancia con respecto a los valores reportados en la base de datos Flavornet (Acree & Arn, 2004). Por otro lado, cinco moléculas (benzoato de bencilo, benzoato de feniletilo, cinamato de feniletilo, cinamil cinamato, antranilato de feniletilo) tienen valores de *I* predicho o reportado por encima del valor máximo del modelo (2640).

La relativa baja precisión en la predicción de los índices de retención presentados en la Tabla 4.9 se pueden atribuir a la baja reproducibilidad entre laboratorios, la cual se puede deber a diferentes factores que afectan el análisis GC (Soják *et al.*, 1981; Wittkowski & Matissek, 1993; Woloszyn & Jurs, 1993; Barwick, 1999; Babushok *et al.*, 2007):

1. Interacción polar entre el compuesto volátil analizado y las paredes de la fase estacionaria.
2. Identificación errónea de los compuestos.
3. Sensibilidad del equipo GC.
4. Variaciones de las condiciones de medición de los *I*, por ejemplo, tipo de gas usado, condiciones de temperatura y parámetros del detector.
5. Variaciones en las propiedades intrínsecas de la fase estacionaria: tipo de columna, espesor de película de la fase estacionaria, dimensión de la columna y antigüedad de la misma.
6. Propiedades de la muestra (estabilidad térmica y la presencia de impurezas) y factores involucrados durante la preparación de la muestra.
7. Factores que afectan la introducción de la muestra en el equipo GC: selección del sistema apropiado de inyección (poco o excesivo volumen y partición de la muestra).
8. Fuentes de errores “espurios” (mal funcionamiento del equipo o errores del analista), errores y variabilidad asociados con la integración de los picos; así como errores normales de experimentación y errores de entrada de datos.

Finalmente, Gramatica (Gramatica, 2007) indicó que las predicciones erróneas posiblemente se pueden deber a datos experimentales incorrectos.

#### **4.2.4. Conclusiones**

En este estudio se ha desarrollado un modelo QSPR con buena capacidad predictiva para los índices de retención medidos en la columna polar Carbowax 20M. De esta forma, se han predicho los  $k'$  para 22 compuestos volátiles para los cuales no se reportó la propiedad en la base de datos. Se compararon las predicciones con valores reportados en la literatura y se estableció que el modelo falla principalmente en la predicción del  $k'$  de derivados del antranilato y salicilato. Por lo tanto, se discutieron las posibles causas de discrepancias entre los valores predichos y experimentales, las cuales se deberían a factores que afectan la reproducibilidad del análisis GC, datos erróneos reportados en la literatura y las limitaciones intrínsecas del modelo QSPR. Sin embargo, el modelo desarrollado en este estudio permite una relación cuantitativa estructura–propiedad más general y, al igual que para la columna apolar OV–101, los descriptores independientes de la conformación son suficientes para generar modelos predictivos.

### **4.3 Modelo QSPR para índices de retención de aromas medidos en la columna polar DB–225MS**

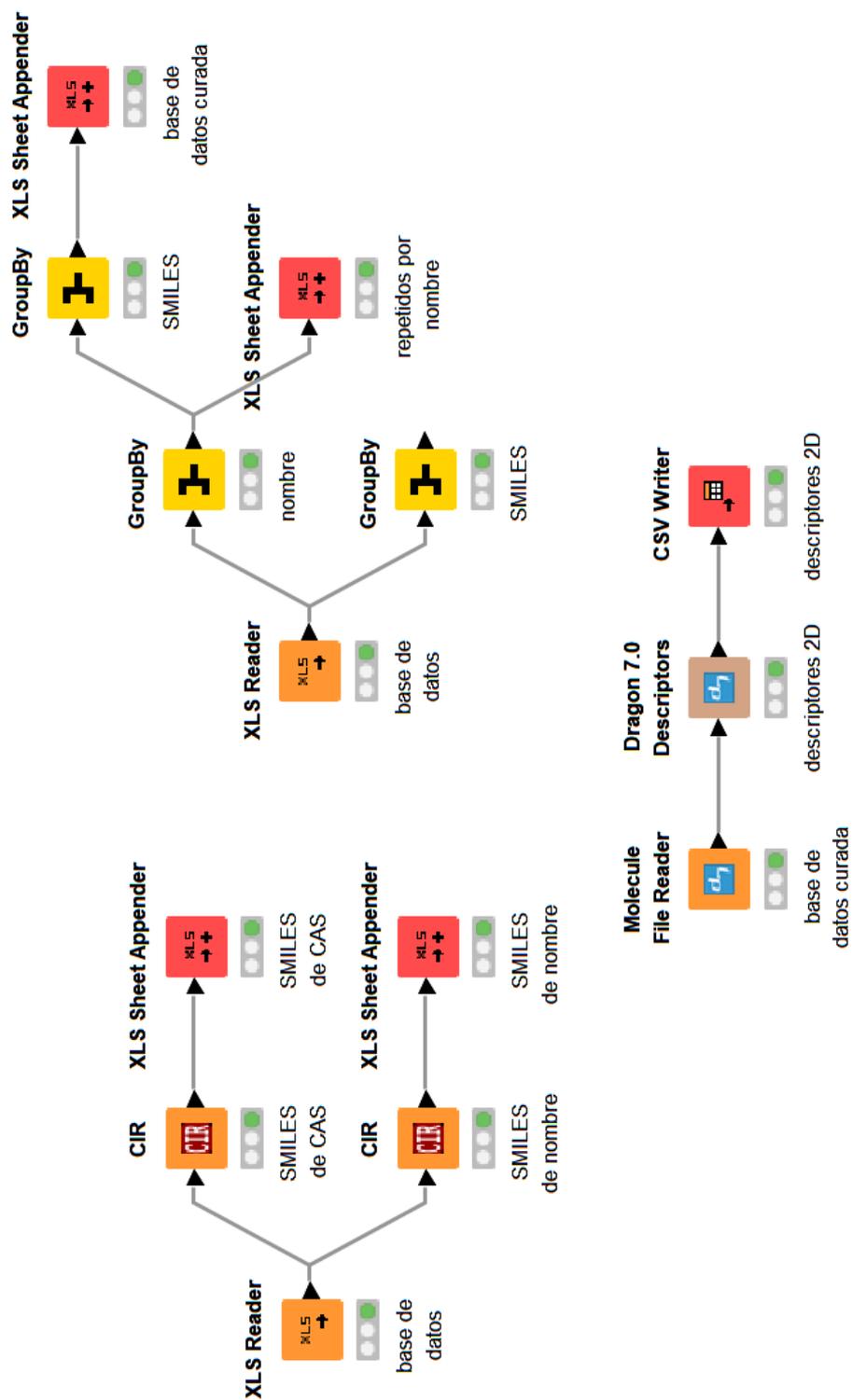
#### **4.3.1. Introducción**

Hasta donde se conoce, luego de una búsqueda exhaustiva en la literatura, no existen modelos QSPR desarrollados para índices de retención de compuestos volátiles en la columna polar DB–225MS. Con este antecedente, el objetivo de este estudio ha sido desarrollar la primera relación cuantitativa estructura–propiedad para los índices de retención medidos en la columna polar DB–225MS en GC. Los resultados de este trabajo han sido recientemente aceptados para su publicación en una revista científica con referato (Rojas *et al.*, 2017).

#### **4.3.2. Materiales y métodos**

Los índices de retención experimentales de 312 compuestos aromáticos se tomaron de la base de datos desarrollada por Yan *et al.* (Yan *et al.*, 2015). Esta propiedad fue medida en un cromatógrafo de Gas Agilent 7890A acoplado con un espectrómetro de masas Agilent 5975C. Los nombres químicos, el número de registro CAS y los índices de retención experimental se importaron en el programa KNIME (Berthold *et al.*, 2008). La notación de cadena SMILES de cada estructura se obtuvo a partir del nombre y del CAS, mediante el nodo del solucionador de identificación química (CIR) (NCI/CADD Group, 2013). Para cada molécula, se verificó que la notación SMILES coincidiera con el obtenido a partir del nombre y del número CAS; de esta manera, aquellos compuestos que tienen notación SMILES diferente fueron

chequeados manualmente en las bibliotecas químicas PubChem (Kim *et al.*, 2015), ChemSpider (Pence & Williams, 2010) y NIST Chemistry WebBook (Linstrom & Mallard, 2001). Durante el curado de la base de datos, se identificaron 41 compuestos repetidos por nombre o CAS. Debido a que el objetivo es desarrollar un modelo QSPR independiente de la conformación, 2 compuestos que se identificaron como duplicados fueron removidos. Se usó el valor promedio *I* para los compuestos repetidos por nombre o CAS, así como los compuestos similares en notación de cadena SMILES. De esta manera, la base de datos curada está constituida de 269 moléculas (Tabla 3A que se encuentra en el CD que acompaña a esta tesis). El diagrama de flujo programado en KNIME para el filtrado y curado de la base de datos se presenta en la Figura 4.5.



**Figura 4.5.** Diagrama de flujo KNIME para el curado de las bases de datos de índices de retención medidos en las columnas DB–225MS, HP5–MS y HP–1

Posteriormente, se calcularon 3808 descriptores moleculares independientes de la conformación usando el nodo de DRAGON 7 (Kode srl., 2016) implementado en KNIME (Figura 4.5). Se excluyeron aquellos descriptores con valores contantes o casi constantes y aquellos con al menos un valor faltante. Así, 1819 descriptores fueron analizados mediante el método de reemplazo para la selección de descriptores y encontrar el modelo óptimo. Para la validación del modelo, la base de datos se dividió en conjuntos de calibración, validación y predicción mediante el BSM. El grupo de calibración se usó para la selección de descriptores mediante RM y el ajuste de los modelos; mientras que el grupo de validación se usó para evitar la presencia de sobreajuste durante la selección supervisada. También se han usado las técnicas de validación cruzada de dejar–uno–fuera (loo) y dejar–varios–fuera (lmo) mediante la exclusión aleatoria del 20% de las moléculas, con 50000 repeticiones. La ausencia de correlación casual se estableció mediante la aleatorización–Y con 10000 repeticiones. Adicionalmente, se analizaron los criterios de validación propuestos por Golbraikh y Tropsha. Finalmente, se definió el dominio de aplicabilidad siguiendo el enfoque del influencia crítico ( $h^*$ ), así como la contribución de cada descriptor molecular y la interpretación mecanicista del modelo.

**Tabla 4.10.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$    | $R^2_{val}$ | $S_{val}$    | $R^2_{ij\ max}$ | descriptores   |
|----------|-------------|--------------|-------------|--------------|-----------------|--|
| 1        | 0.72        | 253.6        | 0.74        | 216.5        | 0               | <i>SNar</i>  |
| 2        | 0.79        | 218.2        | 0.82        | 183.1        | 0.01            | <i>J_Dt, SpAD_B(s)</i>   |
| 3        | 0.78        | 224.7        | 0.83        | 180.6        | 0.64            | <i>Chi_B(i), MATS1p, P_VSA_m_1</i>                             |
| <b>4</b> | <b>0.82</b> | <b>205.2</b> | <b>0.84</b> | <b>173.1</b> | <b>0.12</b>     | <b><i>SpAbs_B(e), MATS1v, C-001, C-008</i></b>                 |
| 5        | 0.86        | 181.7        | 0.84        | 172.6        | 0.56            | <i>SpMaxA_Dz(p), J_Dz(i), SpAbs_B(m), P_VSA_v_2, nCrt</i>      |
| 6        | 0.88        | 168.0        | 0.84        | 174.7        | 0.69            | <i>SCBO, J_D, P_VSA_m_2, nCrt, C-006, CATS2D_08_DA</i>         |
| 7        | 0.91        | 148.5        | 0.83        | 180.6        | 0.70            | <i>SCBO, RFD, X5sol, GATS6v, P_VSA_m_2, nCrt, CATS2D_08_DA</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\ max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

### 4.3.3. Resultados y discusión

La base de datos de 269 compuestos fue dividida mediante el BSM en grupos de calibración y validación de 90 moléculas cada uno, y un grupo de predicción de 89 compuestos. Las asignaciones realizadas por BSM se detallan en la Tabla 3A. El método de reemplazo permitió desarrollar modelos a partir de 1819 descriptores independientes de la

conformación. La Tabla 4.10 resume los mejores modelos QSPR de 1 a 7 descriptores, en los cuales se observa la ausencia de sobreajuste.

La selección del modelo óptimo (en negrita) propone como el candidato más apropiado al modelo constituido por 4 descriptores:

$$I = 702.6 + 65.9 \text{SpAbs}_- B(e) + 1465.6 \text{MAT S1v} - 105.9C - 001 + 235.7C - 008 \quad (4.3)$$

$$N_{cal} = 90, R_{cal}^2 = 0.82, S_{cal} = 205.2$$

$$N_{val} = 90, R_{val}^2 = 0.84, S_{val} = 173.1$$

$$N_{pred} = 89, R_{pred}^2 = 0.76, S_{pred} = 179.9$$

$$R_{loo}^2 = 0.82, S_{loo} = 191.5, R_{lmo}^2 = 0.88, S_{lmo} = 212.4$$

$$S_{rand} = 418.8, \alpha(3S) = 3, R_{ij \max}^2 = 0.16$$

Los resultados del modelo y los criterios adicionales de validación que se detallan a continuación confirman que se ha obtenido un modelo QSPR predictivo para los índices de retención medidos en la columna DB-225MS:

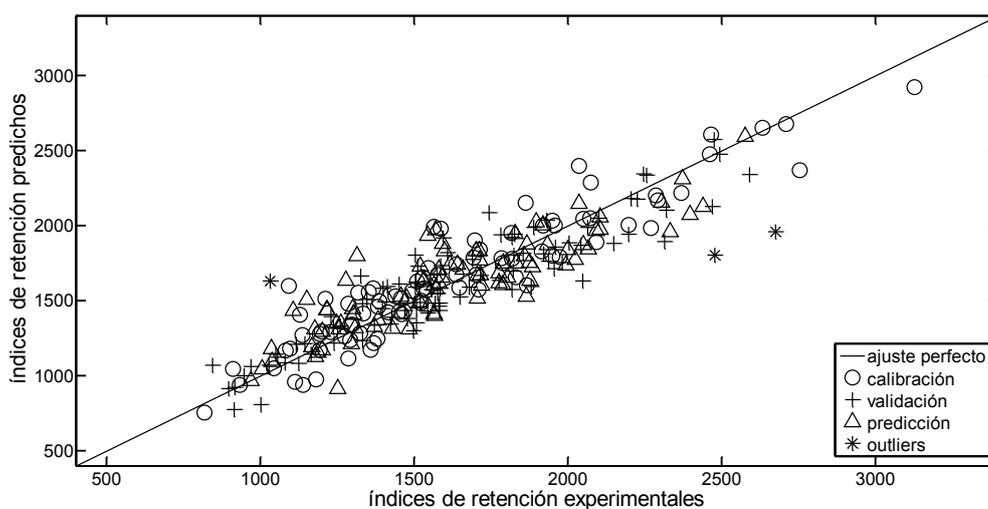
$$R_{loo}^2 > 0.5 \text{ (0.82) y } R_{pred}^2 > 0.6 \text{ (0.76)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.000) o } 1 - R_0'^2 / R_{pred}^2 < 0.1 \text{ (0.10)}$$

$$0.85 \leq k(0.99) \leq 1.15 \text{ y } 0.85 \leq k'(0.99) \leq 1.15$$

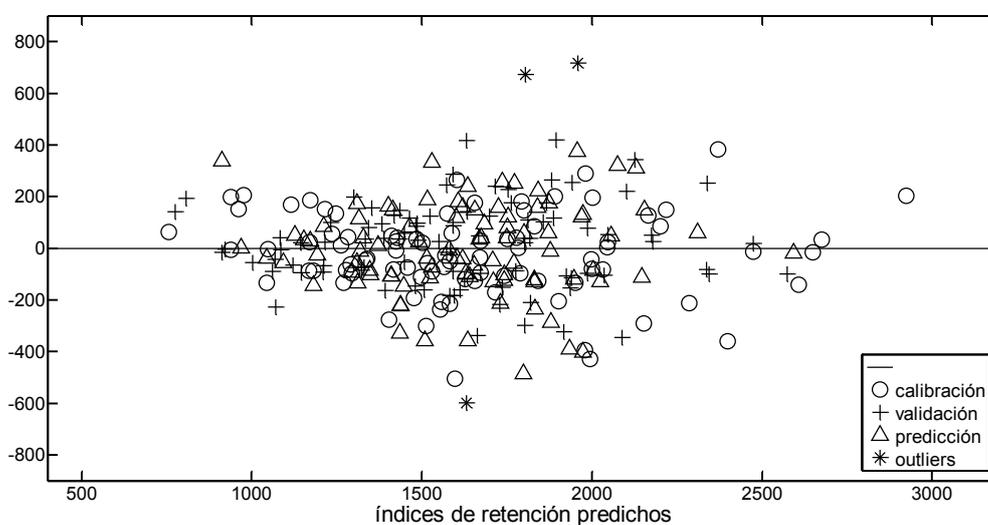
$$R_m^2 > 0.5 \text{ (0.75)}$$

En la Figura 4.6 se observa la tendencia lineal alrededor de la recta de ajuste perfecto de los valores de índices de retención experimentales en función de los predichos por el modelo. Por otra parte, en la Figura 4.7 se muestra el gráfico de los valores de los residuos, el cual revela que dichos valores tienden a seguir un patrón aleatorio, con lo cual se acepta la suposición de que los índices de retención se modelan mediante una regresión lineal múltiple.



**Figura 4.6.** Índices de retención experimentales y predichos para aromas medidos en la columna DB-225MS

Existen tres moléculas identificadas como valores atípicos: *dihidroactinolida*, *4-(4-Hidroxifenil)-2-butanona*, y *alcohol de bencilo*. Se comprobó que la fórmula química y el índice de retención sean correctos. Como se había indicado, este comportamiento se puede deber a la heterogeneidad de las estructuras químicas consideradas.



**Figura 4.7.** Gráfico de dispersión para el modelo QSPR de la Ec. 4.3

La máxima correlación ( $R_{ij \max}^2 = 0.16$ ) entre los descriptores *SpAbs\_B(e)* y *MATS1v* indica una baja correlación entre los descriptores del modelo, por lo que cada descriptor refleja aspectos particulares en el modelado de la propiedad (Duchowicz *et al.*, 2010). Por otra parte, el grado de contribución relativo de cada descriptor indica que *SpAbs\_B(e)* (0.78) > *C-001* (0.28) > *MATS1v* (0.20) > *C-008* (0.17).

La energía del grafo de la matriz de Burden ponderada por la electronegatividad de Sanderson ( $SpAbs\_B(e)$ ) es un índice topológico que se calcula como la suma de los valores absolutos de todos los autovalores (invariante del grafo) de la matriz de Burden ponderada por la electronegatividad de Sanderson (Gutrnan *et al.*, 2003; Consonni & Todeschini, 2008). Este descriptor indica que la presencia de átomos electronegativos (O, N, S, P, Se o halógenos) o grupos de átomos potencialmente involucrados en enlaces de hidrógeno tienden a incrementar el índice de retención debido a la habilidad que tienen para atraer electrones hacia ellos. En consecuencia, la fragancia es capaz de crear interacciones polares con los componentes de la fase estacionaria. Este hecho se confirma por la presencia de CH enlazados a dos átomos de carbono y un átomo electronegativo en la molécula (C-008) (Ghose *et al.*, 1998).

La autocorrelación de Moran a desplazamiento 1 ponderada por el volumen de van der Waals ( $MATS1v$ ) se calcula aplicando el coeficiente de Moran (Moran, 1950) a un grafo molecular completo y ponderado por el volumen de van der Waals. Este descriptor provee información concerniente a la distribución de esta propiedad a lo largo de la estructura topológica de la molécula. Índices de retención altos están relacionados a valores positivos del coeficiente de Moran (autocorrelaciones espaciales positivas), es decir, el compuesto posee átomos con volumen de van der Waals similar a desplazamiento 1. Finalmente, el coeficiente negativo del descriptor C-001 (Ghose *et al.*, 1998) indica que los índices de retención están inversamente relacionados con la presencia de fragmentos  $CH_3$  en la molécula. Este comportamiento de los índices de retención se ha observado en hidrocarburos ramificados (Carlson *et al.*, 1998; Katritzky *et al.*, 2000).

Finalmente, el dominio de aplicabilidad del modelo permitió establecer la limitación de los 4 descriptores moleculares en la generación de predicciones confiables; es decir, los  $I$  predichos están restringidos a moléculas que presentan valor de influencia inferior al valor crítico ( $h^* = 0.083$ ). En este estudio no existen moléculas del grupo de predicción fuera del dominio de aplicabilidad, por lo tanto, el modelo fue capaz de predecir los índices de retención de todas las fragancias. Ver Tabla 3A.

#### 4.3.4. Conclusiones

El modelo QSPR desarrollado en este estudio constituye el primero para el estudio de los índices de retención medidos en la columna polar DB-225MS. La aplicación del enfoque QSPR independiente de la conformación resultó ser apropiado y este hecho confirma lo previamente expuesto en el estudio de los  $I$  para la columna polar Carbowax 20M. El modelo basado en 4 descriptores refleja que la energía del grafo de la matriz de Burden ponderada por la electronegatividad de Sanderson, la autocorrelación de Moran a

desplazamiento 1 ponderada por el volumen de van der Waals, y la presencia de fragmentos CH<sub>3</sub> y fragmentos CH enlazados a dos átomos electronegativos en la molécula son suficientes para predecir esta propiedad.

## 4.4 Modelo QSPR para índices de retención de aromas medidos en la columna ligeramente polar HP5–MS

### 4.4.1. Introducción

Diversos estudios QSPR han sido reportados en la literatura para modelar y predecir los índices de retención en la fase estacionaria ligeramente polar HP5–MS. En el año 2003 Eckel y Kind (Eckel & Kind, 2003) realizaron un modelo de regresión simple entre el punto de ebullición y el índice de retención de Lee de 56 compuestos ( $R_{cal}^2 = 0.97$ ). El mismo año, Wang et al. (Wang *et al.*, 2003) desarrollaron un modelo QSPR para 18 hidrocarburos aromáticos policíclicos clorados (Cl-PAHs) con buenos resultados ( $R_{cal}^2 = 0.99$ ).

Posteriormente, en el año 2009 Qin et al. (Qin *et al.*, 2009) desarrollaron 3 modelos QSPR para los tiempos de retención de 96 aceites esenciales. El primer modelo fue en base a regresión MLR con 6 descriptores ( $R_{cal}^2 = 0.98$ ,  $S_{cal} = 1.64$ ,  $R_{val}^2 = 0.94$  y  $S_{val} = 2.65$ ), el segundo modelo se obtuvo mediante regresión en componentes principales (PCR) con 4 componentes ( $R_{cal}^2 = 0.97$ ,  $S_{cal} = 1.92$ ,  $R_{val}^2 = 0.95$  y  $S_{val} = 2.44$ ), mientras que el tercer modelo se desarrolló usando PLS con 5 variables latentes ( $R_{cal}^2 = 0.97$ ,  $S_{cal} = 1.81$ ,  $R_{val}^2 = 0.96$  y  $S_{val} = 2.19$ ). El mismo año, Riahi et al. (Riahi *et al.*, 2009) trabajaron con 100 índices de retención de aceites esenciales de *Citrus sudachi*. Usaron GAs para calibrar un modelo MLR ( $R_{cal}^2 = 0.95$ ,  $S_{cal} = 48.3$ ,  $R_{val}^2 = 0.93$  y  $S_{val} = 60.4$ ) y otro PLS ( $R_{cal}^2 = 0.94$ ,  $S_{cal} = 50.3$ ,  $R_{val}^2 = 0.91$  y  $S_{val} = 67.0$ ) con seis descriptores y 3 LVs. Adicionalmente desarrollaron dos modelos no lineales: un modelo polinomial PLS con 3 LVs ( $R_{cal}^2 = 0.95$ ,  $S_{cal} = 46.0$ ,  $R_{val}^2 = 0.93$  y  $S_{val} = 56.9$ ) y un modelo basado en máquinas de soporte vectorial (SVM) ( $R_{cal}^2 = 0.99$ ,  $S_{cal} = 24.7$ ,  $R_{val}^2 = 0.96$  y  $S_{val} = 51.4$ ).

Por otro lado, en el año 2011 Mohammadhosseini et al. (Mohammadhosseini *et al.*, 2011) establecieron un modelo QSPR con

tres descriptores para 29 compuestos de calibración ( $R_{cal}^2 = 0.98$ ) y 14 moléculas en el grupo de validación ( $R_{val}^2 = 0.95$ ). Al siguiente año, Noorizadeh y Noorizadeh (Noorizadeh & Noorizadeh, 2012) estudiaron los tiempos de retención de 69 opiáceos y sedantes mediante el uso de descriptores DRAGON seleccionados mediante GAs. Se desarrollaron dos modelos lineales: 1) MLR con seis descriptores ( $R_{cal}^2 = 0.91$ ,  $S_{cal} = 95.6$ ,  $R_{val}^2 = 0.87$  y  $S_{val} = 163.6$ ); y 2) PLS con ocho descriptores y 5 LVs ( $R_{cal}^2 = 0.92$ ,  $S_{cal} = 90.5$ ,  $R_{val}^2 = 0.88$  y  $S_{val} = 157.9$ ). También desarrollaron dos modelos no lineales: 1) kernel PLS ( $R_{cal}^2 = 0.93$ ,  $S_{cal} = 81.9$ ,  $R_{val}^2 = 0.90$  y  $S_{val} = 141.3$ ) y 2) red neuronal artificial de Levenberg–Marquardt (L–M ANN) ( $R_{cal}^2 = 0.96$ ,  $S_{cal} = 68.4$ ,  $R_{val}^2 = 0.93$  y  $S_{val} = 117.7$ ). Ese mismo año, Zhao et al. (Zhao *et al.*, 2012) utilizaron una base de datos de 178 compuestos orgánicos volátiles para predecir los índices de retención a temperatura programada usando 9 bloques de índices topológicos en subespacios ortogonales con validación cruzada Monte–Carlo ( $R_{cal}^2 = 0.99$ ,  $S_{cal} = 36.0$ ,  $S_{val} = 32.6$ ). Posteriormente, Qin et al. (Qin *et al.*, 2013) desarrollaron un modelo QSPR para los índices de retención de los constituyentes de aceite esencial mediante el uso del algoritmo de Kennard–Stone para dividir el conjunto de datos. Se seleccionaron tres descriptores DRAGON mediante la selección de predictores ordenados ( $R_{cal}^2 = 0.94$  y  $R_{val}^2 = 0.90$ ).

En un estudio reciente, Goudarzi et al. (Goudarzi *et al.*, 2014) predijeron los índices de retención de 83 hidrocarburos aromáticos policíclicos (PAHs) mediante el método de bosques al azar (RF) con 5 descriptores DRAGON ( $R_{cal}^2 = 0.99$ ), los cuales se usaron también para desarrollar una red neuronal artificial de retropropagación (BPANN) ( $R_{cal}^2 = 0.99$ ). Al mismo tiempo, Mohammadhosseini (Mohammadhosseini, 2014) calibró un modelo MLR para los índices de retención de 80 derivados terpénicos basado en la selección de descriptores de optimización por enjambre de partículas (PSO), el cual se usó para predecir los  $I$  de 28 compuestos ( $R_{cal}^2 = 0.94$  y  $R_{val}^2 = 0.90$ ).

En el trabajo que se presenta a continuación, se ha desarrollado un modelo QSPR para los índices de retención de 266 compuestos volátiles medidos en la columna HP5–MS en cromatografía de gases. El modelo desarrollado ha sido aplicado para la predicción del índice de retención para los tres compuestos que no presentaban dicha propiedad. Los resultados de este trabajo han sido recientemente aceptados para su publicación en una revista científica con referato (Rojas *et al.*, 2017).

#### 4.4.2. Materiales y métodos

Para los índices de retención medidos en la columna HP5–MS se siguió la misma metodología y diagrama de flujo usado para la base de datos DB–225MS (Figura 4.5). Los índices de los compuestos *Isobutanol*, *Acetato de etilo* y *Acetato de isopropilo* no fueron reportados, por lo que se usaron para predecir la propiedad experimental de los mismos. Por lo tanto, se utilizaron 266 compuestos para el desarrollo del modelo QSPR (Tabla 4A que se encuentra en el CD que acompaña a esta tesis).

**Tabla 4.11.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d        | $R^2_{cal}$ | $S_{cal}$   | $R^2_{val}$ | $S_{val}$   | $R^2_{ij\ max}$ | descriptores   |
|----------|-------------|-------------|-------------|-------------|-----------------|--|
| 1        | 0.91        | 98.7        | 0.93        | 90.6        | 0               | <i>SNar</i>  |
| 2        | 0.93        | 84.6        | 0.94        | 80.3        | 0.05            | <i>SpPos_A, MATS1p</i>                                       |
| <b>3</b> | <b>0.94</b> | <b>79.4</b> | <b>0.95</b> | <b>74.8</b> | <b>0.03</b>     | <b><i>SpPos_B(p), C-001, C-008</i></b>                       |
| 4        | 0.95        | 73.5        | 0.95        | 75.6        | 0.14            | <i>SpPos_A, SM6_B(s), MATS1p, nCrt</i>                       |
| 5        | 0.96        | 64.4        | 0.94        | 77.6        | 0.44            | <i>RFD, Ho_A, SM1_Dz(p), MATS1p, F04[C-O]</i>                |
| 6        | 0.97        | 61.7        | 0.95        | 76.2        | 0.87            | <i>Sp, AVS_B(p), P_VSA_m_1, Eig15_AEA(bo), nCs, B04[O-O]</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R^2_{ij\ max}$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

#### 4.4.3. Resultados y discusión

Inicialmente se aplicó el BSM para dividir la base de datos de 266 compuestos en un grupo de calibración de 88 moléculas y grupos de validación y predicción con 89 moléculas cada uno. Estas asignaciones se encuentran disponibles en la Tabla 4A. Seguidamente, el método de reemplazo permitió obtener modelos MLR a partir de 1819 descriptores independientes de la conformación. La Tabla 4.11 resume los mejores modelos QSPR de 1 a 6 descriptores, en los cuales no se aprecia sobreajuste. El modelo óptimo resulta aquel que tiene 3 descriptores:

$$I = 286.6 + 74.3 SpPos\_B(p) - 56.6C - 001 + 108.7C - 008 \quad (4.4)$$

$$N_{cal} = 88, R^2_{cal} = 0.94, S_{cal} = 76.9$$

$$N_{val} = 89, R^2_{val} = 0.95, S_{val} = 72.8$$

$$N_{pred} = 89, R^2_{pred} = 0.90, S_{pred} = 76.2$$

$$R^2_{loo} = 0.94, S_{loo} = 75.1, R^2_{lmo} = 0.96, S_{lmo} = 83.0$$

$$S_{rand} = 294.0, \alpha(3S) = 1, R^2_{ij\ max} = 0.03$$

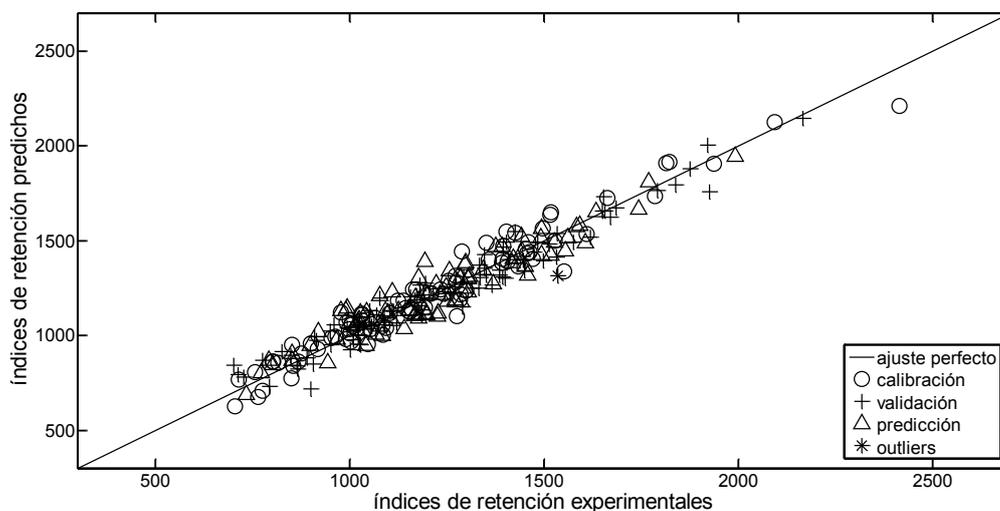
Los resultados de la Ec. 4.4 indican que se ha obtenido un modelo QSPR predictivo para los índices de retención medidos en la columna DB-225MS. Este hecho también se confirma con los criterios de validación que se indican a continuación:

$$R_{100}^2 > 0.5 \text{ (0.94)} \text{ y } R_{pred}^2 > 0.6 \text{ (0.90)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.00)} \text{ o } 1 - R_0'^2 / R_{pred}^2 < 0.1 \text{ (0.01)}$$

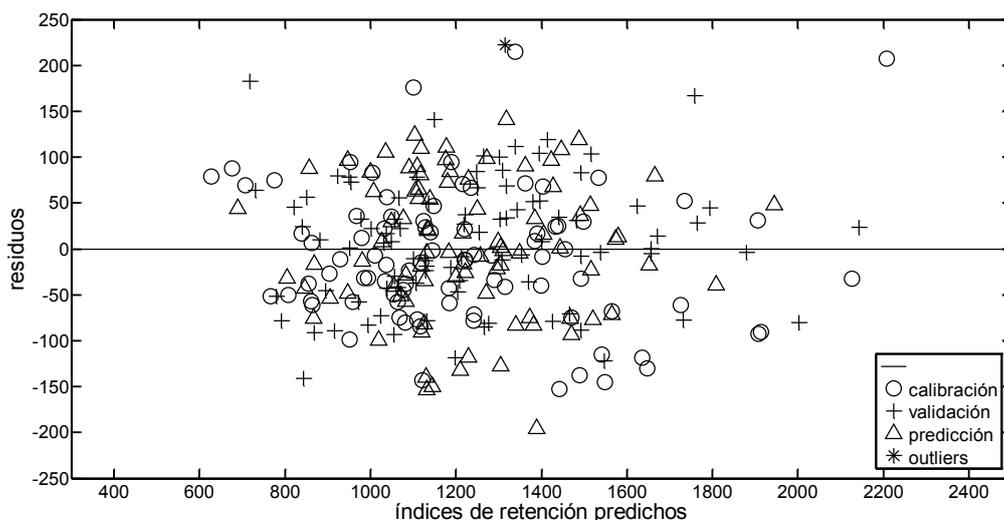
$$0.85 \leq k(1.00) \leq 1.15 \text{ y } 0.85 \leq k'(0.99) \leq 1.15$$

$$R_m^2 > 0.5 \text{ (0.89)}$$



**Figura 4.8.** Índices de retención experimentales y predichos para aromas medidos en la columna HP5-MS

Los valores de índices de retención experimentales en función de los predichos por el modelo (Figura 4.8) y el gráfico de los valores de los residuos (Figura 4.9) indican que un modelo MLR es apropiado para los  $I$ . Por otra parte, la única molécula que presenta valor de  $I$  mayor a 3 veces la desviación estándar es el *dihidroactinolida*.



**Figura 4.9.** Gráfico de dispersión para el modelo QSPR de la Ec. 4.4

Los descriptores involucrados en la Ec. 4.4. corresponden a los bloques de descriptores basados en la matriz 2D ( $SpPos_B(p)$ ) y fragmentos centrados en el átomo (C-001 y C-008). La baja correlación ( $R_{ij\max}^2 = 0.03$ ) entre los descriptores  $SpPos_B(p)$  y C-001 indica ausencia de multicolinealidad (Duchowicz *et al.*, 2010) y por consiguiente que los tres descriptores capturan aspectos particulares en el modelado del  $I$ . El grado de contribución relativo de cada descriptor indica que  $SpPos_B(p)$  (0.98) > C-001 (0.21) > C-008 (0.11).

La suma espectral positiva de la matriz de Burden ponderada por la polarizabilidad ( $SpPos_B(p)$ ) (Gutman *et al.*, 2003; Consonni & Todeschini, 2008) es un índice topológico calculado como la suma de los autovalores positivos (invariante del grafo) de la matriz de Burden que ha sido ponderada con la polarizabilidad. Valores grandes de esta suma indican la presencia de átomos con alta polarizabilidad y, por consiguiente, mayor habilidad de interactuar con la fase estacionaria mediante dipolos siendo retenida más tiempo. Como se presentó para la columna DB-225MS, los índices de retención se encuentran inversamente relacionados a la presencia de fragmentos  $CH_3$  (C-001) y relacionados sinérgicamente a la existencia de fragmentos CH enlazados a dos átomos de carbono y un átomo electronegativo (C-008) (Ghose *et al.*, 1998).

El análisis del dominio de aplicabilidad del modelo establece que las predicciones del  $I$  son confiables para los compuestos que presentan valor de influencia inferior al valor crítico ( $h^* = 0.068$ ). En este estudio no existen moléculas del grupo de predicción fuera del dominio de aplicabilidad, por lo tanto, el modelo fue capaz de predecir los índices de retención de todas las moléculas del grupo de predicción. Finalmente, el modelo QSPR de la Ec. 4.4 se usó para predecir el índice de retención de los tres compuestos para los cuales Yan *et al.* no presentaron los

valores experimentales: *Isobutanol* ( $I = 589.9$  y  $h_{ii} = 0.031$ ), *Acetato de etilo* ( $I = 681.0$  y  $h_{ii} = 0.024$ ) y *Acetato de isopropilo* ( $I = 831.5$  y  $h_{ii} = 0.076$ ). Se observa que la molécula *Acetato de isopropilo* presenta valor de influencia superior al valor crítico por lo que su predicción no sería confiable.

#### 4.4.4. Conclusiones

Se ha desarrollado un modelo predictivo para los índices de retención medidos en la fase estacionaria ligeramente polar HP5–MS. El modelo independiente de la conformación refleja que la suma espectral positiva de la matriz de Burden ponderada por la polarizabilidad y los dos fragmentos CH<sub>3</sub> y CH (enlazados a dos átomos de carbono y uno electronegativo) presentados son los apropiados para generar un modelo QSPR con buena capacidad predictiva.

### 4.5 Modelo QSPR para índices de retención de aromas medidos en la columna apolar HP–1

#### 4.5.1. Introducción

Algunos trabajos QSPR se han desarrollado para predecir los índices de retención medidos en la fase estacionaria apolar HP–1. En el año 2000, Héberger et al. (Héberger *et al.*, 2000) usaron cetonas y aldehídos alifáticos para desarrollar un modelo PLS2 con 4 LVs y 16 índices de retención ( $R_{cal}^2 = 0.99$  y  $R_{val}^2 = 0.97$ ). Al siguiente año, Körtvélyesi et al. (Körtvélyesi *et al.*, 2001) construyeron dos modelos QSPR usando la base de datos de Héberger (Héberger *et al.*, 2000). El primer modelo está conformado por dos descriptores cuánticos ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 12.04$ ) y el segundo por tres descriptores fisicoquímicos ( $R_{cal}^2 = 1.00$  y  $S_{cal} = 10.93$ ). En el año 2003, da Silva Junkes et al. (da Silva Junkes *et al.*, 2003) aplicaron un nuevo índice topológico semiempírico ( $I_{ET}$ ) a 632 índices de retención de compuestos orgánicos ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 17.71$ ); posteriormente recalibraron el modelo excluyendo 84 valores atípicos ( $R_{cal}^2 = 1.00$  y  $S_{cal} = 7.01$ ). En el mismo año, Ren (Ren, 2003) propuso un modelo QSPR para 33 compuestos (14 aldehídos y 19 cetonas) usando un índice modificado y tres tipos de átomos ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 7.73$ ). Un año más tarde, Junkes et al. (Junkes *et al.*, 2004) utilizó 31 compuestos de la base de datos de Héberger (Héberger *et al.*, 2000) para relacionar el índice de retención con el  $I_{ET}$  ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 5.47$ ).

Zhou y Nie (Zhou & Nie, 2007) en el año 2007 usaron la base de datos de Héberger (Héberger *et al.*, 2000) para proponer un modelo basado en un nuevo índice topológico y un número de trayecto como descriptores ( $R_{cal}^2 = 1.00$  y  $S_{cal} = 10.8$ ). Un año después, Konoz *et al.* (Konoz *et al.*, 2008), usando 140 compuestos de la misma base de datos, calibró un modelo MLR ( $R_{cal}^2 = 0.94$ ,  $S_{cal} = 47.4$ ,  $R_{val}^2 = 0.94$  y  $S_{val} = 50.5$ ) y un modelo ANN ( $R_{cal}^2 = 1.00$ ,  $S_{cal} = 8.0$ ,  $R_{val}^2 = 1.00$  y  $S_{val} = 10.5$ ). Finalmente, Souza *et al.* (Souza *et al.*, 2009) aplicaron el índice electrotopológico semiempírico  $I_{SET}$  para desarrollar modelos QSPR para 15 aldehídos ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 10.31$ ) y 42 cetonas ( $R_{cal}^2 = 0.99$  y  $S_{cal} = 11.72$ ) de la misma base de datos de Héberger (Héberger *et al.*, 2000).

Por consiguiente, el objetivo de este trabajo fue el desarrollo de un modelo QSPR para los índices de retención de 262 compuestos volátiles medidos en la columna apolar HP-1 en cromatografía de gases. Seguidamente, la relación cuantitativa ha sido aplicada para la predicción del  $I$  de 7 moléculas para las cuales esta propiedad no fue reportada. Los resultados de este trabajo han sido recientemente aceptados para su publicación en una revista científica con referato (Rojas *et al.*, 2017).

#### 4.5.2. Materiales y métodos

Para los índices de retención medidos en la fase estacionaria apolar HP-1 se siguió la misma metodología y diagrama de flujo previamente descrito para la base de datos DB-225MS y HP5-MS (Figura 4.5). En este caso, los  $I$  experimentales de los compuestos *Acrilato de etilo*, *Acetoína*, *Pentanal*, *Ácido propanoico*, *Isobutanol*, *Acetato de etilo* y *Acetato de isopropilo* no fueron reportados por Yan *et al.* Por lo tanto, la base de datos está conformada por 262 moléculas más los 7 compuestos que se usaron para predecir el índice de retención (Tabla 5A que se encuentra en el CD que acompaña a esta tesis).

**Tabla 4.12.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d | $R_{cal}^2$ | $S_{cal}$   | $R_{val}^2$ | $S_{val}$   | $R_{ij\ max}^2$ | descriptores  |
|---|-------------|-------------|-------------|-------------|-----------------|---|
| 1 | 0.94        | 84.1        | 0.92        | 78.1        | 0               | <i>SpPos_A</i>  |
| 2 | 0.94        | 80.9        | 0.94        | 65.7        | 0.01            | <i>MATS1p</i> , <i>Chi0_EA(ri)</i>  |
| 3 | <b>0.95</b> | <b>75.1</b> | <b>0.95</b> | <b>62.2</b> | <b>0.55</b>     | <b><i>SpPos_A</i>, <i>MATS1p</i>, <i>F02[C-C]</i></b>   |
| 4 | 0.96        | 68.8        | 0.95        | 62.4        | 0.11            | <i>Ho_A</i> , <i>MATS1p</i> , <i>nCt</i> , <i>Hy</i>  |
| 5 | 0.97        | 58.6        | 0.95        | 60.4        | 0.21            | <i>J_D</i> , <i>GATS1s</i> , <i>P_VSA_v_2</i> , <i>nCt</i> , <i>AMR</i>                                     |
| 6 | 0.97        | 57.4        | 0.95        | 63.6        | 0.68            | <i>ATS2s</i> , <i>P_VSA_m_1</i> , <i>nCs</i> ,<br><i>CATS2D_04_AA</i> , <i>CATS2D_06_LL</i> ,<br><i>AMR</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

### 4.5.3. Resultados y discusión

Los 262 compuestos medidos en la columna HP-1 se dividieron usando el BSM en un grupo de calibración de 88 fragancias y grupos de validación y predicción con 87 compuestos cada uno. Estas asignaciones están disponibles en la Tabla 5A. Seguidamente, se usaron los grupos de calibración y validación para la búsqueda de los mejores modelos de 1 a 6 descriptores que se detallan en la Tabla 4.12. Al igual que para las columnas DB-225MS y HP5-MS, no existe sobreajuste de los modelos. El modelo QSPR óptimo para estos datos es el que tiene 3 descriptores moleculares independientes de la conformación:

$$I = 296.1 + 153.7 \text{SpPos}_A + 543.4 \text{MATSp} - 9.5 \text{F0}[C - C] \quad (4.5)$$

$$N_{cal} = 88, R_{cal}^2 = 0.95, S_{cal} = 75.1$$

$$N_{val} = 87, R_{val}^2 = 0.95, S_{val} = 62.2$$

$$N_{pred} = 87, R_{pred}^2 = 0.92, S_{pred} = 65.1$$

$$R_{loo}^2 = 0.95, S_{loo} = 69.7, R_{lmo}^2 = 0.97, S_{lmo} = 74.4$$

$$S_{rand} = 282.4, \alpha(3S) = 2, R_{ij\max}^2 = 0.60$$

Los resultados del modelo desarrollado y los criterios adicionales de validación que se detallan a continuación, confirman que se ha obtenido un modelo QSPR predictivo:

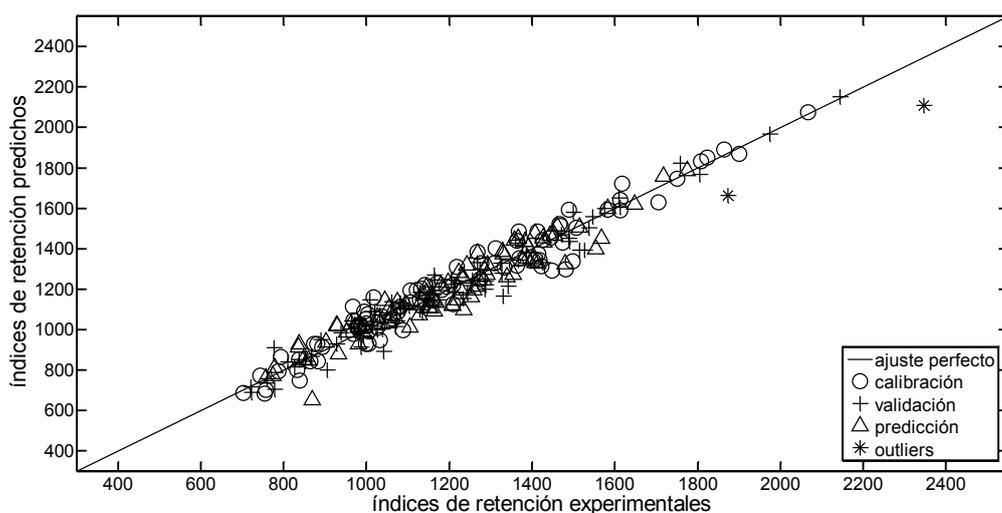
$$R_{loo}^2 > 0.5 \text{ (0.95)} \text{ y } R_{pred}^2 > 0.6 \text{ (0.92)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.00)} \text{ o } 1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.00)}$$

$$0.85 \leq k(1.00) \leq 1.15 \text{ y } 0.85 \leq k'(0.99) \leq 1.15$$

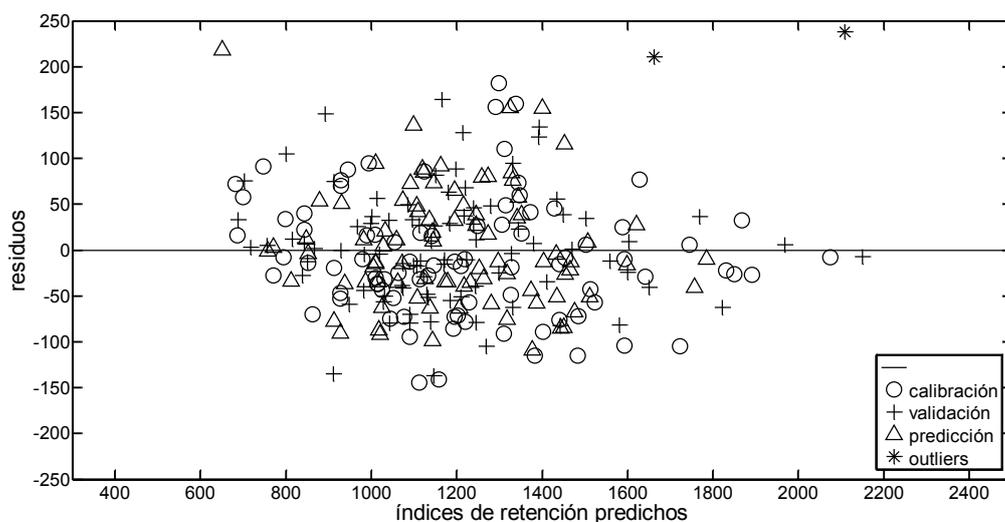
$$R_m^2 > 0.5 \text{ (0.89)}$$

En la Figura 4.10 se observa la tendencia lineal alrededor de la recta de ajuste perfecto entre los índices de retención experimentales y los predichos por el modelo, mientras que en la Figura 4.11 se confirma la distribución normal de los residuos. Las dos figuras afirman el supuesto de que los  $I$  se modelan apropiadamente por la Ec. 4.5.



**Figura 4.10.** Índices de retención experimentales y predichos para aromas medidos en la columna HP-1

En este modelo existen dos moléculas que presentan valores de residuo mayores a 3 desviaciones estándar:  $\delta$ -Tetradecalactona y Cinamil cinamato.



**Figura 4.11.** Gráfico de dispersión para el modelo QSPR de la Ec. 4.5

Por otra parte, la máxima correlación entre los descriptores  $SpPos\_A$  y  $F02[C-C]$  ( $R^2_{ij\max} = 0.60$ ) indica una correlación intermedia entre los mismos; mientras que la contribución en la predicción del  $l$  en la fase apolar muestra que  $SpPos\_A$  (1.06) >  $MATS1p$  (0.18) >  $MATS1v$  (0.20) >  $F02[C-C]$  (0.15).

La suma espectral positiva de la matriz de adyacencia ( $SpPos\_A$ ) (Consonni & Todeschini, 2008) es un índice topológico que se obtiene al sumar los autovalores positivos (invariante del grafo) de la matriz de

adyacencia, indicando de esta forma que el incremento en el  $I$  está relacionado con la complejidad de la molécula.

La autocorrelación de Moran a desplazamiento 1 ponderada por la polarizabilidad ( $MATS1p$ ) es un descriptor que se obtiene por aplicación del coeficiente de Moran (Moran, 1950) a un grafo molecular libre de hidrógenos y ponderado por la polarizabilidad atómica. Este descriptor provee información relacionada a la distribución de la polarizabilidad a lo largo de la estructura topológica de un compuesto. En otras palabras, altos valores de  $I$  están relacionados a autocorrelaciones espaciales positivas (la molécula tiene átomos con polarizabilidad similar a una distancia topológica de 1).

La frecuencia de enlaces C–C a distancia topológica 2 (Carhart *et al.*, 1985) pertenece a los pares de átomos 2D e indica que el índice de retención se encuentra inversamente correlacionado con la presencia de átomos de carbono separados por 2 unidades topológicas; es decir, que la propiedad experimental tiende a disminuir a medida que aumenta el tamaño de la cadena alifática en el compuesto.

El estudio de dominio de aplicabilidad indica que la única molécula que presenta un valor de influencia superior al  $h^* = 0.069$  es el  $\beta$ -Elemene ( $h_{ij} = 0.072$ ). Por otra parte, el modelo de la Ec. 4.5 se usó para predecir el índice de retención de los 7 compuestos para los cuales no se reportó el valor experimental: *Acrilato de etilo* ( $I = 700.9$  y  $h_{ij} = 0.048$ ), *Acetoína* ( $I = 691.4$  y  $h_{ij} = 0.004$ ), *Pentanal* ( $I = 725.2$  y  $h_{ij} = 0.007$ ), *Ácido propanoico* ( $I = 619.7$  y  $h_{ij} = 0.006$ ), *Isobutanol* ( $I = 611.5$  y  $h_{ij} = 0.004$ ), *Acetato de etilo* ( $I = 524.8$  y  $h_{ij} = 0.097$ ) e *Acetato de isopropilo* ( $I = 617.5$  y  $h_{ij} = 0.055$ ). La fragancia *Acetato de etilo* es el único compuesto que no pertenece al dominio del modelo, por lo que su predicción resulta de una extrapolación y por consiguiente, poco confiable.

#### 4.5.4. Conclusiones

El modelo QSPR desarrollado para los índices de retención medidos en la columna apolar HP–1 indica que son tres descriptores los indicados para generar un modelo estable y predictivo: suma espectral positiva de la matriz de adyacencia, autocorrelación de Moran a desplazamiento 1 ponderada por la polarizabilidad y frecuencia de enlaces carbono – carbono a distancia topológica 2. Este modelo apoya la conclusión de utilidad del uso de modelos independientes de la conformación para los  $I$  medidos en fases estacionarias apolares expuestos en el estudio en el que se usaron datos de la columna OV–101.

## 4.6 Referencias

- Acree, T., & Arn, H. (2004). Flavornet and Human Odor Space. *Available from: <http://www.flavornet>*, vol. 2015).
- Anker, L. S., Jurs, P. C., & Edwards, P. A. (1990). Quantitative structure-Retention Relationship Studies of Odor-Active Aliphatic Compounds with Oxygen-Containing Functional Groups. *Analytical Chemistry*, 62(24), 2676-2684.
- Babushok, V., Linstrom, P., Reed, J., Zenkevich, I., Brown, R., Mallard, W., & Stein, S. (2007). Development of a Database of Gas Chromatographic Retention Properties of Organic Compounds. *Journal of Chromatography A*, 1157(1), 414-421.
- Babushok, V., Linstrom, P., & Zenkevich, I. (2011). Retention Indices for Frequently Reported Compounds of Plant Essential oils. *Journal of Physical and Chemical Reference Data*, 40(4), 043101.
- Barwick, V. J. (1999). Sources of Uncertainty in Gas Chromatography and High-performance Liquid Chromatography. *Journal of Chromatography A*, 849(1), 13-33.
- Berthold, M. R., Cebren, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz Information Miner. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications*, (pp. 319-326): Springer.
- Bicchi, C., Rubiolo, P., Saranz Camargo, E. E., Vilegas, W., de Souza Gracioso, J., & Monteiro Souza Brito, A. R. (2003). Components of *Turnera Diffusa* Willd. var. *Afrodisiaca* (Ward) Urb. Essential Oil. *Flavour and fragrance journal*, 18(1), 59-61.
- Buydens, L., & Massart, D. L. (1981). Prediction of Gas Chromatography Retention Indexes from Linear Free Energy and Topological Parameters. *Analytical Chemistry*, 53(13), 1990-1993.
- Carhart, R. E., Smith, D. H., & Venkataraghavan, R. (1985). Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *Journal of chemical information and computer sciences*, 25(2), 64-73.
- Carlson, D. A., Bernier, U. R., & Sutton, B. D. (1998). Elution Patterns from Capillary GC for Methyl-branched Alkanes. *Journal of Chemical Ecology*, 24(11), 1845-1865.
- Ciubotariu, D., Medeleanu, M., Vlaia, V., Olariu, T., Ciubotariu, C., Dragos, D., & Corina, S. (2004). Molecular van der Waals Space and Topological Indices from the Distance Matrix. *Molecules*, 9(12), 1053-1078.
- Consonni, V., & Todeschini, R. (2008). New Spectral Indices for Molecule Description. *MATCH Communications in Mathematical and in Computer Chemistry*, 60, 3-14.
- Consonni, V., & Todeschini, R. (2012). Multivariate Analysis of Molecular Descriptors. In M. Dehmer, K. Varmuza & D. Bonchev (Eds.),

- Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, (pp. 111-147): Wiley-Blackwell.
- Cserhádi, T. (2010). *Chromatography of Aroma Compounds and Fragrances*: Springer.
- da Silva Junkes, B., Amboni, R. D. d. M. C., & Augusto, R. (2003). Semiempirical Topological Index: A Novel Molecular Descriptor for Quantitative Structure–Retention Relationship Studies. *Internet Electronic Journal of Molecular Design*, 2, 33-49.
- Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. (2005). A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. *Journal of chemical information and modeling*, 45(4), 839-849.
- Dimov, N., Osman, A., Mekenyan, O., & Papazova, D. (1994). Selection of Molecular Descriptors Used in Quantitative Structure-Gas Chromatographic Retention Relationships: I. Application to Alkylbenzenes and Naphthalenes. *Analytica Chimica Acta*, 298(3), 303-317.
- Duchowicz, P. R., Fernández, F. M., & Castro, E. A. (2010). Orthogonalization Methods in QSPR-QSAR Studies. In E. A. Castro (Ed.), *QSPR-QSAR Studies on Desired Properties for Drug Design*, (pp. 189-203 ): Research Signpost.
- Eckel, W. P., & Kind, T. (2003). Use of Boiling Point–Lee Retention Index Correlation for Rapid Review of Gas Chromatography–Mass Spectrometry Data. *Analytica Chimica Acta*, 494(1), 235-243.
- Edwards, P. A., Anker, L. S., & Jurs, P. C. (1991). Quantitative Structure–Property Relationship Studies of the Odor Threshold of Odor Active Compounds. *Chemical senses*, 16(5), 447-465.
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687-698.
- Garcia, J., Duchowicz, P. R., & Castro, E. A. (2016). Considering the Molecular Conformational Flexibility in QSAR Studies. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research: QSAR in Medicinal Chemistry*, (pp. 129-158): CRC Press.
- Gerasimenko, V. A., & Nabivach, V. M. (1990). Relationships between Gas Chromatographic Retention Indices and Molecular Structure of Aromatic Hydrocarbons. *Journal of Chromatography A*, 498(0), 357-366.
- Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1998). Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *The Journal of Physical Chemistry A*, 102(21), 3762-3772.

- Golbraikh, A., & Tropsha, A. (2002). Beware of  $q^2$ ! *Journal of Molecular Graphics and Modelling*, 20(4), 269-276.
- Goodner, K. L. (2008). Practical Retention Index Models of OV-101, DB-1, DB-5, and DB-Wax for Flavor and Fragrance Compounds. *LWT-Food Science and Technology*, 41(6), 951-958.
- Goudarzi, N., Shahsavani, D., Emadi-Gandaghi, F., & Chamjangali, M. A. (2014). Application of Random Forests Method to Predict the Retention Indices of Some Polycyclic Aromatic Hydrocarbons. *Journal of Chromatography A*, 1333, 25-31.
- Gramatica, P. (2007). Principles of QSAR Models Validation: Internal and External. *QSAR & Combinatorial Science*, 26, 694-701.
- Gutran, I., Vidović, D., Cmiljanović, N., Milosavljević, S., & Radenković, S. (2003). Graph Energy-A Useful Molecular Structure-Descriptor. *Indian Journal of Chemistry*, 42A, 1309-1311.
- Hall, L. H., & Kier, L. B. (1995). Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *Journal of chemical information and computer sciences*, 35(6), 1039-1045.
- Héberger, K., Görgényi, M., & Sjöström, M. (2000). Partial Least Squares Modeling of Retention Data of Oxo Compounds in Gas Chromatography. *Chromatographia*, 51(9-10), 595-600.
- Héberger, K. (2007). Quantitative Structure-(Chromatographic) Retention Relationships. *Journal of Chromatography A*, 1158(1-2), 273-305.
- Hoffmann, R., Minkin, V. I., & Carpenter, B. K. (1996). Ockham's Razor and Chemistry. *Bulletin de la Société chimique de France*, 133(2), 117-130.
- Hypercube Inc. HyperChem. <http://www.hyper.com>.
- Ivanciuc, O. (2000). QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs. *Journal of chemical information and computer sciences*, 40(6), 1412-1422.
- Jalali-Heravi, M., & Fatemi, M. (2001). Artificial Neural Network Modeling of Kovats Retention Indices for Noncyclic and Monocyclic Terpenes. *Journal of Chromatography A*, 915(1), 177-183.
- Jennings, W., & Shibamoto, T. (1980). *Qualitative Analysis of Flavor and Fragrance Volatiles by Glass Capillary Gas Chromatography*. London: Academic Press, Inc.
- Junkes, B. d. S., Amboni, R. D. d., Yunes, R. A., & Heinzen, V. E. (2004). Application of the Semi-Empirical Topological Index in Quantitative Structure-Chromatographic Retention Relationship (QSRR) Studies of Aliphatic Ketones and Aldehydes on Stationary Phases of Different Polarity. *Journal of the Brazilian Chemical Society*, 15(2), 183-189.
- Kaliszan, R. (1977). Correlation between the Retention Indices and the Connectivity Indices of Alcohols and Methyl Esters with Complex Cyclic Structure. *Chromatographia*, 10(9), 529-531.

- Kaliszan, R., & Foks, H. (1977). The Relationship between the  $R_M$  Values and the Connectivity Indices for Pyrazine Carbothioamide Derivatives. *Chromatographia*, 10(7), 346-349.
- Kaliszan, R. (2007). QSRR: Quantitative Structure-(Chromatographic) Retention Relationships. *Chemical reviews*, 107, 3212-3246.
- Katritzky, A. R., Chen, K., Maran, U., & Carlson, D. A. (2000). QSPR Correlation and Predictions of GC Retention Indexes for Methyl-Branched Hydrocarbons Produced by Insects. *Analytical Chemistry*, 72(1), 101-109.
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2015). PubChem Substance and Compound databases. *Nucleic acids research*, 44(D1), D1202-D1213.
- Kode srl. (2016). Dragon version 7. Software for Molecular Descriptor Calculation. <http://chm.kode-solutions.net/>.
- Konoz, E., Fatemi, M. H., & Faraji, R. (2008). Prediction of Kovats Retention Indices of Some Aliphatic Aldehydes and Ketones on Some Stationary Phases at Different Temperatures Using Artificial Neural Network. *Journal of Chromatographic Science*, 46(5), 406-412.
- Körtvélyesi, T., Görgényi, M., & Héberger, K. (2001). Correlation Between Retention Indices and Quantum-Chemical Descriptors of Ketones and Aldehydes on Stationary Phases of Different Polarity. *Analytica Chimica Acta*, 428(1), 73-82.
- Linstrom, P. J., & Mallard, W. G. (2001). NIST Chemistry WebBook, NIST Standard Reference Database Number 69. In). Gaithersburg MD: National Institute of Standards and Technology.
- Michotte, Y., & Massart, D. (1977). Molecular Connectivity and Retention Indexes. *Journal of pharmaceutical sciences*, 66(11), 1630-1632.
- Mohammadhosseini, M., Zamani, H. A., Akhlaghi, H., & Nekoei, M. (2011). Hydrodistilled Volatile Oil Constituents of the Aerial Parts of Prangos Serpentinica (Rech. f., Aell. Esfand.) Herznstadt and Heyn from Iran and Quantitative Structure-Retention Relationship Simulation. *Journal of Essential Oil Bearing Plants*, 14(5), 559-573.
- Mohammadhosseini, M. (2014). Prediction of the GC-MS Retention Indices for a Diverse Set of Terpenes as Constituent Components of Camu-camu (Myrciaria dubia (HBK) Mc Vaugh) Volatile Oil, Using Particle Swarm Optimization-Multiple Linear Regression (PSO-MLR). *Journal of Chemical Health Risks*, 4(1), 75-95.
- Moran, P. A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17-23.
- NCI/CADD Group. (2013). Chemical Identifier Resolver. In).
- Noorizadeh, H., & Noorizadeh, M. (2012). QSRR-based Estimation of the Retention Time of Opiate and Sedative Drugs by Comprehensive Two-Dimensional Gas Chromatography. *Medicinal Chemistry Research*, 21(8), 1997-2005.

- Pence, H. E., & Williams, A. (2010). ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education*, 87(11), 1123-1124.
- Porto, L. C., Souza, É. S., da Silva Junkes, B., Yunes, R. A., & Heinzen, V. E. F. (2008). Semi-Empirical Topological Index: Development of QSPR/QSRR and Optimization for Alkylbenzenes. *Talanta*, 76(2), 407-412.
- Qin, L.-T., Liu, S.-S., Liu, H.-L., & Tong, J. (2009). Comparative Multiple Quantitative Structure–Retention Relationships Modeling of Gas Chromatographic Retention Time of Essential Oils Using Multiple Linear Regression, Principal Component Regression, and Partial Least Squares Techniques. *Journal of Chromatography A*, 1216(27), 5302-5312.
- Qin, L.-T., Liu, S.-S., Chen, F., Xiao, Q.-F., & Wu, Q.-S. (2013). Chemometric Model for Predicting Retention Indices of Constituents of Essential Oils. *Chemosphere*, 90(2), 300-305.
- Ren, B. (2003). Atom-Type-Based AI Topological Descriptors for Quantitative Structure–Retention Index Correlations of Aldehydes and Ketones. *Chemometrics and Intelligent Laboratory Systems*, 66(1), 29-39.
- Riahi, S., Pourbasheer, E., Ganjali, M. R., & Norouzi, P. (2009). Investigation of Different Linear and Nonlinear Chemometric Methods for Modeling of Retention Index of Essential Oil Components: Concerns to Support Vector Machine. *Journal of Hazardous Materials*, 166(2–3), 853-859.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015a). QSPR Analysis for the Retention Index of Flavors and Fragrances on a OV-101 Column. *Chemometrics and Intelligent Laboratory Systems*, 140, 126-132.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015b). Quantitative Structure–Property Relationship Analysis for the Retention Index of Fragrance-Like Compounds on a Polar Stationary Phase. *Journal of Chromatography A*, 1422, 277-288.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2017). Quantitative Structure–Property Relationships for Predicting the Retention Indices of Fragrances on Stationary Phases of Different Polarity. *Journal of the Argentine Chemical Society, In Press*.
- Shiratsuchi, H., Shimoda, M., Imayoshi, K., Noda, K., & Osajima, Y. (1994). Off-Flavor Compounds in Spray-Dried Skim Milk Powder. *Journal of Agricultural and Food Chemistry*, 42(6), 1323-1327.
- Soják, L., Berezkin, V. G., & Janák, J. (1981). Effect of Adsorption on the Reproducibility of Retention Indices of Hydrocarbons in Capillary Gas-Liquid Chromatography. *Journal of Chromatography A*, 209(1), 15-20.
- Souza, É. S., Kuhnen, C. A., Junkes, B. d. S., Yunes, R. A., & Heinzen, V. E. F. (2009). Modeling the Semi-Empirical Electrotopological

- Index in QSPR Studies for Aldehydes and Ketones. *Journal of chemometrics*, 23(5), 229-235.
- Stanton, D. T., & Jurs, P. C. (1989). Computer-Assisted Prediction of Gas Chromatographic Retention Indexes of Pyrazines. *Analytical Chemistry*, 61(13), 1328-1332.
- Surburg, H., & Panten, J. (2016). *Common Fragrance and Flavor Materials: Preparation, Properties and Uses*: John Wiley & Sons.
- Sutter, J. M., Peterson, T. A., & Jurs, P. C. (1997). Prediction of Gas Chromatographic Retention Indices of Alkylbenzenes. *Analytica Chimica Acta*, 342(2-3), 113-122.
- Talete srl. (2015). Dragon version 6. Software for Molecular Descriptor Calculation. <http://www.talete.mi.it/>.
- The MathWorks Inc. MatLab. <http://www.mathworks.com>.
- Touhami, I., Mokrani, K., & Messadi, D. (2012). Modèles QSRR Hybrides Algorithme Génétique-Régression Linéaire Multiple des Indices de Rétenion de Pyrazines en Chromatographie gazeuse. *Lebanese Science Journal*, 13(1), 75-88.
- Wang, D., Xu, X., Chu, S., & Zhang, D. (2003). Analysis and Structure Prediction of Chlorinated Polycyclic Aromatic Hydrocarbons Released from Combustion of Polyvinylchloride. *Chemosphere*, 53(5), 495-503.
- Wittkowski, R., & Matissek, R. (1993). *Capillary Gas Chromatography in Food Control and Research*: Technomic Publishing Co., Inc.
- Woloszyn, T. F., & Jurs, P. C. (1993). Prediction of Gas Chromatographic Retention Data for Hydrocarbons from Naphthas. *Analytical Chemistry*, 65(5), 582-587.
- Yan, A., Jiao, G., Hu, Z., & Fan, B. T. (2000). Use of Artificial Neural Networks to Predict the Gas Chromatographic Retention Index Data of Alkylbenzenes on Carbowax-20M. *Computers & chemistry*, 24(2), 171-179.
- Yan, A., & Hu, Z. (2001). Linear and Non-Linear Modeling for the Investigation of Gas Chromatography Retention Indices of Alkylbenzenes on Cit.A-4, SE-30 and Carbowax 20M. *Analytica Chimica Acta*, 433(1), 145-154.
- Yan, J., Cao, D.-S., Guo, F.-Q., Zhang, L.-X., He, M., Huang, J.-H., Xu, Q.-S., & Liang, Y.-Z. (2012). Comparison of Quantitative Structure-Retention Relationship Models on Four Stationary Phases with Different Polarity for a Diverse Set of Flavor Compounds. *Journal of Chromatography A*, 1223, 118-125.
- Yan, J., Liu, X.-B., Zhu, W.-W., Zhong, X., Sun, Q., & Liang, Y.-Z. (2015). Retention Indices for Identification of Aroma Compounds by GC: Development and Application of a Retention Index Database. *Chromatographia*, 78(1-2), 89-108.
- Zefirov, N. S., & Palyulin, V. A. (2001). QSAR for Boiling Points of "Small" Sulfides. Are the "High-Quality Structure-Property-Activity

- Regressions" the Real High Quality QSAR Models? *Journal of chemical information and computer sciences*, 41(4), 1022-1027.
- Zhao, C., Liang, Y., Wang, X., Liu, L., Yuan, D., Wu, T., & Xu, Q. (2012). Modeling of Programmed-Temperature Retention Indices of a Diverse Set of Natural Compounds by Subspace Orthogonal Projection. *Current Analytical Chemistry*, 8(1), 168-179.
- Zhou, C., & Nie, C. (2007). Modeling Quantitative Structure Property Relationships with a Semi-Empirical Topological Descriptor and Path Numbers. *Chromatographia*, 66(7-8), 545-554.



# Capítulo 5

---

## APLICACIONES EN QUÍMICA DE LOS ALIMENTOS

---

En este Capítulo se presentarán las diversas aplicaciones desarrolladas en el área de la Química de los Alimentos. Se ha utilizado la teoría QSPR para desarrollar un modelo predictivo para el control de calidad de arroz basado en índices de retención en el espacio de cabeza en microextracción en fase sólida/cromatografía de gases/espectrometría de masas (SPME/GC/MS). Por otra parte, se ha estudiado exhaustivamente el dulzor de moléculas orgánicas, tanto para la predicción del dulzor relativo como para la discriminación entre el gusto dulce y no dulce (amargo e insípido). Los valores experimentales del dulzor se los puede considerar ruidosos, debido al hecho de que su medición está sometida a la evaluación de panelistas (entrenados o no entrenados), los cuales catalogan el tipo de gusto y, eventualmente, su intensidad. La percepción del gusto es un aspecto importante que ayuda a mejorar la calidad sensorial de alimentos y fármacos.

### 5.1 Modelo QSPR basado en índices de retención para el control de calidad de arroz

#### 5.1.1. Introducción

El arroz (*Oriza sativa* L.) es uno de los cereales mayormente producidos en el mundo y constituye un alimento básico en varios países (Fukuda *et al.*, 2014). Por este motivo, es necesario que se mejore el control de calidad de este cereal de tal forma que se aseguren características organolépticas óptimas para su aceptación por parte de los consumidores (Grimm *et al.*, 2002). El perfil aromático del arroz es producido por la presencia de diversos compuestos aromáticos, incluso en concentraciones bajas. El compuesto aromático principal del arroz es la 2-acetil-1-pirrolina (2-AP), la cual se genera durante el crecimiento de la planta. Sin embargo, durante la post-cosecha y almacenamiento, su concentración decrece (Grimm *et al.*, 2011).

El perfil aromático del arroz y el estudio de su calidad se han enfocado en la identificación de estos compuestos aromáticos. Con este propósito, la microextracción en fase sólida/cromatografía de gases/espectrometría de masas (SPME-GC-MS) ha demostrado ser un método eficiente de

análisis (Grimm *et al.*, 2002; Champagne, 2008). Compuestos volátiles con alto peso molecular son retenidos mejor en fibras con polidimetilsiloxano (PDMS), mientras que fases con Carboxen (CAR) o divinilbenceno (DVB) son más apropiadas para moléculas pequeñas. De esta forma, la fibra que combina DVB–CAR–PDMS ha sido probada como la más adecuada para analizar componentes complejos de los aromas y con diferente polaridad, así como para trabajar con altas temperaturas durante el análisis SPME–GC–MS (Grimm *et al.*, 2001; Bryant & McClung, 2011).

En el año 2014, Fatemi y Malekzadeh (Fatemi & Malekzadeh, 2014) desarrollaron un modelo QSPR para predecir el índice de retención de 96 compuestos volátiles identificados en tres variedades de arroz durante 4 etapas de cocción. La propiedad experimental fue medida en GC–MS mediante espacio de cabeza modificado en microextracción en fase sólida y usando la columna DB Wax. Ellos usaron la notación de cadena SMILES y una representación grafo–molecular para calcular descriptores moleculares en el programa CORAL. Se dividió la base de datos en grupos de calibración ( $n = 70$ ,  $R_{cal}^2 = 0.97$  y  $S_{cal} = 79.5$ ), validación ( $n = 13$ ,  $R_{val}^2 = 0.97$  y  $S_{val} = 125.6$ ) y predicción ( $n = 13$ ,  $R_{pred}^2 = 0.95$  y  $S_{pred} = 191.6$ ). También se analizó la validación cruzada de dejar–uno–fuera ( $R_{loo}^2 = 0.93$ ) y la aleatorización–Y ( $0.0003 \leq R^2 \leq 0.271$ ), así como otros criterios.

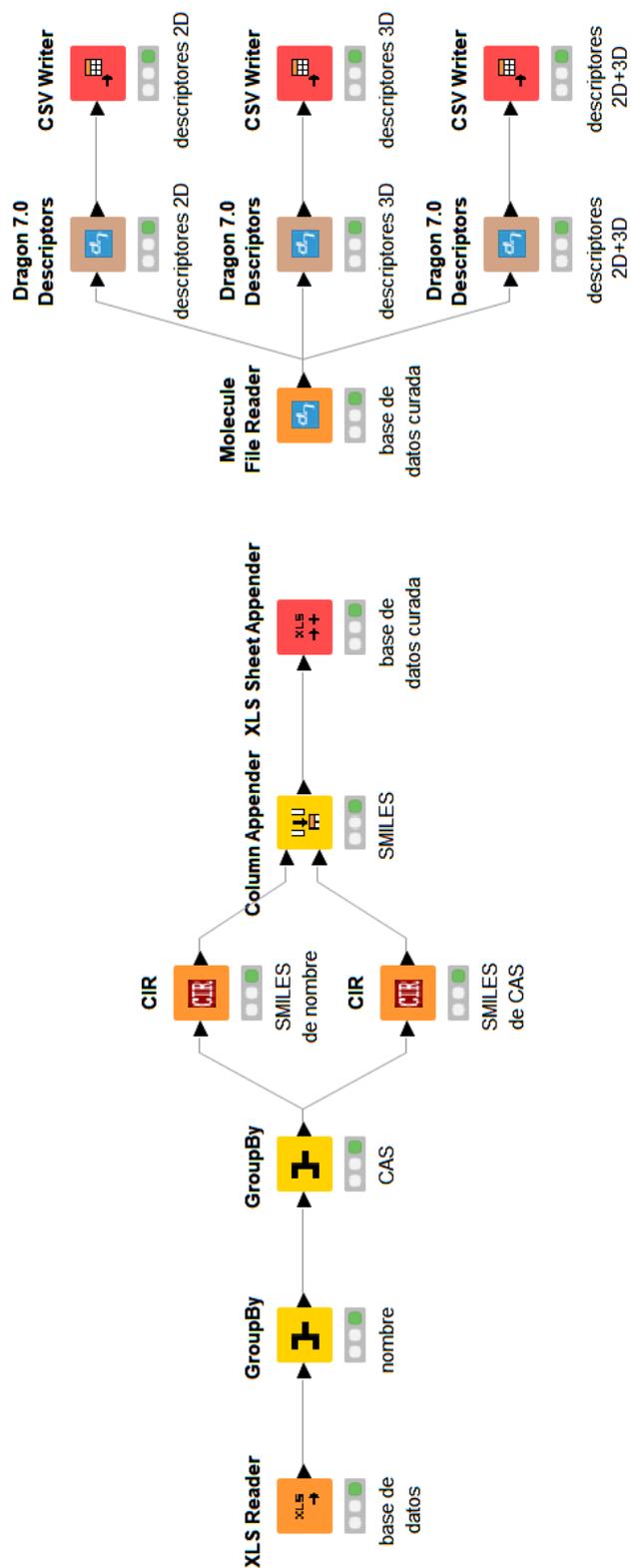
Por consiguiente, el objetivo de este trabajo fue el desarrollar una relación cuantitativa estructura–propiedad para los índices de retención de 137 compuestos orgánicos volátiles detectados en el espacio de cabeza de arroz en SPME–GC–MS. Posteriormente, el modelo ha sido aplicado para la predicción del I de 46 contaminantes que son frecuentemente reportados en muestras de arroz. De esta manera, se presenta la utilidad del modelo QSPR como herramienta para el control de calidad de arroz. Los resultados de este trabajo se encuentran actualmente en revisión en una revista científica con referato.

### 5.1.2. Materiales y métodos

Los índices de retención de 138 compuestos volátiles fueron tomados de Grimm *et al.* (Grimm *et al.*, 2002). La propiedad experimental corresponde al índice de retención medido en el espacio de cabeza en microextracción en SPME/GC/MS usando la fibra DVB–CAR–PDMS para trabajar a altas temperaturas.

Los nombres químicos, el número de registro CAS y los índices de retención experimental se exportaron al programa KNIME (Berthold *et al.*, 2008), para obtener notación de cadena SMILES de cada estructura mediante el nodo del solucionador de identificación química (CIR)

(NCI/CADD Group, 2013). Para cada molécula, se verificó que la notación SMILES coincidiera con el obtenido a partir del nombre o del CAS. De esta manera, aquellos compuestos que tienen notación SMILES diferente, fueron chequeados manualmente en las bibliotecas químicas PubChem (Kim *et al.*, 2015), ChemSpider (Pence & Williams, 2010) y NIST Chemistry WebBook (Linstrom & Mallard, 2001). Durante el curado de la base de datos, se identificó que el compuesto *2,2,4-trimetilheptano* (número CAS 14720-74-2) estaba duplicado como el compuesto *trimetilheptano*. Por lo tanto, este último se eliminó y se usó el / promedio para el *2,2,4-trimetilheptano*. De esta manera, la base de datos curada está constituida por 137 compuestos volátiles (Tabla 6A que se encuentra en el CD que acompaña a esta tesis.). El diagrama de flujo implementado en KNIME para el curado de la base de datos se muestra en la Figura 5.1.



**Figura 5.1.** Diagrama de flujo KNIME para el curado de la base de datos de índices de retención de aromas de arroz

Posteriormente, las geometrías de los compuestos fueron optimizadas utilizando el método MM+ de la mecánica molecular, y posteriormente refinadas mediante el método semiempírico PM7, implementado en el programa MOPAC (Stewart, 2016). Las geometrías moleculares se optimizaron hasta que el elemento máximo del vector gradiente de la energía total con respecto a las coordenadas atómicas sea menor que  $1 \text{ kcal} \times (\text{Å mol})^{-1}$ . A continuación, se calcularon 5239 descriptores moleculares usando el programa DRAGON versión 7 (Kode srl., 2016), con los cuales se construyeron tres conjuntos de descriptores: 1) todos los bloques de descriptores, 2) descriptores independientes de la conformación y 3) descriptores dependientes de la conformación (3D). En todos los casos, se han excluido descriptores no informativos; es decir, aquellos con valores constantes o casi constantes y descriptores con al menos un valor faltante.

Seguidamente, el método de reemplazo permitió explorar los tres conjuntos de descriptores para encontrar el modelo óptimo en cada caso. Para efectos de validación, la base de datos se dividió en conjuntos de calibración, validación y predicción mediante el BSM. El grupo de calibración se utilizó para la selección de descriptores mediante RM y el ajuste de los modelos, mientras que el grupo de validación se usó para evitar la presencia de sobreajuste durante el proceso de selección. El grupo de predicción se utilizó para medir la capacidad predictiva del modelo QSPR seleccionado. También se han usado las técnicas de validación cruzada de dejar–uno–fuera (loo) y dejar–varios–fuera (lmo) mediante la exclusión aleatoria del 20% de las moléculas y 50000 repeticiones. La ausencia de correlación casual se estableció mediante la aleatorización–Y con 10000 repeticiones. Adicionalmente, se analizaron los criterios de validación propuestos por Golbraikh y Tropsha (Golbraikh & Tropsha, 2002). Finalmente, se definió el dominio de aplicabilidad siguiendo el enfoque de influencia crítica ( $h^*$ ), así como la contribución de cada descriptor molecular y la interpretación del mecanismo de acción de los descriptores presentes en el modelo.

### 5.1.3. Resultados y discusión

Al igual que en los modelos presentados para las columnas OV–101 y Carbowax 20M en el capítulo 4, se buscó identificar si los descriptores conformacionales tienen importancia para la predicción del  $I$ . De esta manera, se evaluaron 3 conjuntos de descriptores: 1) 2898 descriptores de todos los tipos, 2) 1753 descriptores independientes de la conformación y 3) 1145 descriptores dependientes de la conformación.

El BSM particionó la base de datos de 137 compuestos volátiles en grupos de calibración y validación con 46 moléculas cada uno y un grupo de predicción con 45 compuestos. Estas asignaciones se encuentran disponibles en la Tabla 6A. En los tres conjuntos de

descriptores el RM se usó para seleccionar modelos de 1 a 6 descriptores.

**Tabla 5.1.** Mejores modelos QSPR obtenidos mediante RM usando todos los bloques de descriptores moleculares

| d | $R_{cal}^2$ | $S_{cal}$   | $R_{val}^2$ | $S_{val}$   | $R_{ij\ max}^2$ | descriptores   |
|---|-------------|-------------|-------------|-------------|-----------------|--|
| 1 | 0.95        | 96.1        | 0.94        | 98.6        | 0.00            | G2   |
| 2 | 0.97        | 83.3        | 0.95        | 90.2        | 0.04            | XMOD, AVS_B(e)                                       |
| 3 | <b>0.98</b> | <b>67.3</b> | <b>0.96</b> | <b>79.9</b> | <b>0.32</b>     | <b>X0Av, XMOD, MATS1p</b>                            |
| 4 | 0.98        | 58.8        | 0.96        | 82.1        | 0.66            | X1sol, Vindex, Mor12s, Hy                            |
| 5 | 0.99        | 48.2        | 0.95        | 88.7        | 0.77            | Psi_i_t, X1sol, VE2_Dz(e), MATS1v, Mor32e            |
| 6 | 0.99        | 39.9        | 0.94        | 94.7        | 0.62            | X1sol, Vindex, P_VSA_i_2, Mor13s, RTs+, CATS2D_02_NL |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\ max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

**Tabla 5.2.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d | $R_{cal}^2$ | $S_{cal}$   | $R_{val}^2$ | $S_{val}$   | $R_{ij\ max}^2$ | descriptores                                |
|---|-------------|-------------|-------------|-------------|-----------------|---|
| 1 | 0.95        | 97.9        | 0.94        | 99.8        | 0.00            | SpPos_B(m)                                  |
| 2 | 0.97        | 83.3        | 0.95        | 90.2        | 0.04            | XMOD, AVS_B(e)                              |
| 3 | <b>0.98</b> | <b>67.3</b> | <b>0.96</b> | <b>79.9</b> | <b>0.32</b>     | <b>X0Av, XMOD, MATS1p</b>                   |
| 4 | 0.98        | 60.7        | 0.96        | 78.2        | 0.18            | X1sol, SM3_B(s), VE2sign_B(s), MATS1v       |
| 5 | 0.99        | 55.4        | 0.96        | 81.5        | 0.80            | Psi_i_t, X1sol, Xindex, VE2_Dz(Z), MATS1p   |
| 6 | 0.99        | 49.9        | 0.96        | 78.7        | 0.54            | nR06, X1sol, Xindex, BIC0, ATSC1s, nFuranes |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\ max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

Los resultados obtenidos en los tres casos se encuentran en las Tablas 5.1, 5.2 y 5.3. La selección del modelo óptimo (se muestra en negrita) se realizó considerando los parámetros de calidad del grupo de calibración y validación (menor S y mayor  $R^2$ ), para evitar la presencia de sobreajuste en el modelo. Asimismo, se busca que el coeficiente de correlación máximo entre descriptores ( $R_{ij\ max}^2$ ) sea el más bajo posible, al igual que la dimensión del modelo sea apropiada, de acuerdo al principio de parsimonia de Ockham (Hoffmann *et al.*, 1996).

**Tabla 5.3.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores dependientes de la conformación.

| d | $R_{cal}^2$ | $S_{cal}$   | $R_{val}^2$ | $S_{val}$   | $R_{ij\ max}^2$ | descriptores              |
|---|-------------|-------------|-------------|-------------|-----------------|---------------------------|
| 1 | 0.95        | 96.1        | 0.94        | 98.6        | 0.00            | G2                        |
| 2 | 0.96        | 88.6        | 0.96        | 89.8        | 0.49            | G2, RDF060p               |
| 3 | <b>0.97</b> | <b>77.8</b> | <b>0.96</b> | <b>85.5</b> | <b>0.61</b>     | <b>G2, Mor26e, Mor22i</b> |
| 4 | 0.98        | 63.1        | 0.93        | 106.8       | 0.65            | G2, Mor32s, R4u+, G(O..O) |

|   |      |      |      |      |      |  |
|---|------|------|------|------|------|--|
| 5 | 0.98 | 61.0 | 0.94 | 98.0 | 0.97 | <i>G1, G2, SM2_G/D, Mor13s, RTs+</i>           |
| 6 | 0.99 | 54.6 | 0.94 | 96.7 | 0.61 | <i>G2, TDB05u, Mor32e, Mor27p, Mor13i, E3e</i> |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

En los tres casos, se observa la ausencia de sobreajuste para el modelo seleccionado. En efecto, cuando se consideran todos los bloques de descriptores (Tabla 5.1) y los descriptores independientes de la conformación (Tabla 5.2), el modelo óptimo coincide. Esto indica claramente que los descriptores conformacionales no aportan información significativa para predecir el índice de retención de compuestos volátiles del arroz. Asimismo, el mejor modelo dependiente de la conformación no refleja mejoras en la predicción de dicha propiedad. Por lo tanto, el modelo QSPR independiente de la conformación es el más apropiado, debido a que evita ambigüedades debidas a la optimización de la geometría molecular y los costos computacionales involucrados (García *et al.*, 2016):

$$I = 729.9 - 826.1XOA_v + 29.6XMOD + 492.2MATSp \quad (5.1)$$

$$N_{cal} = 46, R_{cal}^2 = 0.98, S_{cal} = 67.3$$

$$N_{val} = 46, R_{val}^2 = 0.97, S_{val} = 79.9$$

$$N_{pred} = 45, R_{pred}^2 = 0.97, S_{pred} = 80.0$$

$$R_{loo}^2 = 0.97, S_{loo} = 74, R_{lmo}^2 = 0.98, S_{lmo} = 86$$

$$S_{rand} = 353.8, \alpha(3S) = 1, R_{ij\max}^2 = 0.32$$

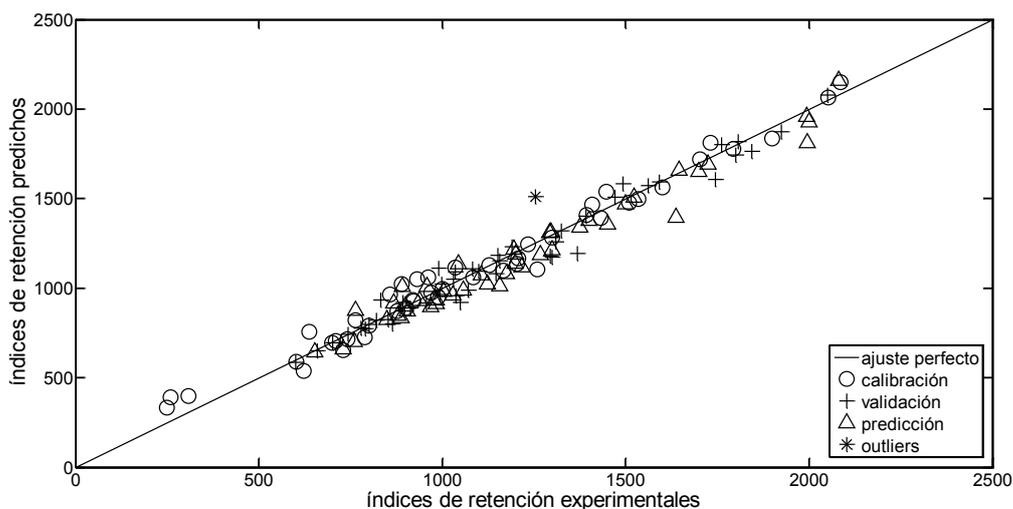
La validación cruzada de dejar–uno–fuera (loo) y dejar–varios–fuera (lmo), mediante la exclusión de una muestra a la vez y del 20% de moléculas, respectivamente, indica que el modelo es estable. Asimismo, la aleatorización–Y refleja que el  $S_{cal} < S_{rand}$ , indicando la ausencia de correlación casual en el modelo de la Ec. 5.1. Finalmente, los parámetros adicionales de validación que se presentan a continuación, indican que una relación cuantitativa estructura–propiedad predictiva se ha alcanzado:

$$R_{loo}^2 > 0.5 \text{ (0.97) y } R_{pred}^2 > 0.6 \text{ (0.97)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.000) o } 1 - R_0'^2 / R_{pred}^2 < 0.1 \text{ (0.000)}$$

$$0.85 \leq k(1.03) \leq 1.15 \text{ y } 0.85 \leq k'(0.97) \leq 1.15$$

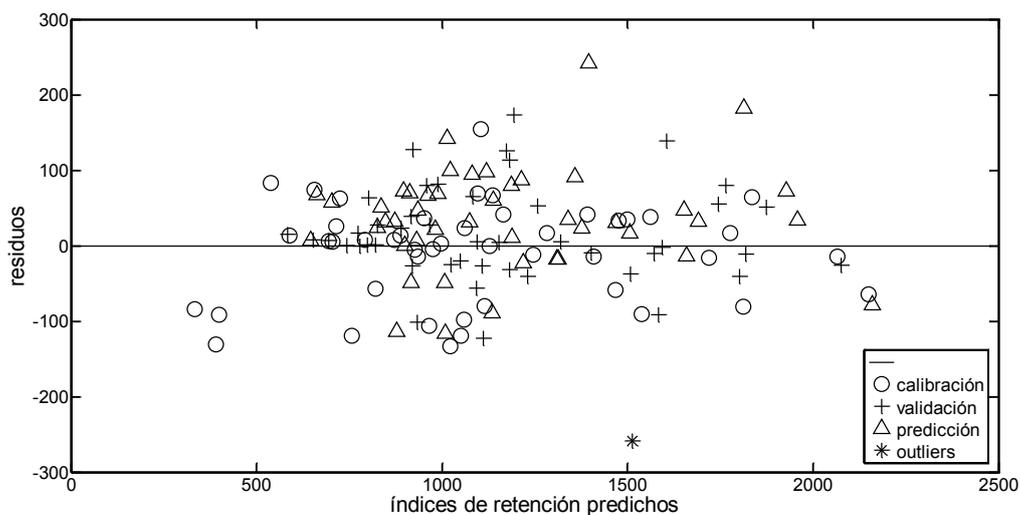
$$R_m^2 > 0.5 \text{ (0.95)}$$



**Figura 5.2.** Índices de retención experimentales y predichos para aromas medidos en el espacio de cabeza de arroz

La Figura 5.2 muestra la gráfica de los valores de índices de retención experimentales en función de los predichos por el modelo. Se observa la tendencia lineal alrededor de la recta de ajuste perfecto. Por otra parte, en la Figura 5.3 se muestra el gráfico de los valores de los residuos, el cual revela que dichos valores tienden a seguir un patrón aleatorio alrededor de la línea cero, sugiriendo que los índices de retención se modelan mediante un MLR.

El único compuesto que tiene valor de residuo mayor a 3 veces la desviación estándar es el *2-hexilo-1-octanol*. Se verificó que tanto su fórmula química como su índice de retención experimental sean correctos. Se asume su comportamiento atípico a la diversidad de compuestos químicos considerados en esta base de datos, así como a otros aspectos inherentes a la medición de la propiedad experimental (Rojas *et al.*, 2015).



**Figura 5.3.** Gráfico de dispersión del modelo QSPR de la Ec. 5.1

Los descriptores  $X0Av$  y  $MATS1p$  presentan la máxima correlación ( $R_{ij \max}^2 = 0.32$ ), indicando que no existe multicolinealidad en el modelo y por lo tanto, que cada descriptor explica aspectos particulares en el mecanismo de retención de los compuestos volátiles en la fibra DVB–CAR–PDMS (Duchowicz *et al.*, 2010). Así, la contribución de cada descriptor medida por los coeficientes de regresión estandarizados es:  $XMOD (0.95) > X0Av (0.19) > MATS1p (0.13)$ .

El índice modificado de Randić ( $XMOD$ ) es un índice de conectividad que se calcula mediante una fórmula tipo Randić sobre un grafo molecular libre de hidrógenos, el cual considera los electrones de valencia y la conectividad (Lohninger, 1993). El índice de conectividad de Randić mide el grado de ramificación y compactación de moléculas y se encuentra bien correlacionado con los tiempos de retención cromatográficos (Randić, 1975). De esta forma, moléculas con alto grado de ramificación (compuestos compactos) muestran valores altos de  $XMOD$  (efecto sinérgico). Esta relación ha sido descrita previamente por Yan *et al.* (Yan *et al.*, 2013).

Por otro lado, el índice de conectividad de valencia promedio de orden 0 ( $X0Av$ ) describe la presencia de heteroátomos en los compuestos; así como la presencia de enlaces dobles o triples. El efecto antagónico de este descriptor sobre la predicción del  $I$  ha sido descrito por Riahi *et al.* (Riahi *et al.*, 2008). Finalmente, la autocorrelación de Moran a desplazamiento 1 ponderado por la polarizabilidad ( $MATS1p$ ) es una autocorrelación 2D, la cual se calcula aplicando el coeficiente de Moran (Moran, 1950) a un grafo molecular ponderado por las polarizabilidades atómicas ( $p$ ). Este descriptor proporciona información acerca de la distribución de la polarizabilidad a lo largo de un compuesto aromático. En otras palabras, moléculas con altos índices de retención poseen valores positivos del coeficiente de Moran (autocorrelaciones espaciales positivas), lo cual indica que contienen átomos con polarizabilidad similar a desplazamiento topológico 1.

Las predicciones de la Ec. 5.1 están limitados a aquellos compuestos volátiles para los cuales el índice de retención esté dentro del rango de 250 a 2086 unidades y cuyo valor de influencia sea menor al valor crítico ( $h^* = 0.130$ ). El análisis del dominio de aplicabilidad (AD) indica que ninguna molécula del grupo de predicción posee valores superiores a dicho umbral, por lo que todas las predicciones son confiables.

**Tabla 5.4.** Contaminantes frecuentemente detectados en arroz: nombre químico,  $I$  predicho, valores de influencia y referencia

| Fuente      | nombre                    | $I_{pred}$ | $h_{ij}$           | Ref.                         |
|-------------|---------------------------|------------|--------------------|------------------------------|
| Desconocida | <i>Tetracloroetileno</i>  | 678.1      | 0.297 <sup>a</sup> | (Grimm <i>et al.</i> , 2002) |
|             | <i>Metoxi-fenil-oxima</i> | 1360.4     | 0.055              |                              |
|             | <i>1,4-diclorobenceno</i> | 1077.1     | 0.016              |                              |

|                               |                               |                   |                    |                                 |
|-------------------------------|-------------------------------|-------------------|--------------------|---------------------------------|
| Plasticantes                  | <i>Ftalato de dietilo</i>     | 1690.4            | 0.052              |                                 |
|                               | <i>Ftalato de dibutilo</i>    | 2044.1            | 0.080              |                                 |
|                               | <i>DEHP</i>                   | 2715.4            | 0.195 <sup>a</sup> |                                 |
| Antioxidante                  | <i>BHA</i>                    | 2443.0            | 0.129              |                                 |
| PAHs                          | <i>Acenaftileno</i>           | 1422.7            | 0.077              | (Liu & Korenaga, 2001;          |
|                               | <i>Fluoreno</i>               | 1502.0            | 0.071              | Tao <i>et al.</i> , 2006;       |
|                               | <i>Fenantreno</i>             | 1594.7            | 0.075              | Escarrone <i>et al.</i> , 2014) |
|                               | <i>Antraceno</i>              | 1591.7            | 0.075              |                                 |
|                               | <i>Fluoranteno</i>            | 1782.8            | 0.089              |                                 |
|                               | <i>Pireno</i>                 | 1779.8            | 0.089              |                                 |
|                               | <i>Benz[a]antraceno</i>       | 1949.0            | 0.096              |                                 |
|                               | <i>Criseno</i>                | 1952.0            | 0.096              |                                 |
|                               | <i>Perileno</i>               | 2137.4            | 0.117              | (Liu & Korenaga, 2001;          |
|                               |                               |                   |                    | Tao <i>et al.</i> , 2006)       |
|                               | <i>Benzo[b]fluoranteno</i>    | 2137.4            | 0.117              | (Tao <i>et al.</i> , 2006;      |
|                               | <i>Benzo[k]fluoranteno</i>    | 2134.4            | 0.117              | Escarrone <i>et al.</i> , 2014) |
|                               | <i>Benzo[a]pireno</i>         | 2134.4            | 0.117              |                                 |
|                               | <i>Indeno[1,2,3-cd]pireno</i> | 2316.6            | 0.140 <sup>a</sup> |                                 |
|                               | <i>Dibenz[a,h]antraceno</i>   | 2304.9            | 0.134 <sup>a</sup> |                                 |
|                               | <i>Benzo[g,h,i]perileno</i>   | 2316.6            | 0.140 <sup>a</sup> |                                 |
|                               | <i>Naftalina</i>              | 1230.6            | 0.067              |                                 |
| Fungicidas                    | <i>Validamicina A</i>         | 3549.9            | 0.439 <sup>a</sup> | (Berg, 2001)                    |
|                               | <i>Propiconazol</i>           | 2386.0            | 0.137 <sup>a</sup> |                                 |
|                               | <i>Hexaconazole</i>           | 2199.0            | 0.088              |                                 |
|                               | <i>Isoprotilano</i>           | 2063.1            | 0.070              |                                 |
|                               | <i>Iprodiona</i>              | 2252.1            | 0.109              |                                 |
| Herbicidas                    | <i>Ciproconazol</i>           | 2150.9            | 0.086              |                                 |
|                               | <i>2,4-D</i>                  | 1630.2            | 0.044              |                                 |
|                               | <i>Pretilaclor</i>            | 2171.3            | 0.100              |                                 |
|                               | <i>Fenclorim</i>              | 1667.4            | 0.036              |                                 |
|                               | <i>Fenoxaprop-p-etil</i>      | 2602.0            | 0.190 <sup>a</sup> |                                 |
|                               | <i>MCPA</i>                   | 1528.8            | 0.037              |                                 |
|                               | <i>Pirazosulfuron-etil</i>    | 2989.2            | 0.385 <sup>a</sup> |                                 |
|                               | <i>Butacloro</i>              | 2171.3            | 0.100              |                                 |
|                               | <i>Propanilo</i>              | 1555.4            | 0.027              |                                 |
|                               | Insecticidas                  | <i>Fenobucarb</i> | 1562.4             | 0.029                           |
| <i>Hidrocloruro de Cartap</i> |                               | 1753.8            | 0.058              |                                 |
| <i>λ-cialotrina</i>           |                               | 3129.0            | 0.291 <sup>a</sup> |                                 |
| <i>Deltametrina</i>           |                               | 3168.3            | 0.303 <sup>a</sup> |                                 |
| <i>Buprofezina</i>            |                               | 2135.8            | 0.084              |                                 |
| <i>Isoprocarb</i>             |                               | 1467.6            | 0.025              |                                 |
| <i>Alfamestrina</i>           |                               | 2905.8            | 0.229 <sup>a</sup> |                                 |
| <i>Fipronil</i>               |                               | 3014.9            | 0.262 <sup>a</sup> |                                 |
|                               | <i>Etofenprox</i>             | 2753.7            | 0.198 <sup>a</sup> |                                 |

<sup>a</sup> moléculas con valor de influencia superior a  $h^* = 0.130$

Durante el análisis de los compuestos volátiles de arroz, algunos contaminantes han sido detectados. Por ejemplo, plasticantes (ésteres del ácido ftálico) y antioxidantes (BHA), los cuales pudieron migrar de los materiales de empaque usados para el transporte del arroz (Grimm *et al.*, 2002). Otros contaminantes frecuentemente detectados durante el

análisis del perfil aromático del arroz son los hidrocarburos aromáticos policíclicos (PAHs) (Liu & Korenaga, 2001; Tao *et al.*, 2006; Escarrone *et al.*, 2014), que son compuestos químicos originados a partir de la combustión incompleta de combustibles fósiles. Los PAHs son contaminantes ambientales que se encuentran ampliamente distribuidos en el aire, suelo y agua, por lo que la exposición humana a dichos compuestos es inevitable, particularmente en la dieta alimentaria a base de cereales y vegetales (Phillips, 1999). Adicionalmente, se han considerado también pesticidas derivados del piretroide y carbamato (Berg, 2001) usados en el cultivo de arroz. De esta forma, se usó el modelo QSPR de la Ec. 5.1 para predecir el índice de retención de 46 contaminantes del arroz (Tabla 5.4).

Existen 14 compuestos con valor de influencia superior al umbral de corte de 0.130 y por tanto corresponden a compuestos que caen fuera del AD (extrapolaciones del modelo QSPR). Contrariamente, 32 contaminantes pertenecen al AD, por lo que sus índices de retención predichos son confiables. De esta manera, el modelo QSPR de la Ec. 5.1 se lo puede utilizar para identificar contaminantes presentes en el arroz mediante uso de la técnica GC acoplado con la fibra DVB-CAR-PDMS. Por ejemplo, si se identifica un  $I = 2137.4$ , se puede sospechar la presencia de perileno o Benzo[b]fluoranteno. Finalmente, este modelo puede ser de utilidad para investigadores que trabajan en cromatografía gaseosa como herramienta para la identificación rápida de compuestos volátiles con estructura molecular conocida y cuyo valor  $I$  no se conozca.

#### **5.1.4. Conclusiones**

Los índices de retención de compuestos volátiles medidos en el espacio de cabeza de arroz fueron descritos y predichos por un modelo QSPR de tres descriptores DRAGON independientes de la conformación. Este estudio confirma los resultados obtenidos para los índices de retención medidos en las columnas OV-101 y Carbowax 20M, es decir, modelos independientes de la conformación son apropiados para modelar y predecir esta propiedad. El modelo desarrollado en este estudio constituye una herramienta útil para las personas que trabajan en el control de calidad de arroz mediante la identificación de contaminantes en GC. En efecto, la versatilidad del modelo se ha usado para predecir los  $I$  de 46 contaminantes comunes del arroz.

### **5.2 Modelo QSPR para el dulzor relativo**

#### **5.2.1. Introducción**

El dulzor es uno de los gustos más importantes para los seres humanos y ha sido usado ampliamente en la industria de alimentos por producir una sensación agradable. Entre los diversos compuestos

dulcificantes, la sacarosa es el edulcorante usado con mayor frecuencia y es la sustancia estándar utilizada para la medición del dulzor relativo (RS) de otros edulcorantes (Singh *et al.*, 2014). El dulzor relativo se define como la relación de la concentración de sacarosa (estándar) con respecto a la concentración de otro edulcorante (Bassoli *et al.*, 2001). En otras palabras, una solución estándar de sacarosa tiene una percepción de dulzor de 1 (o 100) y el dulzor de un edulcorante determinado se compara con relación a la sacarosa.

El desarrollo y búsqueda de nuevos edulcorantes es complejo. Por un lado, existen múltiples factores que afectan el dulzor; por ejemplo, solubilidad, estabilidad en un amplio rango de pH y temperatura, dulzor puro sin retrogusto desagradable, beneficio dulzor/costo en comparación con la sacarosa y, el más importante, seguridad para la salud (Belitz *et al.*, 2009). Por otra parte, la medición del RS involucra altos costos, debido al hecho que se usan panelistas entrenados para la comparación del RS del edulcorante en estudio con respecto al estándar. Debido a todos estos factores, existen ventajas en el desarrollo de relaciones cuantitativas estructura–propiedad/actividad para el desarrollo y síntesis de nuevos y potentes edulcorantes (Yang *et al.*, 2011; Zhong *et al.*, 2013).

El primer estudio QSAR para modelar el dulzor relativo fue realizado por Deutsch y Hansch (Deutsch & Hansch, 1966), en el cual usaron derivados del *2-amino-4-nitrobenzoceno* y concluyeron que esta propiedad está muy relacionada con la hidrofobicidad y la constante de Hammett. A partir de este estudio, ha existido un creciente interés en el desarrollo de modelos QSAR/QSPR para modelar y predecir el RS de diversos edulcorantes (Rojas *et al.*, 2016a). Así, en el año 2002, Katritzky *et al.* (Katritzky *et al.*, 2002) estudiaron 238 compuestos constituidos por aldoximas, sulfamatos y acesulfamatos, ácidos cicloalcanóicos ( $\alpha$ -arilsulfonil) y edulcorantes naturales (azúcares, péptidos, guanidinas) para desarrollar un modelo MLR basado en cinco descriptores moleculares. Posteriormente, excluyeron los 87 péptidos para desarrollar otro modelo QSPR también formado por 5 descriptores moleculares.

Más adelante, Kelly *et al.* (Kelly *et al.*, 2005) sintetizaron 19 derivados monosustituidos del fenilsulfamato (ciclamos), los cuales fueron estudiados junto con otros 63 edulcorantes. El dulzor se categorizó en tres clases (dulce, no dulce y dulce/no dulce). Nueve descriptores del grupo de calibración (75 moléculas) se usaron para desarrollar modelos basados en DA, LDA y QDA. Para mejorar los resultados incorporaron el valor de dulzor de los panelistas y usaron el método de regresión CART (árboles de clasificación y regresión) con los seis descriptores más relevantes. El mismo año, Tarko *et al.* (Tarko *et al.*, 2005) trabajaron con una base de datos de 123 compuestos para desarrollar un modelo MLR basado en dos descriptores: masa molecular y el producto del porcentaje de oxígeno y de la máxima carga de los átomos de oxígeno.

Un año después, el mismo grupo (Tarko *et al.*, 2006) presentaron un nuevo estudio QSPR para modelar el dulzor de 136 derivados del ácido 3-aminosuccinámico basado en diversos fragmentos moleculares. En una primera etapa consideraron todas las moléculas y posteriormente desarrollaron un nuevo modelo QSPR al excluir 15 valores atípicos.

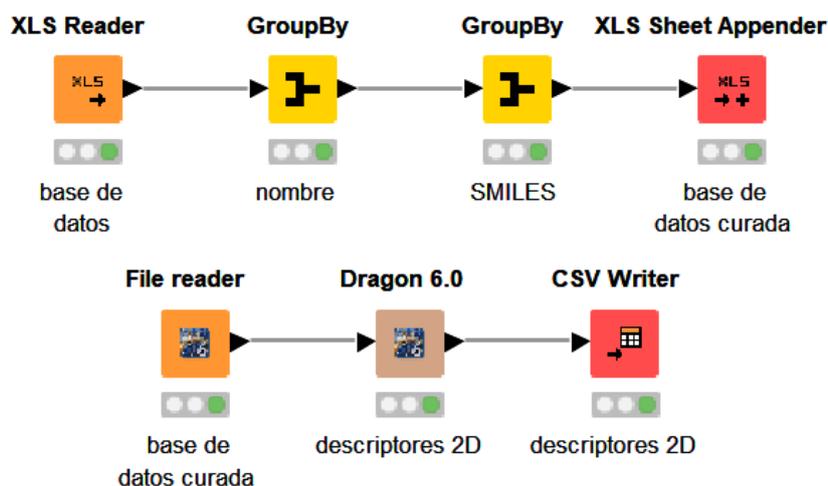
En el año 2007, Vepuri *et al.* (Vepuri *et al.*, 2007) usaron 53 moléculas de análogos del aspartamo para desarrollar un modelo 3D-QSAR basado en un algoritmo funcional genético (GFA) y 4 descriptores moleculares. Posteriormente desarrollaron modelos basados en CoMFA y CoMSIA. Como conclusión importante, los autores indicaron que las contribuciones de CoMFA y CoMSIA son las mismas. Posteriormente, Yang *et al.* (Yang *et al.*, 2011) utilizaron 103 edulcorantes para desarrollar 3 modelos QSPR: el primero basado en mínimos cuadrados ordinarios, el segundo usando una red neuronal artificial (ANN) y el tercero mediante aplicación de máquinas de soporte vectorial (SVM). Finalmente, Zhong *et al.* (Zhong *et al.*, 2013) ampliaron la base de datos de su estudio previo (Yang *et al.*, 2011) para formar una base de datos de 320 edulcorantes con la cual calcularon 1235 descriptores moleculares y 6776 combinaciones a partir de autocorrelaciones 2D. Como resultado de este estudio, 12 descriptores fueron seleccionados para desarrollar modelos basados en MLR y SVM.

El objetivo del presente estudio es el uso de una base de datos de 233 edulcorantes naturales y sintéticos para el desarrollo de una relación cuantitativa estructura-dulzor relativo. El modelo ha sido apropiadamente validado, lo cual permite su uso para la predicción del RS de nuevos potenciales edulcorantes. Este modelo QSPR ha sido previamente publicado (Rojas *et al.*, 2016b).

### 5.2.2. Materiales y métodos

Los datos experimentales del RS se tomaron de varios trabajos publicados. Para garantizar el desarrollo de un modelo QSPR confiable e independiente de la conformación, se ha realizado un curado de base de datos siguiendo los siguientes criterios programados en el programa KNIME (Berthold *et al.*, 2008):

1. Se han removido estructuras iónicas (sales). Esto debido al hecho de que el cálculo de descriptores moleculares no es posible para este tipo de compuestos.
2. También se removieron edulcorantes cuyo RS fue medido usando una solución de glucosa como estándar.
3. Cuando se encontraron estereoisómeros con una misma notación SMILES, solo uno de ellos fue retenido y se usó el valor RS promedio. Por ejemplo, D-ramnosa y L-ramnosa.



**Figura 5.4.** Diagrama de flujo KNIME para el curado de la base de datos de dulzor relativo

La Figura 5.4 muestra el diagrama de flujo KNIME utilizado para el curado. Por consiguiente, la base de datos curada está compuesta de 233 edulcorantes cuyo rango de variación del RS está entre 0.10 y 200000. Debido a este rango amplio, el RS fue transformado a una escala logarítmica,  $\log(\text{RS})$ , con lo que el rango de variación está comprendido entre  $-1.00$  y  $5.30$ . Detalles de esta base de datos se encuentran en la Tabla 7A, que se encuentra en el CD que acompaña a esta tesis. Finalmente, se usó la notación de cadena SMILES como representación molecular de cada compuesto.

A continuación, se calcularon 3736 descriptores moleculares independientes de la conformación usando el programa DRAGON versión 6 (Talete srl, 2015). Se excluyeron descriptores con valores constantes (1525), casi constantes (160) y al menos con un valor faltante (16). Así, 2062 descriptores fueron explorados con el RM para encontrar el modelo óptimo. Para efectos de validación, el BSM se usó para dividir la base de datos en conjuntos de calibración, validación y predicción. El grupo de calibración se usó para la selección de descriptores mediante RM y el ajuste de los modelos; mientras que el grupo de predicción se utilizó para medir la capacidad predictiva del modelo QSPR óptimo. También se han usado las técnicas de validación cruzada de dejar-uno-fuera (loo) y dejar-varios-fuera (lmo) mediante la exclusión aleatoria del 20% de las moléculas y 50000 repeticiones. Se utilizó la aleatorización-Y con 10000 repeticiones para asegurar la ausencia de correlación casual. Asimismo, se analizaron los criterios de validación propuestos por Golbraikh y Tropsha. Finalmente, se definió el dominio de aplicabilidad siguiendo el enfoque del influencia crítica ( $h^*$ ), así como la contribución de cada descriptor molecular y la interpretación del mecanismo de acción de los mismos.

**Tabla 5.5.** Mejores modelos QSPR obtenidos mediante RM usando los descriptores independientes de la conformación.

| d        | $R_{cal}^2$ | $S_{cal}$    | $R_{pred}^2$ | $S_{pred}$   | $R_{ij\max}^2$ | descriptores  |
|----------|-------------|--------------|--------------|--------------|----------------|---|
| 1        | 0.46        | 0.908        | 0.45         | 0.976        | 0.00           | MATS1s  |
| 2        | 0.59        | 0.793        | 0.50         | 0.921        | 0.10           | SM5_B(m), MATS1s  |
| 3        | 0.68        | 0.705        | 0.63         | 0.796        | 0.22           | MATS1e, P_VSA_v_3, B07[C-N]   |
| 4        | 0.73        | 0.652        | 0.62         | 0.816        | 0.99           | Wi_B(m), VE3_B(m), CATS2D_03_DD, B07[C-N]                                     |
| 5        | 0.76        | 0.609        | 0.54         | 0.917        | 0.28           | GATS1e, P_VSA_v_3 SssO, CATS2D_07_AA F04[C-N]                                 |
| <b>6</b> | <b>0.80</b> | <b>0.565</b> | <b>0.70</b>  | <b>0.737</b> | <b>0.79</b>    | <b>AAC, ATSC6p, CATS2D_02_PN, CATS2D_05_LL, B07[C-N], ALOGP</b>               |
| 7        | 0.82        | 0.540        | 0.70         | 0.746        | 0.79           | H%, ATSC6p, nCp, CATS2D_07_AA, CATS2D_05_LL, B07[C-N], ALOGP                  |
| 8        | 0.82        | 0.533        | 0.70         | 0.751        | 0.95           | Mv, X3sol, ATSC6p, nCp, CATS2D_02_PN, CATS2D_05_LL, B07[C-N], ALOGP           |
| 9        | 0.84        | 0.511        | 0.67         | 0.790        | 0.79           | Mv, ATSC6p, nCp, C-019, CATS2D_07_AA, CATS2D_05_LL, B07[C-N], F05[N-O], ALOGP |

d: número de descriptores;  $R^2$ : coeficiente de determinación;  $R_{ij\max}^2$ : coeficiente de determinación máxima entre descriptores; S: desviación estándar.

### 5.2.3. Resultados y discusión

El BSM dividió la base de datos de 233 edulcorantes en grupos de calibración ( $N_{cal} = 163$ ) y predicción ( $N_{pred} = 70$ ), de tal forma que se asegure un diseño con un balance de los compuestos en los dos grupos. Estas asignaciones se encuentran disponibles en la Tabla 7A. Posteriormente, el método de selección de variables RM se utilizó para desarrollar modelos MLR de 1 a 9 descriptores. La Tabla 5.5 resume los mejores modelos QSPR de 1 a 9 descriptores, en los cuales se observa la ausencia de sobreajuste. El modelo óptimo resulta aquel constituido por seis descriptores moleculares:

$$\begin{aligned} \log(RS) = & -2.216 + 2.002AAC + 0.129ATSC6p \\ & -0.912CATS2D_02_PN - 0.119CATS2D_05_LL \quad (5.2) \\ & + 1.136B07[C-N] + 0.405ALOGP \end{aligned}$$

$$N_{cal} = 163, R_{cal}^2 = 0.80, S_{cal} = 0.565$$

$$N_{val} = 70, R_{val}^2 = 0.70, S_{val} = 0.737$$

$$R_{loo}^2 = 0.78, S_{loo} = 0.586, R_{lmo}^2 = 0.75, S_{lmo} = 0.629$$

$$S_{rand} = 1.126, \alpha(3S) = 2, R_{ij\max}^2 = 0.78$$

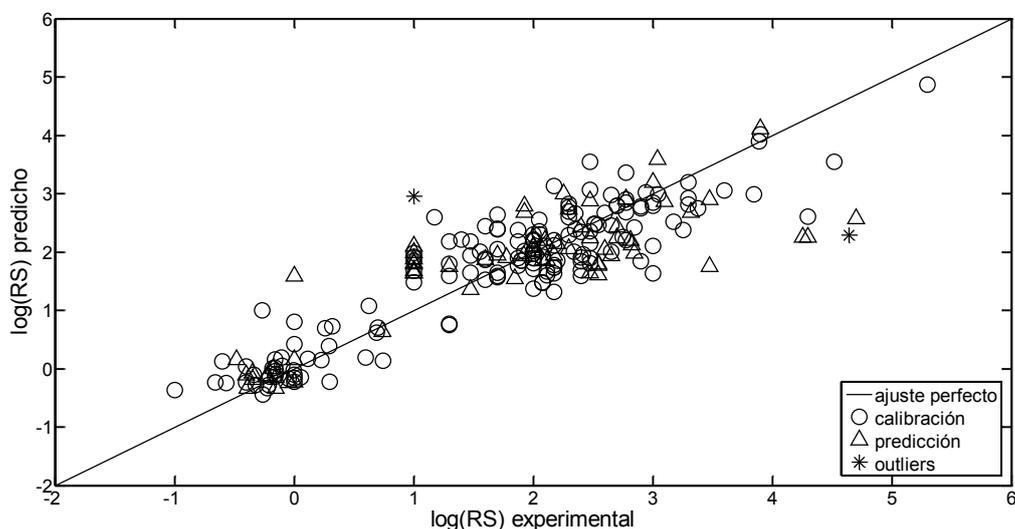
La validación cruzada de dejar–uno–fuera (loo) y dejar–varios–fuera (lmo), indican que el modelo de la Ec. 5.2 es estable. Asimismo, la aleatorización–Y indica que ausencia de correlación casual en el modelo QSPR ( $S_{cal} < S_{rand}$ ). Finalmente, los parámetros adicionales de validación que se presentan a continuación indican que la Ec. 5.2 constituye una relación cuantitativa estructura–propiedad predictiva:

$$R_{loo}^2 > 0.5 \text{ (0.78) y } R_{val}^2 > 0.6 \text{ (0.70)}$$

$$1 - R_0^2 / R_{pred}^2 < 0.1 \text{ (0.0001)}$$

$$0.85 \leq k(1.028) \leq 1.15 \text{ y } 0.85 \leq k'(0.874) \leq 1.15$$

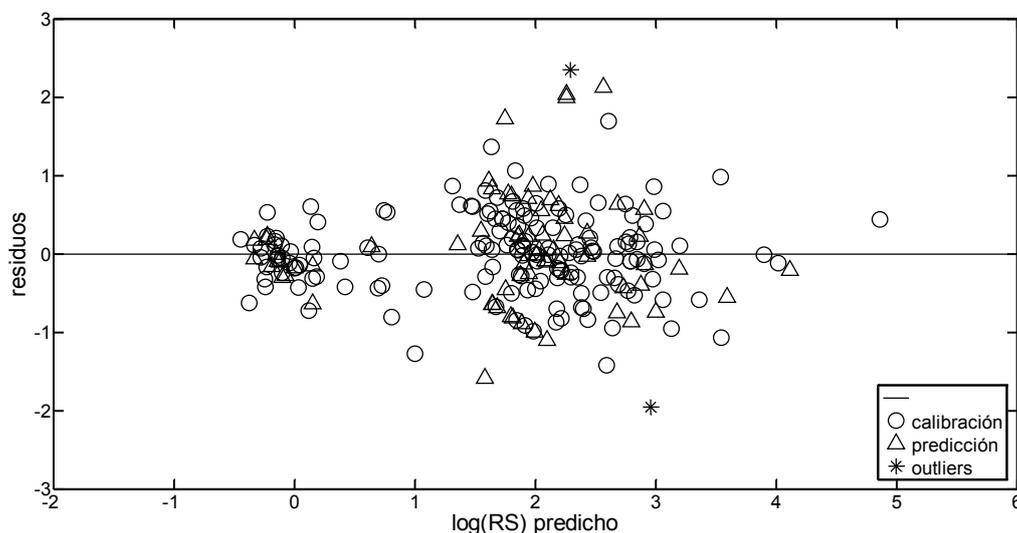
$$R_m^2 > 0.5 \text{ (0.69)}$$



**Figura 5.5.** Logaritmo del dulzor relativo experimental y predicho de diversos edulcorantes

La gráfica de los valores de log(RS) experimentales en función de los predichos por el modelo (Figura 5.5) y la gráfica de la distribución de los valores de los residuos (Figura 5.6) indican que el ajuste MLR es apropiado para predecir los valores del dulzor relativo. Existen dos compuestos con residuo mayor a 3S, particularmente los que tienen valores altos de RS: *compuesto LXXX* (–1.958) y *N–(L–aspartil)–1,1–diaminoalcano 5* (2.350). Al verificar la fórmula química y el dulzor en la fuente, se observó que estos compuestos son correctos. Se puede atribuir este comportamiento a la dificultad para modelar esta propiedad sensorial, debido a los errores humanos durante la determinación del dulzor relativo. De hecho, para su determinación se suelen usar panelistas entrenados o panelistas no entrenados, así como el uso de pocas personas dentro de un grupo de cata. Por otro lado, los humanos tienen limitaciones para discernir diferencias de dulzor debido a la pre–

saturación de las papilas gustativas presentes en la lengua (Birch & Mylvaganam, 1976). Además, existe una amplia gama de moléculas consideradas para el desarrollo del modelo QSPR que han sido sintetizadas y cuyos RS han sido medidos diversos grupos experimentales.



**Figura 5.6.** Gráfico de dispersión del modelo QSPR de la Ec. 5.2

La Ec. 5.2 muestra que cuatro descriptores tienen coeficientes con valor positivo, por lo que su efecto en la predicción del dulzor relativo es sinérgico. Estos descriptores son: *AAC*, *ATSC6p*, *B07[C-N]* y *ALOGP*. Por otro lado, los descriptores *CATS 2D* tienen efecto antagónico sobre la predicción del dulzor. Los descriptores *ATSC6p* y *CATS2D\_05\_LL* muestran una correlación moderada ( $R_{ij \max}^2 = 0.78$ ). El grado de contribución de los descriptores se obtiene mediante los coeficientes de regresión estandarizados: *ATSC6p* (1.264) > *CATS2D\_05\_LL* (0.881) > *ALOGP* (0.858) > *B07[C-N]* (0.339) > *AAC* (0.326) > *CATS2D\_02\_PN* (0.201).

El cálculo de las autocorrelaciones 2D se realiza mediante el uso de diferentes propiedades moleculares, de tal forma de describir los átomos (excepto H) presentes en la molécula a un determinado desplazamiento. Una de estas propiedades es la polarizabilidad atómica escalada con respecto al átomo de carbono (carbon-scaled), la cual se usa para ponderar el grafo molecular y calcular la autocorrelación centrada de Broto-Moreau a desplazamiento 6 (*ATSC6p*). El centrado se obtiene restando el valor promedio de la propiedad en la molécula. *ATSC6p* describe la forma en que esta propiedad se distribuye a lo largo de la estructura topológica (Todeschini & Consonni, 2009).

El descriptor pares de átomos, es un tipo simple de subestructura molecular definida en términos de datos binarios, ausencia/presencia, de un determinado par de átomos a una determinada distancia topológica

(Carhart *et al.*, 1985). Así, el RS se encuentra directamente relacionado a la presencia de pares de átomos de carbono y nitrógeno a distancia topológica 7 ( $C07[C-N]$ ). Por otra parte, AAC es el valor promedio del contenido de información total, cuyo cálculo involucra el número total de átomos, el número de átomos de un cierto tipo y la probabilidad de seleccionar de forma casual dicho tipo de átomo (Todeschini & Consonni, 2009).

Los descriptores CATS 2D son similares a los pares de átomos 2D, donde la mayor diferencia radica en la asignación de los átomos a puntos farmacóforos definidos (Fechner *et al.*, 2003). El punto lipofílico (L), el cargado positivamente o ionizable (P) y el cargado negativamente o ionizable (N) se usan para generar los pares de puntos farmacóforos (PPP) LL y PN. De esta forma, el dulzor relativo se encuentra inversamente relacionado a la presencia de pares LL y PN a distancia topológica 2 y 4, respectivamente.

El coeficiente de partición octanol-agua ( $\log P$ ) es una propiedad común en la química computacional y el modelado QSAR/QSPR. Un modelo conocido para el  $\log P$  fue propuesto por Ghose–Crippen ( $A\log P$ ) (Ghose *et al.*, 1988; Ghose *et al.*, 1998). Este descriptor calcula el  $\log P$  a partir de una ecuación de regresión basada en la contribución hidrofóbica de los diversos tipos de átomos.  $A\log P$  es intrínsecamente atomístico y fácil de calcular. Este concepto atomístico de hidrofobicidad (también conocido como lipofilidad) es muy útil en el diseño de nuevas moléculas. En efecto, la presencia de hidrofobicidad en un edulcorante permite una partición favorable de la sustancia entre el agua (saliva) y la membrana lipídica del receptor. Así, la hidrofobicidad de un compuesto dulce permite su difusión a través de la saliva para interactuar rápidamente con el receptor. Por lo tanto, el factor hidrofóbico gobierna la intensidad del dulzor de moléculas que tienen un apropiado glucóforo (Birch, 1987). En el modelo QSPR de la Ec. 5.2,  $A\log P$  indica que la hidrofobicidad de una molécula está directamente relacionado al dulzor relativo. Esta relación ha sido previamente presentada por Barker *et al.* y Vepuri *et al.*, 2007 (Barker *et al.*, 2002; Vepuri *et al.*, 2007). Adicionalmente, el uso del descriptor  $\log P$  para describir el dulzor relativo también ha sido ampliamente discutido en la literatura (Greenberg, 1980; Iwamura, 1980, 1981; Belitz *et al.*, 1990; Drew *et al.*, 1998; Jäger *et al.*, 2000; Spillane *et al.*, 2009; Briciu *et al.*, 2010).

El modelo desarrollado en este estudio está limitado a predecir el dulzor de moléculas cuyo  $\log(RS)$  se encuentre en el rango  $-1.00$  y  $5.30$  y el valor de influencia sea menor al valor crítico ( $h^* = 0.129$ ). El análisis del AD del modelo muestra que los compuestos *Periandrina III* ( $h_{ij} = 0.153$ ) y el éster fencílico del ácido aspártico ( $h_{ij} = 0.133$ ) son los únicos compuestos del grupo de predicción fuera del AD teórico del modelo.

**Tabla 5.6.** Comparación entre varios modelos QSAR/QSPR recientes para la predicción del dulzor relativo

| Referencia                       | Modelo              | <i>N</i> | <i>d</i> | $R^2_{cal}$ | $S_{cal}$ | $R^2_{pred}$   | $S_{pred}$ |
|----------------------------------|---------------------|----------|----------|-------------|-----------|----------------|------------|
| (Katritzky <i>et al.</i> , 2002) | MLR                 | 151      | 5        | 0.84        | 0.108     | — <sup>a</sup> | —          |
|                                  |                     | 238      | 5        | 0.69        | 0.098     | —              | —          |
| (Kelly <i>et al.</i> , 2005)     | CART                | 83       | 6        | 0.63        | —         | —              | —          |
| (Tarko <i>et al.</i> , 2005)     | MLR                 | 123      | 4        | —           | 0.485     | —              | 0.507      |
| (Tarko <i>et al.</i> , 2006)     | MLR                 | 136      | 4        | 0.66        | 0.652     | —              | —          |
|                                  |                     | 121      | 5        | 0.85        | 0.399     | —              | —          |
| (Vepuri <i>et al.</i> , 2007)    | MLR (3D–<br>GFA)    | 53       | 4        | 0.75        | 5.148     | 0.38           | —          |
|                                  | MLR (3D–<br>CoMFA)  |          | 3        | 0.77        |           | 0.54           | —          |
|                                  | MLR (3D–<br>CoMSIA) |          | 6        | 0.93        |           | 0.60           | —          |
| (Yang <i>et al.</i> , 2011)      | MLR                 | 103      | 3        | 0.89        | 1.023     | 0.86           | 1.205      |
|                                  | ANN                 |          |          | 0.90        | 1.049     | 0.87           | 1.162      |
|                                  | SVM                 |          |          | 0.91        | 1.037     | 0.89           | 1.192      |
| (Zhong <i>et al.</i> , 2013)     | MLR                 | 320      | 12       | 0.81        | 0.958     | 0.77           | 1.029      |
|                                  | SVM                 |          |          | 0.83        | 0.979     | 0.78           | 0.994      |
| Presente <sup>b</sup>            | MLR                 | 233      | 6        | 0.80        | 0.565     | 0.70           | 0.737      |

<sup>a</sup> no disponible; <sup>b</sup> (Rojas *et al.*, 2016b)

Por otra parte, la Tabla 5.6 resume los diversos modelos publicados para la predicción del dulzor relativo. El modelo de la Ec. 5.2 posee una cantidad de descriptores apropiada y su calidad en calibración y predicción es comparable a los modelos basados en regresión lineal múltiple y los métodos no lineales basados en la red neuronal artificial (ANN) y máquinas de soporte vectorial (SVM) (Yang *et al.*, 2011; Zhong *et al.*, 2013). Adicionalmente, la calidad predictiva demuestra ser superior con respecto a los modelos QSAR conformacionales que usan el algoritmo funcional genético (GFA), análisis CoMFA y CoMSIA (Vepuri *et al.*, 2007), los cuales muestran desviación estándar más alta con respecto a todos los demás modelos ( $S = 5.148$ ). Por otra parte, varios modelos QSPR no han sido sometidos a una validación externa (Katritzky *et al.*, 2002; Kelly *et al.*, 2005; Tarko *et al.*, 2006), lo cual limita su uso en la predicción del RS de nuevos compuestos y una comparación de la capacidad predictiva con respecto al modelo propuesto en este estudio.

#### 5.2.4. Conclusiones

En este estudio el RM permitió desarrollar un modelo predictivo para el dulzor relativo de 233 edulcorantes de amplia diversidad química. El modelo constituido por seis descriptores independientes de la conformación muestra buena calidad tanto en ajuste como en predicción. Es interesante señalar que es el primer modelo QSPR para la predicción del RS en el que se utilizan descriptores DRAGON y el RM como herramienta para selección de descriptores. Dada la heterogeneidad de los compuestos considerados en este estudio, el presente modelo podría ser utilizado para la predicción del RS de nuevos compuestos como

potenciales edulcorantes. Una vez más, el enfoque QSAR/QSPR independiente de la conformación permite obtener modelos con buenos resultados.

## **5.3 Modelo QSAR para discriminar los gustos dulce y amargo**

### **5.3.1. Introducción**

La química del gusto se ha convertido en un campo de investigación importante en varias disciplinas, particularmente la química de los alimentos. En efecto, existe un creciente interés en investigar la percepción del gusto, el cual se considera como el resultado de sustancias químicas solubles con diferentes propiedades osmóticas, endotérmicas y exotérmicas, que interactúan de diferentes maneras con las membranas biológicas de las papilas gustativas. Sin embargo, el mecanismo de cómo suceden estas reacciones no es bien conocido. Los gustos pueden separarse en cinco grupos básicos: dulce, amargo, salado, ácido y umami (Damodaran *et al.*, 2008). La percepción de estos gustos puede variar de persona a persona y puede estar relacionada a diferencias sutiles en la anatomía, psicología o la funcionalidad del receptor, concentración del compuesto generador del gusto o de interacciones con otras sustancias (Shallenberger, 1993). En el año 2002, Li *et al.* (Li *et al.*, 2002) describieron por primera vez el quimiorreceptor del gusto dulce, el cual es un receptor acoplado a proteínas G (GPCR) constituido por las subunidades T1R2 y T1R3. Este quimiorreceptor es capaz de reconocer estímulos dulces provenientes de diversos edulcorantes.

Entre los diversos gustos, el dulzor es considerado el más importante debido a que evoca una sensación placentera en una gran variedad de alimentos. Así, la sacarosa es el edulcorante utilizado como estándar internacional para medir el dulzor de otros compuestos, debido a que genera un gusto dulce "limpio" y sin retrogustos, incluso a altas concentraciones. Por otro lado, el amargor se percibe comúnmente como un gusto poco agradable, aunque en algunos productos alimenticios, tales como té, chocolate, café, cerveza, aceitunas y otros, es considerado placentero. El amargor es una característica propia de los alcaloides, entre los cuales la quinina es el compuesto mayormente usado como aditivo alimentario (Damodaran *et al.*, 2008).

Varios estudios QSAR han sido desarrollados para discriminar los gustos dulce y amargo. En el año 1980, Iwamura (Iwamura, 1980) propuso una relación cuantitativa para 49 perillartinas y derivados de la anilina mediante el uso de cinco descriptores STERIMOL que caracterizan el tamaño y forma molecular. Ese mismo año, Kier (Kier, 1980) usó 20 derivados dulces y amargos de aldoximas para desarrollar una función discriminante lineal basada en dos índices de conectividad

molecular. Entre 1982 y 1986, Takahashi et al. (Takahashi *et al.*, 1982; Takahashi *et al.*, 1984; Miyashita *et al.*, 1986b) trabajaron con compuestos dulces y amargos derivados de la perillartina y del dipéptido L-aspartil para desarrollar modelos QSAR basados en máquina de aprendizaje lineal (LLA), *k*NN, LDA y modelado suave independiente por analogía de clases (SIMCA). Adicionalmente, Spillane et al. (Spillane *et al.*, 1983; Drew *et al.*, 1998; Spillane *et al.*, 2002) trabajaron en la discriminación de derivados del sulfamato mediante el uso de modelos discriminantes: gráfico, DA, LDA y QDA.

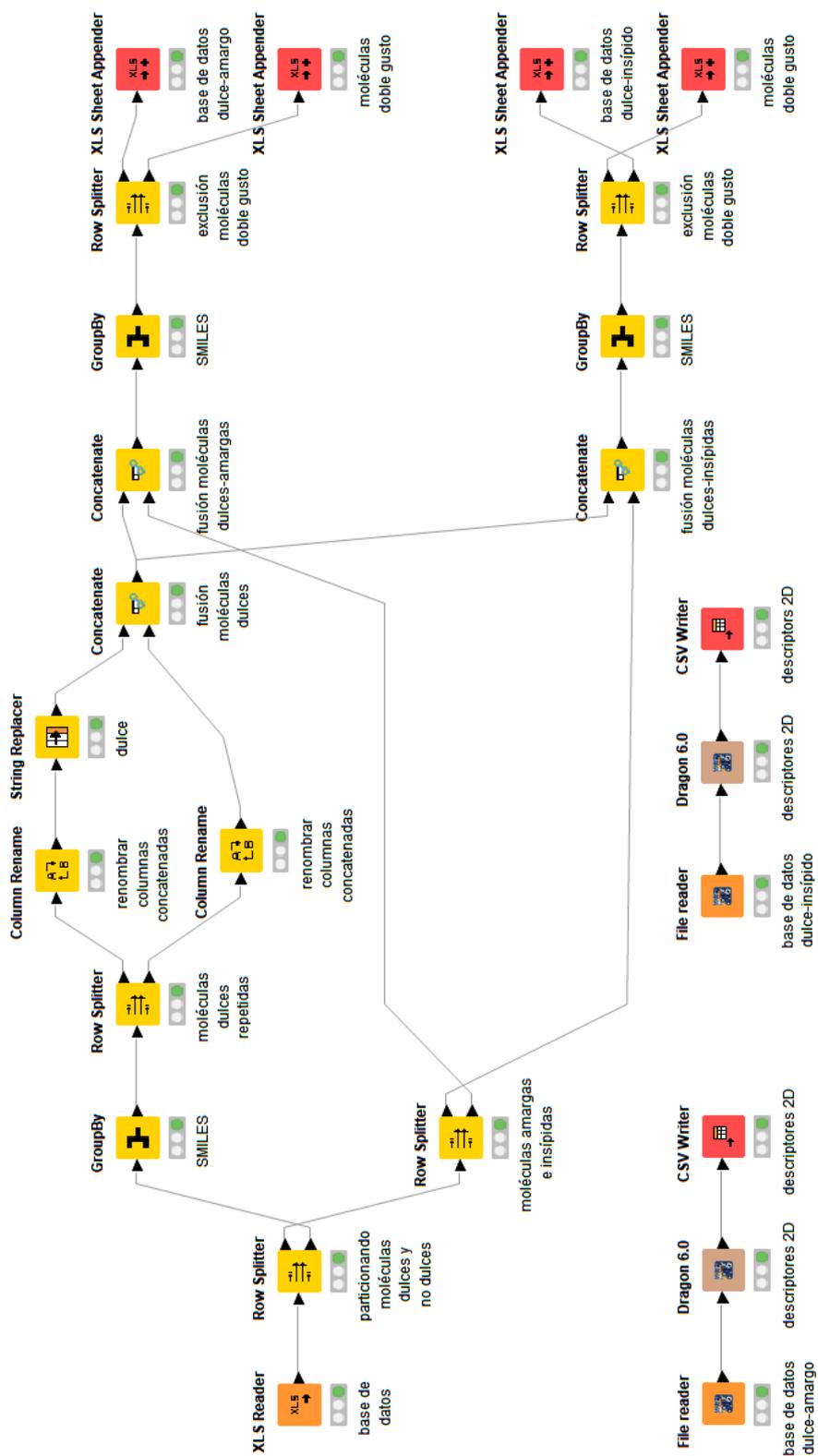
Con estos antecedentes, el objetivo del trabajo fue desarrollar un modelo QSAR para la discriminación de los compuestos dulces y amargos. Para este propósito, se ha puesto atención en el curado de la base de datos para posteriormente desarrollar un modelo de clasificación *k*NN. Los resultados de este trabajo han sido previamente publicados (Rojas *et al.*, 2016c).

### 5.3.2. Materiales y métodos

La base de datos inicial de compuestos dulces y amargos está constituida por 589 moléculas, para las cuales se encuentra disponible una respuesta cualitativa (dulce o amargo). Durante el curado de la base de datos se ha utilizado la notación de cadena SMILES y los siguientes criterios programados en KNIME (Figura 5.7):

1. Se han eliminado estructuras iónicas (sales).
2. Para estereoisómeros que muestran la misma clase, solo uno de ellos ha sido retenido.
3. Estereoisómeros que pertenecen a clases diferentes han sido excluidos.

De esta manera, la base de datos curada está constituida por 508 moléculas (427 dulces y 81 amargos), los cuales se encuentran detallados en la Tabla 8A, que se encuentra en el CD que acompaña a esta tesis. Posteriormente, se han calculado 3736 descriptores independientes de la conformación utilizando el nodo del programa DRAGON (Talete srl, 2015) implementado en KNIME.



**Figura 5.7.** Diagrama de flujo KNIME para el curado de la base de datos de compuestos dulces, amargos e insípidos

Debido a que los compuestos químicos presentan respuesta discreta (clases), se ha usado el método de clasificación no paramétrico *k*NN (Cover & Hart, 1967) para establecer la relación matemática entre la estructura química codificada en los descriptores y las clases modeladas. Este método es particularmente útil cuando la separación entre clases es no lineal. La regla de clasificación *k*NN es conceptualmente simple: un compuesto es clasificado en función de la mayoría de sus *k* vecinos más cercanos (distancia Euclidiana) en el espacio definido por los descriptores moleculares.

Para el desarrollo del modelo, inicialmente se ha aplicado una reducción no supervisada de descriptores basada en el método V–WSP (Ballabio *et al.*, 2014) con la finalidad de eliminar variables redundantes, con multicolinealidad y ruido. Para efectos de validación del modelo la base de datos se dividió en conjuntos de calibración (70%) y predicción (30%). Esta partición se realizó de forma casual y proporcional a la numerosidad de las clases, de tal forma de obtener similar representatividad en los dos grupos. Las moléculas del grupo de calibración se usaron para la selección de descriptores mediante algoritmos genéticos (GAs) (Leardi, 2009) acoplado con el método *k*NN. Durante la selección GAs se usó la validación cruzada basado en ventanas venecianas con 5 grupos y el valor óptimo de *k* se ha seleccionado en función de la mayor tasa de aciertos en esta etapa ( $NER_{cv}$ ). Adicionalmente, el modelo óptimo ha sido también evaluado mediante la especificidad ( $S_p$ ) y sensibilidad ( $S_n$ ) de la clase dulce.

Otro aspecto importante en los estudios QSAR es la interpretación del mecanismo de acción de los descriptores del modelo. Debido a que el método *k*NN no brinda coeficientes para cuantificar la contribución de cada descriptor, el mecanismo de acción ha sido evaluado mediante el análisis de componentes principales (PCA) (Jolliffe, 2002) del grupo de calibración. Se han usado los gráficos de puntuaciones y cargas para evaluar la relación entre los descriptores y las clases modeladas. Finalmente, el AD del modelo QSAR ha sido establecido mediante similitud *k*NN (Sahigara *et al.*, 2013).

### 5.3.3. Resultados y discusión

A partir de los 3763 descriptores calculados, se retuvieron 2164 luego de la exclusión de valores constantes, casi constantes y al menos un valor faltante. Seguidamente, el método de reducción V–WSP permitió obtener una base de datos de 855 descriptores mediante la eliminación de aquellos correlacionados a un umbral de 0.95. La base de datos fue dividida en un grupo de calibración con 356 moléculas y un grupo de predicción con los restantes 152 compuestos.

La selección de descriptores mediante los GAs acoplados con el método *k*NN se desarrolló en dos etapas: se aplicó GAs separadamente sobre cada bloque de descriptores; entonces, los descriptores

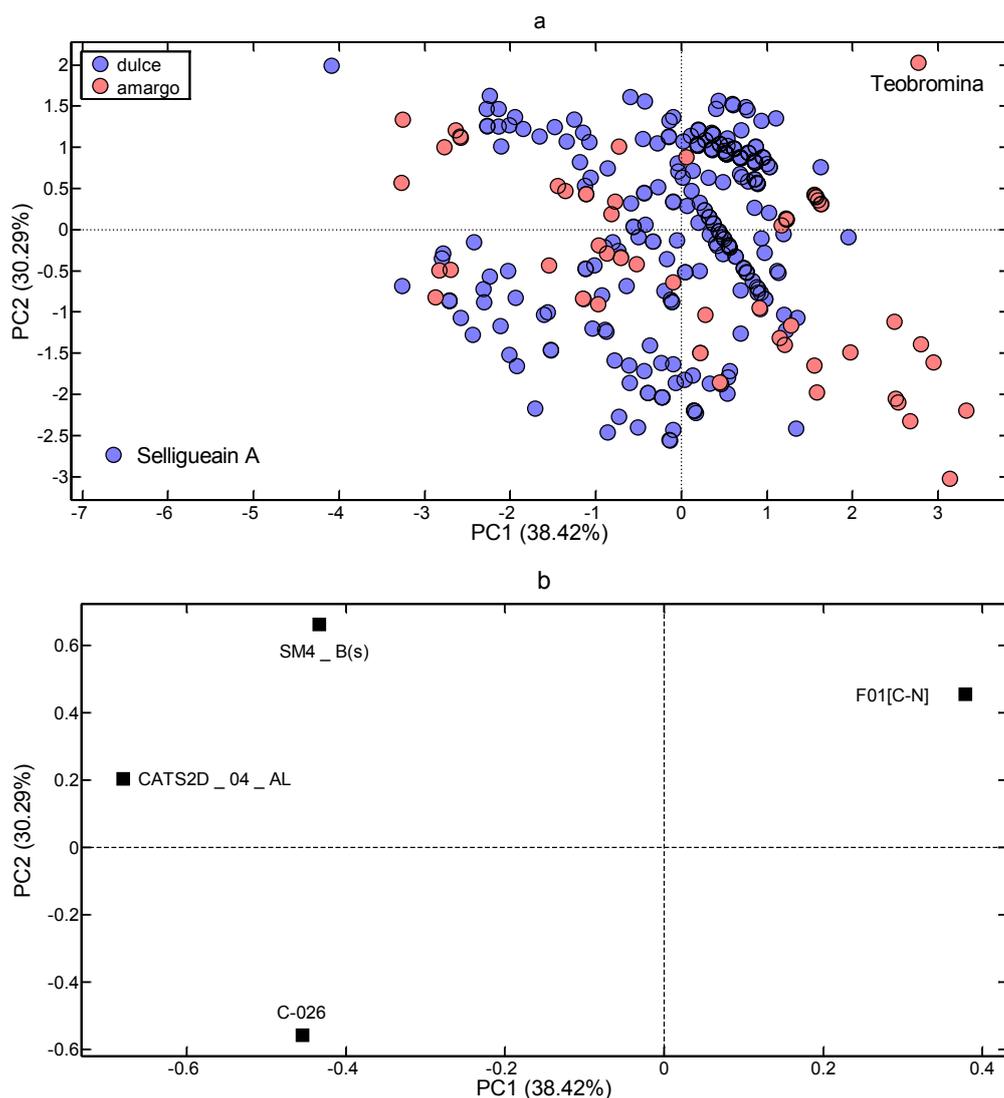
seleccionados en cada bloque se fusionaron para aplicar nuevamente GAs y encontrar el modelo QSAR óptimo. La selección del modelo final se ha alcanzado considerando un balance entre el NER, Sn y Sp de la clase dulce. De esta manera se obtuvo un modelo conformado por 4 descriptores moleculares. En la Tabla 5.7 se muestran los parámetros de calidad del modelo.

**Tabla 5.7.** Parámetros de calidad del modelo QSAR para discriminar los gustos dulce y amargo

|                               | NER   | Sn    | Sp    |
|-------------------------------|-------|-------|-------|
| Calibración                   | 0.864 | 0.957 | 0.772 |
| Validación cruzada (5 grupos) | 0.861 | 0.950 | 0.772 |
| Predicción                    | 0.789 | 0.953 | 0.625 |

El modelo muestra calidad comparable en calibración, validación y predicción, lo que indica que el modelo no sufre sobreajuste. Posteriormente, para evaluar el mecanismo de acción de los descriptores en la discriminación de las dos clases se utilizó el análisis de componentes principales. La combinación de las dos primeras componentes, PC1 y PC2, explica el 69% de la información total. La Figura 5.8a muestra el gráfico de puntuaciones (proyección de los compuestos), mientras que en la Figura 5.8b se ve el gráfico de cargas (proyección de los descriptores).

La mayoría de los compuestos amargos tienen puntuaciones positivas en la PC1 y simultáneamente negativa en la PC2 (cuarto cuadrante); es decir, estos compuestos son fuertemente caracterizados por la presencia de pares de átomos de carbono y nitrógeno separados por 1 distancia topológica (F01[C-N]) (Carhart *et al.*, 1985), así como por la presencia de átomos de carbono enlazados a un átomo electronegativo y cualquier otro grupo a través de dos enlaces aromáticos (C-026) (Ghose *et al.*, 1998). Por otra parte, el alcaloide teobromina se encuentra aislado en el primer cuadrante, es decir, la región de puntuaciones positivas de las dos componentes (dato atípico), indicando que este compuesto es descrito por valores altos de F01[C-N] y SM4\_B(s). Para el cálculo del descriptor SM4\_B(s) se aplica el momento espectral de orden 4 a la matriz de Burden ponderada por el estado intrínseco (s) (Kode srl., 2016). Este descriptor tiene buena relación con el número de átomos (excepto H) de una molécula. De esta forma, valores altos de SM4\_B(s) se encuentran relacionados con moléculas dulces grandes (por ejemplo, mogrosido V, rebaudioside D, mogrosido IV), mientras que valores bajos de dicho descriptor representan mejor a moléculas amargas pequeñas (por ejemplo, pirrolidina, piperazina, pirrol).



**Figura 5.8.** Análisis de componentes principales para el modelo *k*NN para discriminar moléculas dulces y amargas: a) gráfico de puntuaciones y b) gráfico de cargas.

Un gran número de moléculas dulces poseen puntuaciones negativas en la PC1 (cuadrantes dos y tres), indicando que las mismas son descritas por los descriptores CATS2D\_04\_AL, C-026 y SM4\_B(s). Así, el dulzor está descrito por la presencia de fragmentos lineales de átomos aceptores de enlaces de hidrógeno y lipofílicos, es decir, está estrechamente relacionado a la presencia de átomos de N, O y S con al menos un par de electrones libres y átomos lipofílicos separados por 4 unidades topológicas (Fechner *et al.*, 2003). Adicionalmente, los descriptores C-026 y SM4\_B(s) tienen un comportamiento particular, debido a que se ubican en la zona negativa y positiva de la PC2, respectivamente. De esta manera, estos dos descriptores ofrecen información opuesta entre ellos; es decir, moléculas dulces con valores

altos de SM4\_B(s) tienen baja presencia de átomos de carbono enlazados a un átomo electronegativo (C-026). Otro edulcorante atípico es la selligueain A, una proantocianidina trimérica, la cual muestra valores altos de los descriptores C-026 y CATS2D\_04\_AL. En efecto, este edulcorante se caracteriza por tener 6 anillos aromáticos en los cuales 12 átomos de carbono están enlazados a átomos de oxígeno.

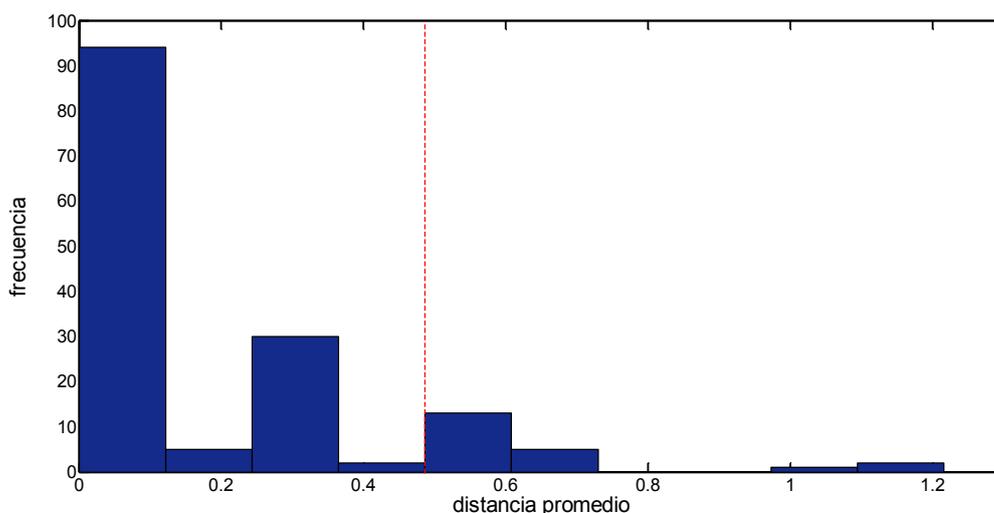
**Tabla 5.8.** Modelos QSAR para discriminar los gustos dulce y amargo.

| Modelo                            | No. Clases | Método    | d               | moléculas |                | NER   |       |
|-----------------------------------|------------|-----------|-----------------|-----------|----------------|-------|-------|
|                                   |            |           |                 | cal       | pred           | cal   | pred  |
| (Iwamura, 1980)                   | 2          | Regresión | 3               | 49        | — <sup>a</sup> | —     | —     |
| (Kier, 1980)                      | 2          | LDA       | 2               | 20        | 9              | 0.850 | 0.775 |
| (Takahashi <i>et al.</i> , 1982)  | 2          | LLA       | 3               | 22        | —              | 1     | —     |
| (Spillane <i>et al.</i> , 1983)   | 2          | kNN       | 6               | 22        | —              | 0.909 | —     |
| (Spillane <i>et al.</i> , 1983)   | 2          | LDA       | 3               | 33        | —              | 0.807 | —     |
| (Takahashi <i>et al.</i> , 1984)  | 2          | LDA       | 3               | 22        | 9              | 1     | 0.775 |
| (Miyashita <i>et al.</i> , 1986b) | 3          | LDA       | 2               | 22        | —              | 0.955 | 0.775 |
| (Miyashita <i>et al.</i> , 1986b) | 3          | SIMCA     | 5               | 91        | —              | 0.840 | —     |
| (Drew <i>et al.</i> , 1998)       | 3          | DA        | 11 <sup>d</sup> | 50        | —              | 1     | —     |
| (Drew <i>et al.</i> , 1998)       | 3          | Gráfico   | 2               | 50        | —              | 0.862 | —     |
| (Spillane <i>et al.</i> , 2002)   | 2          | LDA       | 4               | 23        | —              | 0.850 | —     |
| (Spillane <i>et al.</i> , 2002)   | 2          | QDA       | 4               | 23        | —              | 0.900 | —     |
| Presente <sup>e</sup>             | 2          | kNN       | 4               | 356       | 152            | 0.927 | 0.901 |

<sup>a</sup> no disponible; <sup>b</sup> calculado como exactitud (AC); <sup>c</sup> número de componentes para SIMCA; <sup>d</sup> número de componentes para DA; <sup>e</sup> (Rojas *et al.*, 2016c)

En la Tabla 5.8 se presentan en orden cronológico los modelos QSAR propuestos en la literatura para discriminar los gustos dulce y amargo. Se observa que el único modelo kNN fue publicado en el año 1982 por Takahashi *et al.* (Takahashi *et al.*, 1982). La calidad en ajuste del modelo de Takahashi *et al.* es comparable al obtenido en este estudio; sin embargo, la base de datos usada para su calibración está constituida únicamente por 22 compuestos, lo que lógicamente limita la realización de validación externa. En efecto, la mayoría de modelos no presentan validación externa (Iwamura, 1980; Takahashi *et al.*, 1982; Spillane *et al.*, 1983; Miyashita *et al.*, 1986b; Drew *et al.*, 1998; Spillane *et al.*, 2002), por lo que comparar el rendimiento predictivo con respecto al modelo desarrollado no es factible.

Finalmente, el dominio de aplicabilidad del modelo QSAR consiste en la comparación de la distancia promedio entre cada molécula del grupo de predicción y sus *k* vecinos más cercanos a un umbral de corte definido. En la Figura 5.9 se encuentra la distribución de las distancias promedio para las moléculas del grupo de predicción.



**Figura 5.9.** Histograma de la distancias promedio de las moléculas del grupo de predicción con respecto a sus vecinos para el modelo QSAR dulce–amargo

Dada la distribución de la Figura 5.9, se ha elegido 0.5 como valor umbral para decidir si una molécula predicha pertenece o no al AD del modelo. En este modelo la mayoría de las moléculas de predicción están por debajo de dicho umbral (9 moléculas están fuera del mismo).

### 5.3.4. Conclusiones

La calidad del modelo QSAR es adecuada, considerando la simplicidad del método de clasificación utilizado; así como el número reducido de descriptores seleccionados. Es importante indicar que los valores experimentales relacionados al dulzor son ruidosos debido a que la actividad molecular se mide mediante el uso de paneles de cata. Otro factor que afecta es el hecho de que no todos los edulcorantes presentan un sabor dulce puro. Adicionalmente, el hecho de que los edulcorantes poseen más de un donante de enlace hidrógeno y un aceptor de enlace hidrógeno (AH–B) dificulta el comprender cuál realmente interacciona con el receptor. Es importante indicar que los descriptores DRAGON independientes de la conformación, al igual que las metodologías quimiométricas utilizadas aquí se han aplicado por primera vez para discriminar los gustos dulce y amargo.

## 5.4 Modelo QSAR para discriminar los gustos dulce e insípido

### 5.4.1. Introducción

Se había indicado en la sección anterior que el dulzor es una característica que genera una sensación placentera en varios alimentos

y que la sacarosa es el edulcorante utilizado como estándar para comparar el dulzor de otras moléculas. Por otro lado, la insipidez se la puede definir como la falta o pérdida de gusto (dulce, amargo, ácido o umami) (Damodaran *et al.*, 2008).

Hasta donde se conoce, luego de una búsqueda exhaustiva en la literatura, no existen modelos QSAR desarrollados para discriminar moléculas dulces e insípidas. Por consiguiente, en este trabajo se ha desarrollado un modelo QSAR basado en el método de clasificación *k*NN para la discriminación de los compuestos dulces e insípidos. Los resultados de este trabajo han sido previamente publicados (Rojas *et al.*, 2016c).

### 5.4.2. Materiales y métodos

La base de datos inicial de compuestos estaba constituida por 620 compuestos, para los cuales se encuentra disponible la respuesta cualitativa dulce o insípido. Para el curado de la base de datos se aplicaron los mismos criterios que para la base de datos dulce–amargo. Se usó el mismo diagrama KNIME detallado en la Figura 5.7. De esta manera, se obtuvo una base de datos con 566 moléculas (433 dulces y 133 insípidas), la cual se presenta en la Tabla 9A, que se encuentra en el CD que acompaña a esta tesis. Para el desarrollo del modelo QSAR se utilizaron los mismos métodos descritos en la sección 5.3.2 para el modelo QSAR dulce–amargo, es decir, cálculo de los descriptores moleculares, desarrollo y validación del modelo, interpretación del mecanismo de acción de los descriptores y la definición del dominio de aplicabilidad.

### 5.4.3. Resultados y discusión

Se obtuvieron 855 descriptores independientes de la conformación luego de la eliminación de descriptores no informativos y los correlacionados mediante el método de reducción V–WSP. La base de datos fue dividida en un grupo de calibración con 396 compuestos y un grupo de predicción de 170 moléculas.

La selección de descriptores mediante los GAs acoplados con el método *k*NN sobre cada bloque de descriptores moleculares permitió obtener 141 descriptores que fueron considerados en conjunto para la segunda aplicación de los GAs para la construcción del modelo final. De esta manera se obtuvo un modelo QSAR con 9 descriptores que permiten obtener los resultados que se detallan en la Tabla 5.9.

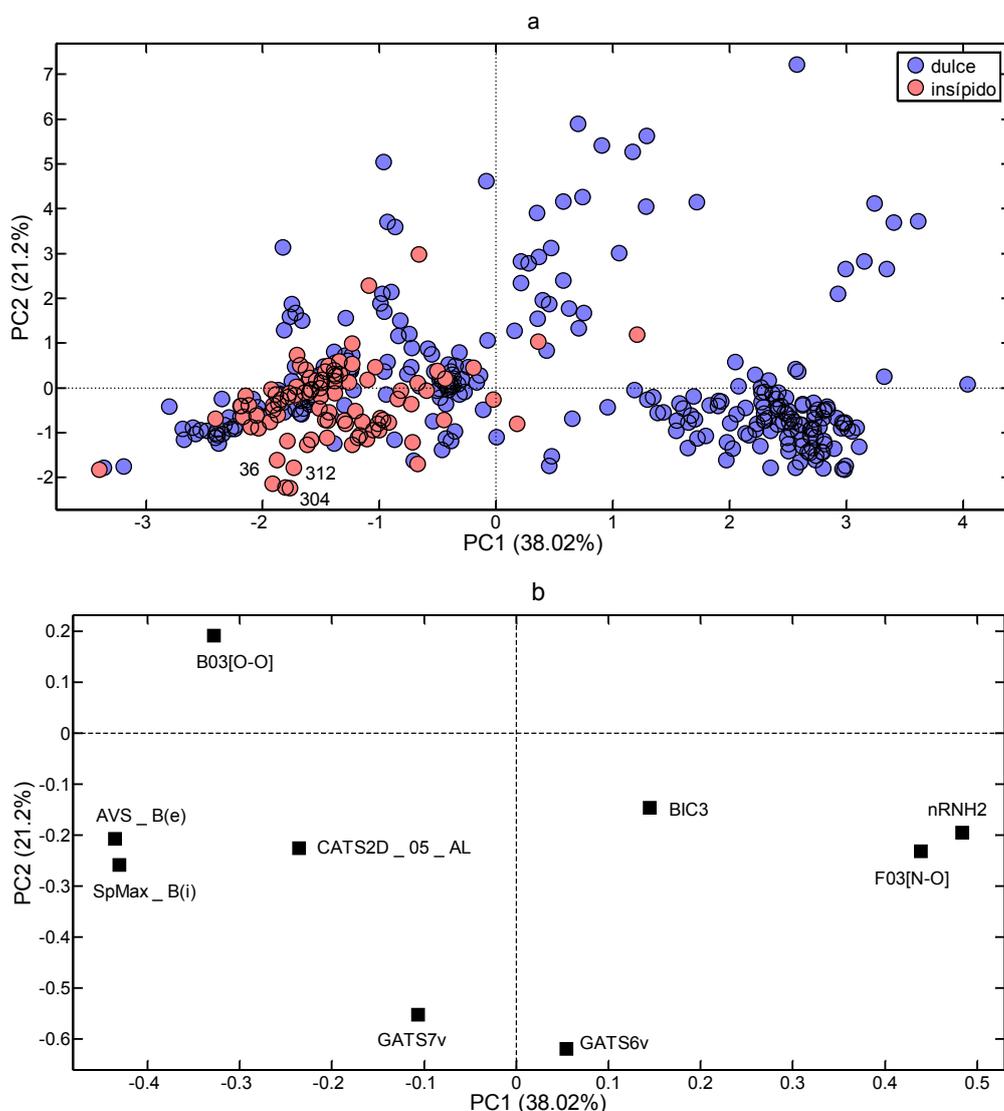
**Tabla 5.9.** Parámetros de calidad del modelo QSAR para discriminar moléculas dulces e insípidas

|                               | NER   | Sn    | Sp    |
|-------------------------------|-------|-------|-------|
| Calibración                   | 0.838 | 0.891 | 0.785 |
| Validación cruzada (5 grupos) | 0.847 | 0.898 | 0.796 |

|            |       |       |       |
|------------|-------|-------|-------|
| Predicción | 0.752 | 0.754 | 0.750 |
|------------|-------|-------|-------|

El modelo QSAR presenta resultados comparables en calibración y validación (interna y externa), con lo que se verifica la ausencia de sobreajuste en el modelo. Por otra parte, la calidad del modelo para la predicción correcta del dulzor también se verifica mediante la sensibilidad y especificidad.

En el PCA, la combinación de las dos primeras componentes permitió evaluar el mecanismo de acción de los 9 descriptores para discriminar las moléculas dulces e insípidas. La combinación de la PC1 y PC2 explica el 59% de la varianza y brinda una separación aceptable entre las dos clases, a pesar de algunas superposiciones. La Figura 5.10a muestra el gráfico de puntuaciones en estas dos componentes, mientras que en la Figura 5.10b se aprecia el gráfico de cargas.



**Figura 5.10.** Análisis de componentes principales para el modelo *k*NN para

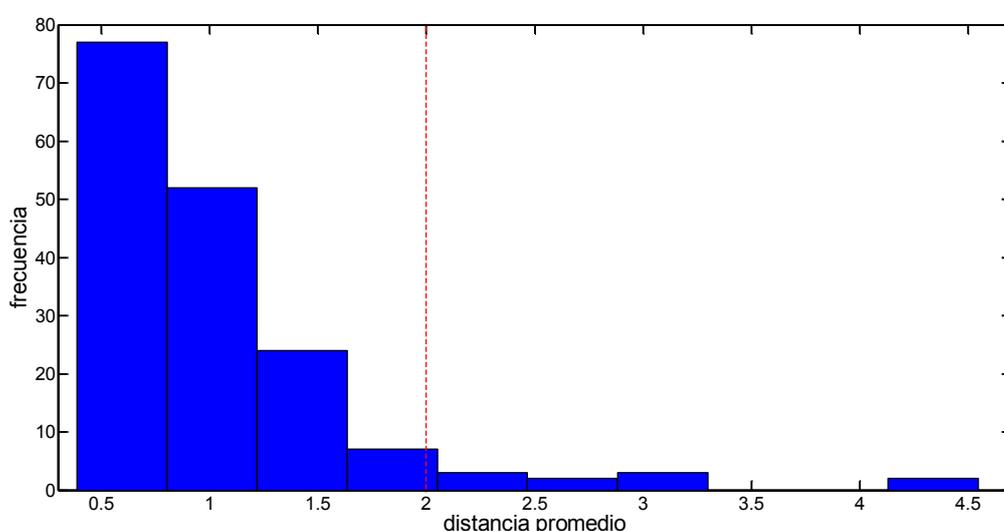
discriminar moléculas dulces e insípidas: a) gráfico de puntuaciones y b) gráfico de cargas.

La mayoría de moléculas dulces tienen puntuaciones positivas en la PC1 (primero y cuarto cuadrante), mientras que las insípidas tienen puntuaciones negativas. Los descriptores que se ubican en la región positiva de la PC1 indican que el dulzor está potencialmente relacionado al número de radicales amino ( $\text{NH}_2$ ) presentes en un esqueleto alifático ( $n\text{RNH}_2$ ) (Kode *srl.*, 2016), y al número de pares de átomos de nitrógeno y oxígeno separados por una distancia topológica de 3 unidades ( $\text{F03}[\text{N}-\text{O}]$ ) (Carhart *et al.*, 1985). Asimismo, el índice del contenido de información de enlace (simetría de entorno de tercer orden) ( $\text{BIC3}$ ) (Magnuson *et al.*, 1983) brinda información relacionada a la complejidad molecular de un compuesto; es decir, este descriptor toma valores altos al incrementar el número de equivalencias de clases de la sustancia (número de átomos similares de orden 3).

Por otra parte, los descriptores  $\text{AVS\_B}(e)$ ,  $\text{SpMax\_B}(i)$ , y  $\text{CATS2D\_05\_AL}$ , que se ubican en la zona negativa de la PC1 caracterizan a los compuestos insípidos, así como a algunas moléculas dulces. Las electronegatividades de Sanderson ( $e$ ) y la energía de ionización ( $i$ ) se usan para ponderar el grafo molecular y obtener las correspondientes matrices de Burden  $\text{B}(e)$  y  $\text{B}(i)$ . A partir de estas matrices se calculan los descriptores  $\text{AVS\_B}(e)$  como la suma del promedio de las filas y  $\text{SpMax\_B}(i)$  como el autovalor principal. Estos dos descriptores se encuentran correlacionados entre sí. Por otra parte, el descriptor  $\text{CATS2D\_05\_AL}$  cuenta los pares de átomos aceptores de enlaces de hidrógeno ( $A$ ) y átomos lipofílicos ( $L$ ) separados por 5 enlaces. De hecho, Spillane *et al.* (Spillane *et al.*, 1996) indicaron que la hidrofobicidad es un parámetro importante para la determinación del dulzor; mientras que Birch *et al.* (Birch *et al.*, 1994) refirieron que el dulzor puede ser atribuido a un equilibrio hidrófilo-lipófilo. El grupo hidrófilo actúa como un ancla que permite a la zona hidrofóbica del edulcorante ajustarse dentro el sitio de enlace hidrofóbico en el receptor (Yuasa *et al.*, 1994). Adicionalmente, en la sección 5.2 se presentó la importancia de algunos descriptores  $\text{CATS2D}$  para la predicción del dulzor relativo.

Por otra parte, la segunda componente brinda información relacionada a otros dos descriptores de autocorrelación: Autocorrelaciones de Geary (Geary, 1954) a desplazamiento 6 ( $\text{GATS6v}$ ) y desplazamiento 7 ( $\text{GATS7v}$ ) ponderados por el volumen de van der Waals. Autocorrelaciones positivas producen valores entre 0 y 1; mientras que autocorrelaciones negativas generan valores mayores a 1. Debido a que estos dos índices de autocorrelación se encuentran ubicados en la región negativa de la PC2, indican que la mayoría de los compuestos insípidos tienen átomos con volumen alto, los cuales se encuentran separados por una distancia topológica de 6 y 7 unidades, siendo el

caso de cuatro derivados de la sacarina (que corresponden a los compuestos numerados como 36, 304 y 312 en la Tabla 9A). Adicionalmente, un número razonable de moléculas dulces que se ubican en la región positiva de la PC2, estarían representados por valores pequeños de estos dos descriptores de autocorrelación. Finalmente, en la región positiva de la PC2, el descriptor presencia/ausencia de pares de átomos oxígeno–oxígeno a distancia topológica 3 (B03[O–O]) (Carhart *et al.*, 1985) describe el dulzor de aquellos compuestos caracterizados por puntuaciones positivas en la segunda componente, y contemporáneamente la ausencia de dicho par de átomos en los compuestos insípidos.



**Figura 5.11.** Histograma de la distancias promedio de las moléculas del grupo de predicción con respecto a sus vecinos para el modelo QSAR dulce–insípido

Al igual que para el modelo QSAR dulce–amargo, el dominio de aplicabilidad se efectuó mediante la comparación de la distancia promedio para cada molécula del grupo de predicción y sus  $k$  vecinos más cercanos a un umbral de corte de 2 (Figura 5.11). Existen 24 moléculas (19 dulces y 4 insípidas) que están fuera del AD.

#### 5.4.4. Conclusiones

La capacidad predictiva del modelo QSAR desarrollado es apropiada para discriminar moléculas dulces e insípidas. Los descriptores DRAGON independientes de la conformación, al igual que las metodologías quimiométricas utilizadas aquí se han aplicado por primera vez para discriminar moléculas dulces e insípidas. De hecho, este modelo es el primero desarrollado con la finalidad de distinguir este tipo de compuestos. Adicionalmente el modelo QSAR dulce–insípido permite entender qué cambios estructurales deben darse para que una molécula

pase de ser dulce a insípida. Por ejemplo, el edulcorante sacarina pierde el dulzor (se convierte en insípido) cuando se reemplaza el grupo imino por los radicales metilo, etilo o bromoetilo.

## **5.5 Modelo QSAR basado en un sistema experto para predecir el dulzor**

### **5.5.1. Introducción**

Las industrias de alimentos y farmacéutica muestran un creciente interés en el desarrollo o descubrimiento de nuevos edulcorantes que puedan tener propiedades beneficiosas. Así, los químicos de alimentos y químicos farmacéuticos tienen el reto de diseñar compuestos que presenten un sabor dulce puro, similar al de la sacarosa. Por ejemplo, el desarrollo de edulcorantes bajos en calorías sin retrogusto ha sido útil para productos alimenticios y medicinas para personas con diabetes (Damodaran *et al.*, 2008; Morini *et al.*, 2011). La medición experimental del gusto se realiza mediante panelistas entrenados y soluciones estándar dulces, amargas, ácidas y saladas. De esta manera, se asigna un gusto y su intensidad (cuando es posible) a cada compuesto (Spillane *et al.*, 1993).

Varias teorías han sido propuestas para describir la relación entre la estructura química y el dulzor. Oertly y Myers (Oertly & Myers, 1919) explicaron la producción del dulzor por medio de la relación entre los grupos funcionales “glucóforos” y “auxoglucos”. Posteriormente, Shallenberger y Acree (Shallenberger & Acree, 1967) propusieron la teoría AH–B, en la que indican que para que un compuesto sea dulce debe tener un donante de enlace hidrógeno (AH) y un aceptor de enlace hidrógeno (B) separados por una distancia de aproximadamente 3 Å. Por otra parte, Lemont Kier (Kier, 1972) propuso la teoría B–X, en la que establece que un compuesto debe tener un tercer sitio de enlace. Finalmente, Nofre y Tinti (Nofre & Tinti, 1996) propusieron la teoría de enlace multipunto (MPA), en la que sugirieron un total de 8 sitios de interacción para el receptor del dulzor, aunque no todos los edulcorantes interaccionan con todos los sitios.

Algunos edulcorantes aceptados para consumo humano han sido descubiertos por casualidad (Birch, 1999), por ejemplo, sacarina, ciclamato y aspartamo. Por otra parte, durante la síntesis de nuevos edulcorantes, algunas variaciones en la estructura química base pueden cambiar el dulzor a no dulce (amargo, insípido, ácido y salado) (Damodaran *et al.*, 2008). Para enfrentar estos inconvenientes, los químicos han desarrollado modelos matemáticos basados en la teoría QSAR/QSPR con la finalidad de predecir el dulzor de los compuestos y optimizar la síntesis los mismos. Además de los modelos QSAR previamente descritos para los gustos dulce–amargo, se han propuesto otros modelos que buscan discriminar entre compuestos dulces y no

dulces de carbosulfamatos (Miyashita *et al.*, 1986a; Okuyama *et al.*, 1988) y otros derivados del sulfamato (Spillane & McGlinchey, 1981; Spillane & Sheahan, 1989; Spillane & Sheahan, 1991; Spillane *et al.*, 1993; Spillane *et al.*, 2000; Spillane *et al.*, 2003; Spillane *et al.*, 2009).

Por lo tanto, el objetivo de este estudio fue el desarrollar un modelo QSAR basado en un sistema experto para la predicción del dulzor. Se ha usado una base de datos extensa de 649 moléculas (435 dulces, 133 insípidas y 81 amargas) mediante la fusión de las bases de datos utilizadas en los estudios presentados en las secciones 5.3 y 5.4. El sistema experto combina un análisis de similitud estructural y dos modelos QSAR de clasificación. El análisis de similitud estructural se ha desarrollado mediante el uso de las huellas digitales moleculares de conectividad ampliada (ECFPs) y el escalado multidimensional (MDS), mientras que los modelos de clasificación están basados en descriptores moleculares independientes de la conformación y los métodos de clasificación de los *N*-vecinos más cercanos (N3) y el análisis discriminante de mínimos cuadrados parciales (PLSDA). Los resultados de este trabajo han sido recientemente publicados (Rojas *et al.*, 2017).

### 5.5.2. Materiales y métodos

En este estudio se han usado las bases de datos dulce-amargo y dulce-insípido conjuntamente. La base de datos inicial estaba compuesta por 727 moléculas, las cuales están descritas por una clase experimental (dulce, amargo, insípido). Debido a que el mayor interés científico es la identificación del dulzor, las clases amargo e insípido se fusionaron en una sola clase etiquetada como “no dulce”. La base de datos fue curada usando un diagrama de flujo programado en KNIME (Figura 5.12) utilizando los siguientes criterios:

1. Exclusión de proteínas dulces, por ejemplo, pentadina, taumanina, monellina.
2. Se retuvieron las estructuras iónicas; por ejemplo, glicirrizinato tripotásico, sales de aspartamo-acesulfamo
3. Para cada estructura se obtuvo la notación lineal SMILES canónico a partir de la estructura diseñada.
4. Las clases amargo e insípido se fusionaron en una sola clase denominada no dulce.
5. Se verificaron los compuestos en función del SMILES canónico, es decir: estereoisómeros pertenecientes a dos clases diversas fueron excluidos (por ejemplo, D-Arginina y L-Arginina) y entre moléculas dulces con el mismo código SMILES, solo una de ellas fue retenida.

De esta manera, la base de datos curada está constituida de 649 moléculas dividida en dos grupos de 435 compuestos dulces y 214 compuestos no dulces (81 amargos y 133 insípidos). La clase no dulce constituye el 33% del total de las moléculas consideradas. Esta base de datos está disponible en la Tabla 10A, que se encuentra en el CD que acompaña a esta tesis.

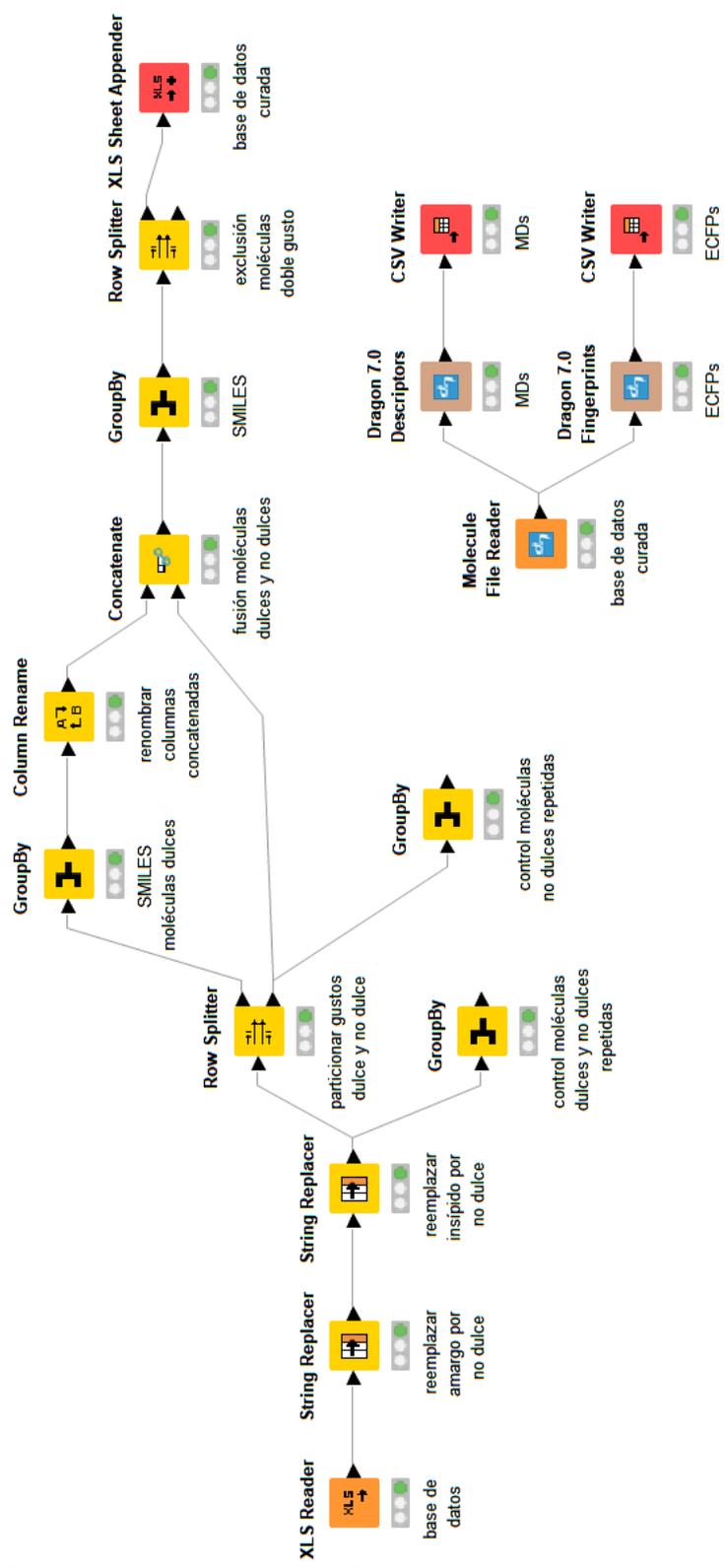


Figura 5.12. Diagrama de flujo KNIME para el curado de la base de datos de gustos dulce y no dulce

Los descriptores moleculares independientes de la conformación y las huellas digitales moleculares de conectividad ampliada (ECFPs) se han calculado usando el nodo del programa DRAGON versión 7 (Kode srl., 2016) implementado en KNIME (Figura 5.12), los cuales se han usado como representación de la estructura molecular. Se ha utilizado el enfoque DRAGON para el cálculo de los descriptores en las estructuras iónicas, el cual consiste en la aplicación de la definición y algoritmo original del descriptor considerado. Adicionalmente, para cada molécula se ha calculado un vector binario (ECFPs) con 2048 bits obtenido mediante 2 bits por estructura y un radio máximo de 2.

Las ECFPs se han usado inicialmente para realizar una exploración de las similitudes entre moléculas mediante el escalado multidimensional (MDS). De esta manera es posible identificar grupos consistentes de estructuras. Posteriormente, y usando el grupo donde están mezcladas las moléculas dulces y no dulces, se han aplicado los métodos de clasificación: *N*-vecinos más cercanos (N3) (Todeschini *et al.*, 2015) y análisis discriminante de mínimos cuadrados parciales (PLSDA) (Wold *et al.*, 2001), los cuales son métodos de discriminación no lineal basado en similitudes locales y lineales, respectivamente. Los modelos de clasificación se han aplicado luego de realizar una reducción no supervisada de descriptores basada en el método V-WSP (Ballabio *et al.*, 2014).

Para validación de los modelos de clasificación, la base de datos se dividió de forma casual y proporcional a la numerosidad de las clases en conjuntos de calibración (70%) y predicción (30%). Así, el grupo de entrenamiento incluye 488 moléculas (327 dulces y 161 no dulces) y el grupo de predicción las restantes 161 moléculas (108 dulces y 53 no dulces). Las moléculas del grupo de calibración y que pertenecen al grupo donde se superponen las moléculas dulces y no dulces se usaron para la selección de descriptores mediante algoritmos genéticos (GAs) (Leardi, 2009) acoplados con los métodos N3 y PLSDA. Durante la selección GAs se usó la validación cruzada basado en ventanas venecianas con 5 grupos.

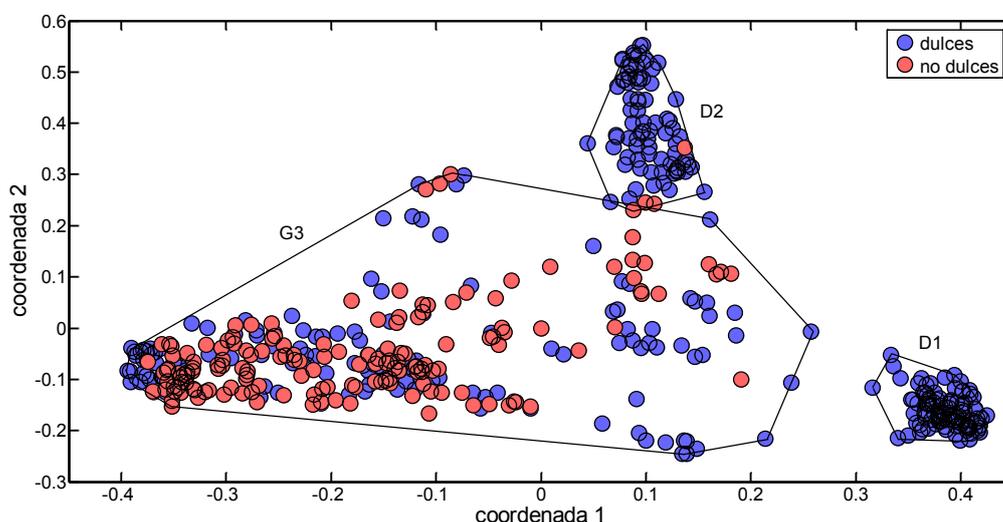
Para el método N3 se usó el escalado de rango de los descriptores y la distancia Euclidiana promedio como medida de distancia, mientras que para el método PLSDA los descriptores fueron autoescalados. El valor del parámetro alfa en N3 y el número óptimo de variables latentes (LVs) en PLSDA se han seleccionado en función de la mayor tasa de aciertos en validación ( $NER_{cv}$ ). También se evaluó la especificidad ( $Sp$ ) y sensibilidad ( $Sn$ ) de la clase dulce.

La interpretación del mecanismo de acción de los descriptores en el modelo N3 se realizó mediante el PCA (Jolliffe, 2002), debido a que este método, al igual que *k*NN, no brinda coeficientes para cuantificar la contribución de los descriptores. Para el método PLSDA se analizaron los coeficientes de las LVs óptimas. Finalmente, se construyó un sistema experto mediante el ensamblado del análisis de similitud MDS y el

consenso entre los métodos PLSDA y N3. El sistema experto QSAR ha sido validado mediante validación cruzada Monte Carlo, y mediante validación externa para las moléculas de grupo de predicción.

### 5.5.3. Resultados y discusión

Las 488 moléculas del grupo de calibración se usaron para realizar una exploración de similitud molecular basada en las huellas digitales moleculares de conectividad ampliada. Debido a que las ECFPs son vectores binarios, las similitudes se han cuantificado mediante el coeficiente de similitud de Jaccard–Tanimoto (Jaccard, 1912; Rogers & Tanimoto, 1960) para producir un escalado multidimensional de la base de datos. La Figura 5.13 muestra las puntuaciones del MDS de las dos primeras coordenadas (varianza explicada igual al 69.85%).



**Figura 5.13.** Proyección de las dos primeras coordenadas del MDS para las moléculas del grupo de calibración

En el MDS se han identificado tres grupos consistentes (D1, D2 y G3) con similitudes estructurales específicas. El grupo D1 está compuesto por 143 moléculas dulces que tienen como característica principal la presencia del ácido amino aspártico dentro de la estructura química. Sin embargo, otros compuestos dulces con la misma estructura se ubican en el grupo G3, por ejemplo aspartamo, súper aspartamo, sal de aspartamo–acesulfamo. Por otra parte, existen 107 compuestos en el grupo D2, los cuales incluyen 100 moléculas dulces y únicamente 7 estructuras no dulces. Finalmente, las restantes 399 sustancias, particularmente, la mayoría de compuestos no dulces se localizan en el grupo G3.

Debido a que el MDS provee un agrupamiento satisfactorio de los compuestos analizados en función del gusto, se ha considerado el desarrollo de un modelo QSAR basado en un sistema experto para

optimizar la discriminación de los compuestos dulces. Este sistema está estructurado de la siguiente manera:

1. Identificar el grupo asociado a la molécula a ser clasificada en función del análisis de similitud basado en las ECFPs. Por ejemplo, si la molécula es asignada a los grupos D1 o D2, es muy probable que pertenezca a la clase dulce.
2. Aplicación de modelos QSAR basados en descriptores moleculares y calibrados usando las moléculas del grupo G3 para discriminar los dos gustos dentro de este espacio químico.

Para el desarrollo de los modelos QSAR basado en descriptores, se han usado las 297 moléculas del grupo de entrenamiento que pertenecen al grupo G3 para calibrar dos modelos basados en los métodos N3 y PLSDA. Cada molécula estaba inicialmente representada por 3763 descriptores moleculares independientes de la conformación. Posteriormente se aplicó la reducción de descriptores basado en el método V-WSP a un umbral de correlación del 0.95. De esta manera 875 descriptores fueron analizados mediante algoritmos genéticos. Este proceso se desarrolló en dos etapas: 1) los GAs acoplados con los métodos de clasificación N3 y PLSDA se aplicaron separadamente sobre cada uno de los 18 bloques de descriptores y 2) los descriptores seleccionados en cada bloque se fusionaron para aplicar nuevamente los GAs. De esta manera, se obtuvieron dos modelos, con un valor óptimo alfa de 1.5 para el modelo N3 y una variable latente para el modelo PLSDA.

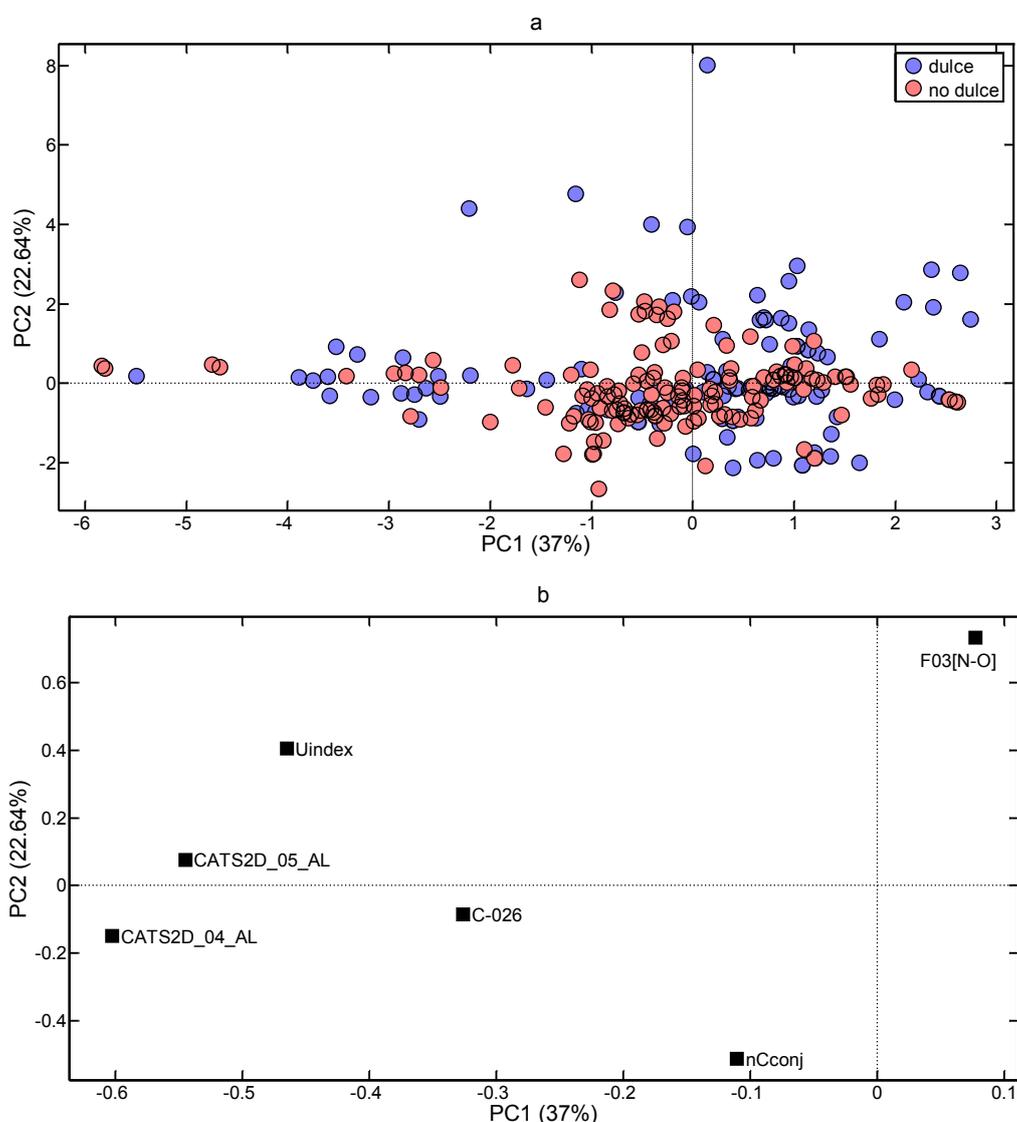
**Tabla 5.10.** Parámetros de calidad de los modelos QSAR basados en los métodos N3, PLSDA y consenso entre N3 y PLSDA

| Modelo   | Calibración |       |       | Validación cruzada (5 grupos) |       |       |
|----------|-------------|-------|-------|-------------------------------|-------|-------|
|          | NER         | Sn    | Sp    | NER                           | Sn    | Sp    |
| N3       | 0.748       | 0.764 | 0.732 | 0.738                         | 0.750 | 0.726 |
| PLSDA    | 0.722       | 0.636 | 0.809 | 0.711                         | 0.607 | 0.815 |
| Consenso | 0.852       | 0.792 | 0.913 | 0.831                         | 0.772 | 0.890 |

La calidad de los modelos presentados en la Tabla 5.10 sugiere una capacidad apropiada de los modelos desarrollados (N3 y PLSDA). Los resultados son comparables en calibración y en validación cruzada para ambos métodos de clasificación, indicando que presentan buena calidad para la discriminación entre compuestos dulces y no dulces.

Una interpretación directa del mecanismo de acción de los descriptores del modelo N3 no es factible, debido a que este método está basado en similitudes locales. Por tal motivo se ha usado el análisis de componentes principales, donde la varianza explicada por las dos primeras componentes es del 59.64%. La Figura 5.14a muestra el gráfico de puntuaciones, mientras que en la Figura 5.14b se ve el gráfico de cargas.

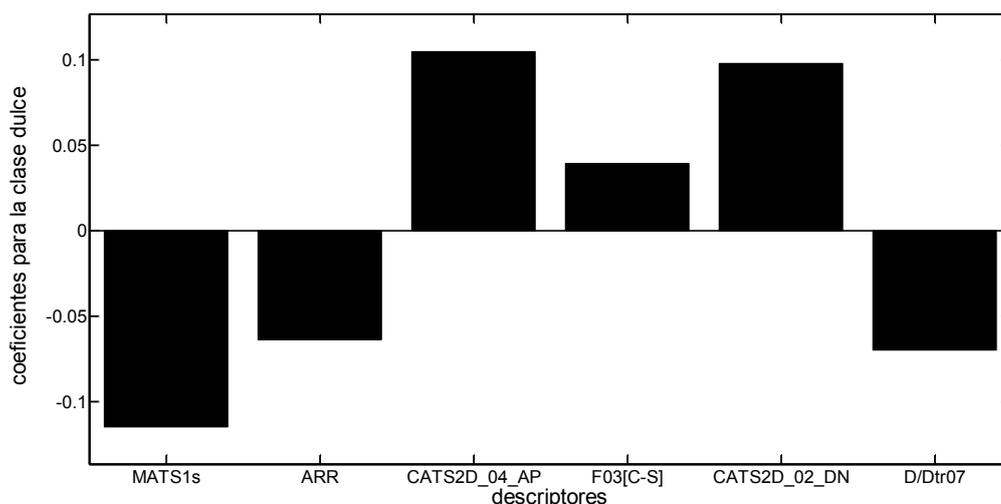
Una clara identificación de las dos clases no es factible debido a la superposición de las moléculas en el grupo G3. Un grupo pequeño y compacto de compuestos dulces se ubican en la zona de puntuaciones positivas de la PC2 (primero y segundo cuadrante). El análisis del correspondiente gráfico de cargas indica que el dulzor de las mismas está influenciada por valores altos del número de pares de átomos de nitrógeno y oxígeno separados por una distancia topológica de 3 enlaces (F03[N-O]) (Carhart *et al.*, 1985). En efecto, se había demostrado la utilidad de este descriptor para la discriminación entre moléculas dulces e insípidas en la sección 5.4 (Rojas *et al.*, 2016c). Por otra parte, el número de átomos de carbonos conjugados no aromáticos (sp<sup>2</sup>) (nCconj) (Kode srl., 2016) tiene alto valor de carga en la PC2, describiendo de esta manera los compuestos no dulces que se ubican en esta región, así como algunos compuestos dulces.



**Figura 5.14.** Análisis de componentes principales para el modelo N3: a) gráfico de puntuaciones y b) gráfico de cargas.

Adicionalmente, los descriptores CATS2D\_04\_AL, CATS2D\_05\_AL, UIndex y C-026 también describen algunas moléculas dulces con puntuaciones negativas en la PC1. Los descriptores CATS2D\_04\_AL y CATS2D\_05\_AL (Fechner *et al.*, 2003) cuentan los pares de aceptores de enlaces de hidrógeno (A) y átomos lipofílicos (L) separados por 4 y 5 enlaces, respectivamente, e indican que el dulzor de dichas moléculas está asociado a la presencia de hidrofobicidad o a un equilibrio hidrófilo-lipófilo (tal como se describió en la sección 5.4). El índice U de Balaban, el cual está relacionado al grado de ramificación molecular (Balaban & Balaban, 1991) y la presencia de átomos de carbono unidos a cualquier átomo electronegativo (O, N, S, P, Se, halógenos) y a otros dos grupos mediante enlaces aromáticos (C-026) (Ghose *et al.*, 1998), también son importantes para la predicción del dulzor en el modelo no lineal local N3.

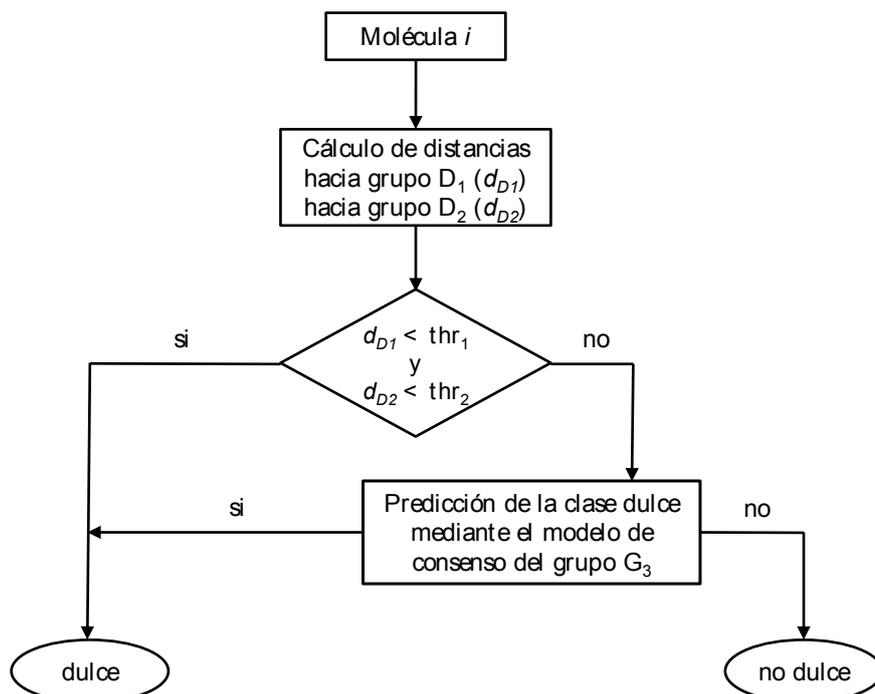
Para el método de clasificación PLSDA, la Figura 5.15 muestra las puntuaciones en la variable latente 1 y los coeficientes de los descriptores para la clase dulce. Los coeficientes del modelo sugieren que el dulzor se encuentra descrito por los descriptores CATS2D\_04\_AP, CATS2D\_02\_DN (Fechner *et al.*, 2003) y F03[C-S] (Carhart *et al.*, 1985). El descriptor CATS2D\_04\_AP codifica la presencia de pares de átomos aceptores de enlaces de hidrógeno (A) y átomos con carga positiva (P) separados por 4 enlaces; mientras que el descriptor CATS2D\_02\_DN indica la presencia de pares de átomos donantes de enlaces de hidrógeno (D) y átomos con carga negativa (N) a distancia topológica 2. En efecto, la presencia de pares de potenciales farmacóforos positivos-negativos a distancia topológica 2 se presentó en la sección 5.2 para la predicción del RS (Rojas *et al.*, 2016b). Por otra parte, el descriptor F03[C-S] indica que el dulzor está también asociado a la frecuencia de pares de átomos carbono y azufre en la estructura molecular separados por 3 enlaces.



**Figura 5.15.** Coeficientes de los descriptores para el grupo de calibración en el modelo PLSDA para la clase dulce.

Los coeficientes para la clase no dulce tienen el mismo valor pero signo contrario con respecto a la clase dulce. De esta manera, los descriptores relacionados con las moléculas no dulces son: la autocorrelación de Moran a desplazamiento 1 ponderada por el estado intrínseco (MATS1s), la relación aromática (ARR) y el índice anillo distancia/detour de orden 7 (D/Dtr07). MATS1s se calcula mediante el coeficiente de Moran (Moran, 1950) a un grafo molecular ponderado por el estado intrínseco como propiedad molecular. Valores positivos de este descriptor indican autocorrelaciones espaciales positivas. Por otra parte, D/Dtr07 (Randić, 1997) es un descriptor topológico que indica la razón entre las longitudes de enlace de la ruta más corta y las longitudes de enlace de la ruta más larga entre cualquier par de vértices que pertenecen a anillos de 7 átomos. Este descriptor en combinación con la proporción del número de enlaces aromáticos con respecto al número total de enlaces, excepto con el H (AAR) (Kode srl., 2016), indica que la clase no dulce está relacionada a la presencia de anillos aromáticos en la estructura.

Debido a que los métodos N3 y PLSDA se basan en diferentes tipos de descriptores y metodologías, se realizó un análisis de consenso (van Rhee, 2003) para combinar la información y capacidad predictiva de los dos modelos. Por lo tanto, cada molécula será predicha si los dos modelos la clasifican en la misma clase. Caso contrario, no será clasificada (método estricto). Los resultados del consenso en calibración (33% de moléculas no clasificadas) y en validación cruzada de ventanas venecianas con 5 grupos (32% de compuestos no clasificados) se muestran en la Tabla 5.10. La calidad del consenso confirma la definición de mejorar la capacidad predictiva. Sin embargo, el número de moléculas no clasificadas es considerable (33% en calibración y 32% en validación cruzada) con respecto a las moléculas del grupo G3.

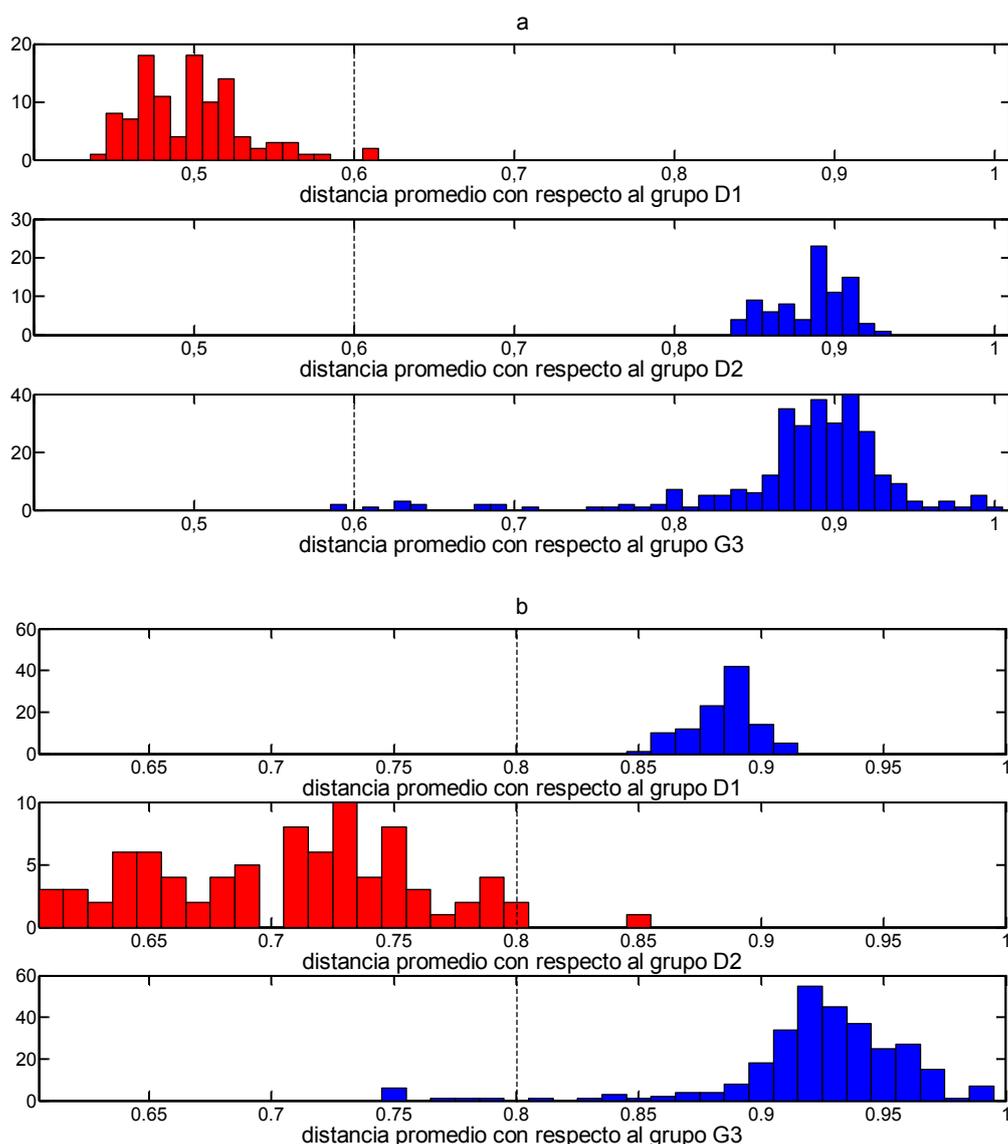


**Figura 5.16.** Diagrama de flujo de las principales etapas del modelo QSAR basado en un sistema experto para la predicción del dulzor.

Una vez que los modelos de clasificación se han calibrado usando las moléculas del grupo G3, se ha desarrollado el modelo QSAR basado en un sistema para la predicción del dulzor de toda la base de datos. En la Figura 5.16 se indica la estructura del sistema experto para la predicción del dulzor. Así, la predicción del dulzor para cualquier molécula de interés se obtiene de acuerdo al siguiente procedimiento:

1. Calcular el vector de ECFPs para la molécula de interés y posteriormente la distancia promedio de Jaccard–Tanimoto con respecto a las moléculas de los grupos D1 ( $d_{D1}$ ) y D2 ( $d_{D2}$ ), respectivamente.
2. Si las distancias  $d_{D1}$  y  $d_{D2}$  son menores que umbrales de corte definidos (0.6 y 0.8, respectivamente), entonces la molécula de interés será directamente clasificada como dulce, debido a su alta similitud estructural con las moléculas de los grupos D1 y D2.
3. Caso contrario, si las distancias  $d_{D1}$  y  $d_{D2}$  son mayores que los umbrales de corte establecidos, entonces el dulzor de la molécula de interés será predicha mediante el modelo de consenso basado en los métodos N3 y PLSDA.

Los umbrales de corte definidos en el paso 2 se han elegido en función de la distribución de las similitudes promedio de cada molécula del grupo de calibración con respecto a las moléculas de los tres grupos. Estas distribuciones permitieron definir un valor umbral de 0.6 (Figura 5.17a) para el grupo D1 y un valor umbral de 0.8 (Figura 5.17b) para el grupo D2.



**Figura 5.17.** Histograma de las similitudes promedio de Jaccard-Tanimoto de las moléculas de grupo de calibración con respecto a las moléculas del grupo D1 (a) y grupo D2 (b).

El modelo QSAR basado en el sistema experto ha sido validado mediante la validación cruzada Monte Carlo (con 1000 interacciones) y validación externa mediante la predicción del dulzor de las 161

moléculas del grupo de predicción. Los resultados confirman la estabilidad del sistema experto en calibración (ausencia de sobreajuste) y la capacidad predictiva del mismo. El porcentaje de moléculas no clasificadas es de aproximadamente el 20%. Los resultados del modelo QSAR basado en el sistema experto se presentan en la Tabla 5.11.

**Tabla 5.11.** Parámetros de calidad del modelo QSAR basado en un sistema experto para la predicción del dulzor con consenso “*estricto*”

|                        | NER   | Sn    | Sp    | % no asignados |
|------------------------|-------|-------|-------|----------------|
| Calibración            | 0.892 | 0.929 | 0.855 | 19.7           |
| Validación Monte Carlo | 0.887 | 0.927 | 0.848 | 20.5           |
| Validación externa     | 0.848 | 0.880 | 0.816 | 19.3           |

Por otra parte, cada predicción del sistema experto debe estar relacionada con la definición apropiada del dominio de aplicabilidad. De esta manera, el AD del sistema experto está integrado por los siguientes pasos:

1. Calcular el vector de ECFPs para cada molécula a ser predicha y seguidamente la distancia promedio de Jaccard–Tanimoto con respecto a las moléculas de los grupos D1 ( $d_{D1}$ ) y D2 ( $d_{D2}$ ), respectivamente.
2. Si las distancias  $d_{D1}$  y  $d_{D2}$  son menores que los umbrales de corte definidos en la Figura 5.17a y Figura 5.17b, entonces la molécula está dentro del AD del sistema experto. Esto se fundamenta en la alta similitud estructural con las moléculas de los grupos D1 y D2.
3. Caso contrario, si las distancias  $d_{D1}$  y  $d_{D2}$  son mayores que los umbrales de corte definidos, la definición del AD de la molécula será realizada por comparación del valor de influencia con respecto al valor crítico para el modelo PLSDA, mientras que para el método N3 se analiza la distribución de las similitudes promedio (similar al AD para los modelos basados en similitud *kNN*).

De esta manera, cualquier molécula de interés debe satisfacer una de estas tres condiciones para estar dentro del AD del sistema experto. Caso contrario su predicción será considerada extrapolación.

En la Tabla 5.12 se presentan los diversos modelos QSAR desarrollados para discriminar moléculas dulces y no dulces. En la mayoría de los casos estos modelos se han establecido mediante el uso de bases de datos con familias de compuestos homogéneas (Iwamura, 1980; Kier, 1980; Spillane & McGlinchey, 1981; Takahashi *et al.*, 1982; Spillane *et al.*, 1983; Takahashi *et al.*, 1984; Miyashita *et al.*, 1986a;

Miyashita *et al.*, 1986b; Okuyama *et al.*, 1988; Spillane & Sheahan, 1989; Spillane & Sheahan, 1991; Spillane *et al.*, 1993; Drew *et al.*, 1998; Spillane *et al.*, 2000; Spillane *et al.*, 2002; Spillane *et al.*, 2003; Spillane *et al.*, 2009). Este hecho limita la generalización de tales modelos a diferentes tipos de compuestos, es decir, el dominio de aplicabilidad de los mismos es restringido. Por otra parte, algunos modelos no han sido sometidos a validación externa (Iwamura, 1980; Takahashi *et al.*, 1982; Spillane *et al.*, 1983; Miyashita *et al.*, 1986a; Miyashita *et al.*, 1986b; Spillane & Sheahan, 1989; Spillane & Sheahan, 1991). Además, el modelo QSAR basado en un sistema experto considera moléculas dulces y no dulces que cubren un espacio químico amplio, por ejemplo, derivados de la sacarosa, sacarina, abrasido, acesulfamo, isovanillina, mogrosido, periandrina, sacarina, rebaudiosido, ciclamato, suosan, aspartamo, así como otros compuestos heterogéneos.

**Tabla 5.12.** Modelos QSAR para discriminar moléculas dulces y no dulces

| Modelos                           | Gustos         | Clases | Método    | d               | moléculas |                | NER                |                    |
|-----------------------------------|----------------|--------|-----------|-----------------|-----------|----------------|--------------------|--------------------|
|                                   |                |        |           |                 | cal       | pred           | cal                | pred               |
| (Iwamura, 1980)                   | Dulce–amargo   | 2      | Regresión | 3               | 49        | — <sup>a</sup> | —                  | —                  |
| (Kier, 1980)                      | Dulce–amargo   | 2      | LDA       | 2               | 20        | 9              | 0.850              | 0.775              |
| (Spillane & McGlinchey, 1981)     | Dulce–no dulce | 2      | Gráfico   | 2               | 35        | 12             | 0.914 <sup>b</sup> | 0.917 <sup>b</sup> |
| (Takahashi <i>et al.</i> , 1982)  | Dulce–amargo   | 2      | LLA       | 3               | 22        | —              | 1                  | —                  |
|                                   |                |        | kNN       | 6               | —         | —              | 0.909              | —                  |
| (Spillane <i>et al.</i> , 1983)   | Dulce–amargo   | 2      | LDA       | 3               | 33        | —              | 0.807              | —                  |
| (Takahashi <i>et al.</i> , 1984)  | Dulce–amargo   | 2      | LDA       | 3               | 22        | 9              | 1                  | 0.775              |
|                                   |                |        |           | 2               | —         | —              | 0.955              | 0.775              |
| (Miyashita <i>et al.</i> , 1986b) | Dulce–amargo   | 3      | SIMCA     | 5               | 91        | —              | 0.840              | —                  |
| (Miyashita <i>et al.</i> , 1986a) | Dulce–no dulce | 2      | SIMCA     | 4               | 50        | —              | 0.798              | —                  |
| (Okuyama <i>et al.</i> , 1988)    | Dulce–no dulce | 2      | SIMCA     | 1 <sup>c</sup>  | 25        | —              | 0.868              | —                  |
|                                   |                |        |           |                 | 20        | —              | 0.808              | —                  |
| (Spillane & Sheahan, 1989)        | Dulce–no dulce | 2      | LDA       | 3               | 23        | —              | 0.642              | —                  |
| (Spillane & Sheahan, 1991)        | Dulce–no dulce | 3      | Gráfico   | 2               | 57        | —              | 0.860              | —                  |
|                                   |                | 2      | LDA       | 3               | 33        | —              | 0.848 <sup>b</sup> | —                  |
|                                   |                |        |           |                 | 23        | —              | 0.870 <sup>b</sup> | —                  |
| (Spillane <i>et al.</i> , 1993)   | Dulce–no dulce | 2      | Gráfico   | 2               | 40        | —              | 0.833              | —                  |
| (Drew <i>et al.</i> , 1998)       | Dulce–amargo   | 3      | DA        | 11 <sup>d</sup> | 50        | —              | 1                  | —                  |
| (Spillane <i>et al.</i> , 2000)   | Dulce–no dulce | 2      | LDA       | 4               | —         | —              | 0.665              | —                  |
|                                   |                |        | QDA       | 4               | 101       | —              | 0.801              | —                  |
|                                   |                |        | CART      | 3               | —         | —              | 0.650              | —                  |
| (Spillane <i>et al.</i> , 2002)   | Dulce–amargo   | 2      | Gráfico   | 2               | 23        | —              | 0.862              | —                  |
|                                   |                |        | LDA       | 4               | —         | —              | 0.850              | —                  |

|                                 |           |                |   |             |    |     |     |                    |       |       |
|---------------------------------|-----------|----------------|---|-------------|----|-----|-----|--------------------|-------|-------|
|                                 |           |                |   | QDA         |    |     | —   | 0.900              | —     |       |
|                                 |           |                |   | LDA         | 4  |     | —   | 0.693              | —     |       |
| (Spillane <i>et al.</i> , 2003) | <i>et</i> | Dulce–no dulce | 2 | QDA         |    | 132 | —   | 0.683              | —     |       |
|                                 |           |                |   | CART        | 3  |     | —   | 0.815              | —     |       |
|                                 |           |                | 2 | LDA         | 2  | 58  | —   | 0.655 <sup>b</sup> | —     |       |
| (Spillane <i>et al.</i> , 2009) | <i>et</i> | Dulce–no dulce | 2 | QDA         | 3  |     | —   | 0.759 <sup>b</sup> | —     |       |
|                                 |           |                | 2 | CART        | 6  | 48  | 10  | 0.950              | 0.700 |       |
|                                 |           |                | 3 | CART        | 6  |     | 10  | 0.908              | 0.611 |       |
| (Rojas <i>et al.</i> , 2016c)   |           | Dulce–insípido | 2 | <i>k</i> NN |    | 9   | 396 | 170                | 0.866 | 0.753 |
|                                 |           | Dulce–amargo   |   |             | 4  | 356 | 152 | 0.927              | 0.901 |       |
| Presente <sup>e</sup>           |           | Dulce–no dulce | 2 | Experto     | 12 | 488 | 161 | 0.892              | 0.848 |       |

<sup>a</sup> no disponible; <sup>b</sup> calculado como exactitud (Accuracy); <sup>c</sup> número de componentes para SIMCA; <sup>d</sup> número de componentes para DA; <sup>e</sup> (Rojas *et al.*, 2017)

El sistema experto se puede considerar como un modelo más general para la predicción del dulzor de compuestos que presentan diversa estructura molecular. Asimismo, este estudio provee por primera vez un modelo QSAR basado en un sistema experto que: 1) considera contemporáneamente el uso de huellas digitales moleculares de conectividad ampliada y descriptores moleculares y 2) integra los resultados de un análisis de similitud molecular y un modelo de clasificación de consenso para la predicción del dulzor.

Varios factores pueden afectar el desarrollo de modelos QSAR para la predicción del dulzor, tales como la presencia de otros gustos a parte del dulzor (es decir, moléculas que presentan multisabores). Por ejemplo, el acesulfamo de potasio, la sacarina sódica, hernaldulcina, esteviosidos y los derivados de la isocoumarina poseen gusto dulce y amargo. El gusto de estas moléculas depende de su concentración en solución (Birch *et al.*, 1994). Por otra parte, la percepción del gusto para los compuestos que presentan más de dos gustos resulta más difícil de explicar (Shamil *et al.*, 1987). Por estos motivos, las personas tienen limitaciones para discriminar las diferencias cuando se trata de evaluar sustancias multisabor. Estas limitaciones se pueden deber a la saturación del receptor presente en las papilas gustativas de la lengua o a la polarización de los receptores del gusto (Birch *et al.*, 1994).

Por otra parte, los edulcorantes pueden existir en varias conformaciones de equilibrio que minimizan su energía (Morini *et al.*, 2011). Además, se había explicado previamente que estos compuestos también poseen más de un sitio AH–B (Spillane & Sheahan, 1989; Damodaran *et al.*, 2008); por lo que resulta complejo definir la conformación activa y cómo los sitios AH–B interactúan con el receptor del gusto dulce. Adicionalmente, la real interacción entre el receptor y el edulcorante no es conocida en su totalidad. Por ejemplo, algunas moléculas interaccionan con el receptor del dulzor pero no son reconocidas como dulces (falsas positivas); mientras que otras

moléculas que no interactúan con dicho receptor son percibidas como dulces (falsas negativas) (Bassoli *et al.*, 2008).

#### **5.5.4. Conclusiones**

En este estudio se ha desarrollado un modelo predictivo QSAR basado en un sistema experto para predecir el dulzor de moléculas orgánicas. El análisis de similitud estructural permitió la identificación de dos grupos consistentes de compuestos dulces, mientras que el tercer grupo, que contiene las moléculas dulces y no dulces, se usó para calibrar los modelos de clasificación N3 y PLSDA mediante el uso de descriptores DRAGON independientes de la conformación. Este estudio brinda el primer estudio QSAR que considera un análisis de similitud estructural basado en las ECFPs e integra los resultados del análisis de consenso entre los dos métodos de clasificación. La facilidad de implementación y la calidad predictiva del modelo permite que el sistema experto constituya una herramienta útil para las personas que trabajan en la búsqueda de nuevos edulcorantes como potenciales aditivos para ser usados en la industria de alimentos y farmacéutica.

## 5.6 Referencias

- Balaban, A. T., & Balaban, T.-S. (1991). New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances. *Journal of Mathematical Chemistry*, 8(1), 383-397.
- Ballabio, D., Consonni, V., Mauri, A., Claeys-Bruno, M., Sergent, M., & Todeschini, R. (2014). A Novel Variable Reduction Method Adapted from Space-Filling Designs. *Chemometrics and Intelligent Laboratory Systems*, 136, 147-154.
- Barker, J. S., Hattotuagama, C. K., & Drew, M. G. B. (2002). Computational Studies of Sweet-Tasting Molecules. *Pure and applied chemistry*, 74(7), 1207-1217.
- Bassoli, A., Drew, M. G. B., Hattotuagama, C. K., Merlini, L., Morini, G., & Wilden, G. R. H. (2001). Quantitative Structure-Activity Relationships of Sweet Isovanillyl Derivatives. *Quantitative Structure-Activity Relationships*, 20(1), 3-16.
- Bassoli, A., Laureati, M., Borgonovo, G., Morini, G., Servant, G., & Pagliarini, E. (2008). Isovanillic Sweeteners: Sensory Evaluation and In Vitro Assays with Human Sweet Taste Receptor. *Chemosensory Perception*, 1(3), 174-183.
- Belitz, H.-D., Rohse, H., Stempf, W., & Gries, H. (1990). Trihalogenated Benzamides: Structure Taste Relationships. *Zeitschrift für Lebensmittel-Untersuchung und Forschung*, 190(4), 319-324.
- Belitz, H.-D., Grosch, W., & Schieberle, P. (2009). *Food Chemistry* (4th ed.). Heidelberg: Springer-Verlag.
- Berg, H. (2001). Pesticide Use in Rice and Rice-Fish Farms in the Mekong Delta, Vietnam. *Crop Protection*, 20(10), 897-905.
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz Information Miner. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications*, (pp. 319-326): Springer.
- Birch, G. G., & Mylvaganam, A. (1976). Evidence for the Proximity of Sweet and Bitter Receptor Sites. *Nature*, 260, 632-634.
- Birch, G. G. (1987). Sweetness and Sweeteners. *Endeavour*, 11(1), 21-24.
- Birch, G. G., Karim, R., & Lopez, A. (1994). Novel Aspects of Structure-Activity Relationships in Sweet Taste Chemoreception. *Food quality and preference*, 5(1), 87-93.
- Birch, G. G. (1999). Modulation of Sweet Taste. *BioFactors*, 9(1), 73-80.
- Briciu, R. D., Kot-Wasik, A., Wasik, A., Namieśnik, J., & Sârbu, C. (2010). The Lipophilicity of Artificial and Natural Sweeteners Estimated by Reversed-Phase Thin-Layer Chromatography and Computed by Various Methods. *Journal of Chromatography A*, 1217(23), 3702-3706.

- Bryant, R., & McClung, A. (2011). Volatile Profiles of Aromatic and Non-Aromatic Rice Cultivars Using SPME/GC-MS. *Food chemistry*, 124(2), 501-513.
- Carhart, R. E., Smith, D. H., & Venkataraghavan, R. (1985). Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *Journal of chemical information and computer sciences*, 25(2), 64-73.
- Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Champagne, E. T. (2008). Rice Aroma and Flavor: A Literature Review. *Cereal Chemistry*, 85(4), 445-454.
- Damodaran, S., Parkin, K. L., & Fennema, O. R. (2008). *Fennema's Food Chemistry* (Fourth ed.): CRC press.
- Deutsch, E. W., & Hansch, C. (1966). Dependence of Relative Sweetness on Hydrophobic Bonding. *Nature*, 211, 75.
- Drew, M. G. B., Wilden, G. R. H., Spillane, W. J., Walsh, R. M., Ryder, C. A., & Simmie, J. M. (1998). Quantitative Structure-Activity Relationship Studies of Sulfamates RNHSO<sub>3</sub>Na: Distinction between Sweet, Sweet-Bitter, and Bitter Molecules. *Journal of Agricultural and Food Chemistry*, 46(8), 3016-3026.
- Duchowicz, P. R., Fernández, F. M., & Castro, E. A. (2010). Orthogonalization Methods in QSPR-QSAR Studies. In E. A. Castro (Ed.), *QSPR-QSAR Studies on Desired Properties for Drug Design*, (pp. 189-203 ): Research Signpost.
- Escarrone, A., Caldas, S., Furlong, E., Meneghetti, V., Fagundes, C., Arias, J., & Primel, E. (2014). Polycyclic Aromatic Hydrocarbons in Rice Grain Dried by Different Processes: Evaluation of a Quick, Easy, Cheap, Effective, Rugged and Safe Extraction Method. *Food chemistry*, 146, 597-602.
- Fatemi, M. H., & Malekzadeh, H. (2014). CORAL: Predictions of Retention Indices of Volatiles in Cooking Rice Using Representation of the Molecular Structure Obtained by Combination of SMILES and Graph Approaches. *Journal of the Iranian Chemical Society*, 12(3), 405-412.
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687-698.
- Fukuda, T., Takeda, T., & Yoshida, S. (2014). Comparison of Volatiles in Cooked Rice with Various Amylose Contents. *Food Science and Technology Research*, 20(6), 1251-1259.
- Garcia, J., Duchowicz, P. R., & Castro, E. A. (2016). Considering the Molecular Conformational Flexibility in QSAR Studies. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications and Research: QSAR in Medicinal Chemistry*, (pp. 129-158): CRC Press.

- Geary, R. C. (1954). The Contiguity Ratio and Statistical Mapping. *The incorporated statistician*, 5(3), 115-146.
- Ghose, A. K., Pritchett, A., & Crippen, G. M. (1988). Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships III: Modeling Hydrophobic Interactions. *Journal of Computational Chemistry*, 9(1), 80-90.
- Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1998). Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *The Journal of Physical Chemistry A*, 102(21), 3762-3772.
- Golbraikh, A., & Tropsha, A. (2002). Beware of  $q^2$ ! *Journal of Molecular Graphics and Modelling*, 20(4), 269-276.
- Greenberg, M. J. (1980). The Importance of Hydrophobic Properties of Organic Compounds on Their Taste Intensities: A Quantitative Structure-Taste-Intensity Study. *Journal of Agricultural and Food Chemistry*, 28(3), 562-566.
- Grimm, C. C., Bergman, C., Delgado, J. T., & Bryant, R. (2001). Screening for 2-Acetyl-1-Pyrroline in the Headspace of Rice Using SPME/GC-MS. *Journal of Agricultural and Food Chemistry*, 49(1), 245-249.
- Grimm, C. C., Champagne, E. T., & Ohtsubo, K. i. (2002). Analysis of Volatile Compounds in the Headspace of Rice Using SPME/GC/MS. In R. Marsili (Ed.), *Flavor, Fragrance, and Odor Analysis*, (pp. 229-248): Marcel Dekker, Inc.
- Grimm, C. C., Champagne, E. T., Lloyd, S. W., Easson, M., Condon, B., & McClung, A. (2011). Analysis of 2-Acetyl-1-Pyrroline in Rice by HSSE/GC/MS. *Cereal Chemistry*, 88(3), 271-277.
- Hoffmann, R., Minkin, V. I., & Carpenter, B. K. (1996). Ockham's Razor and Chemistry. *Bulletin de la Société chimique de France*, 133(2), 117-130.
- Iwamura, H. (1980). Structure-Taste Relationship of Perillartine and Nitro- and Cyanoaniline Derivatives. *Journal of Medicinal Chemistry*, 23(3), 308-312.
- Iwamura, H. (1981). Structure-Sweetness Relationship of L-Aspartyl Dipeptide Analogs. A Receptor Site Topology. *Journal of Medicinal Chemistry*, 24(5), 572-583.
- Jaccard, P. (1912). The Distribution of the Flora in the Alpine Zone. *New phytologist*, 11(2), 37-50.
- Jäger, R., Schmidt, F., Schilling, B., & Brickmann, J. (2000). Localization and Quantification of Hydrophobicity: The Molecular Free Energy Density (MolFESD) Concept and Its Application to Sweetness Recognition. *Journal of Computer-Aided Molecular Design*, 14(7), 631-646.

- Jolliffe, I. T. (2002). *Principal Component Analysis* (Second ed.): Springer.
- Katritzky, A. R., Petrukhin, R., Perumal, S., Karelson, M., Prakash, I., & Desai, N. (2002). A QSPR Study of Sweetness Potency Using the CODESSA Program. *Croatica chemica acta*, 75(2), 475-502.
- Kelly, D. P., Spillane, W. J., & Newell, J. (2005). Development of Structure-Taste Relationships for Monosubstituted Phenylsulfamate Sweeteners Using Classification and Regression Tree (CART) Analysis. *Journal of Agricultural and Food Chemistry*, 53(17), 6750-6758.
- Kier, L. B. (1972). A Molecular Theory of Sweet Taste. *Journal of pharmaceutical sciences*, 61(9), 1394-1397.
- Kier, L. B. (1980). Molecular Structure Influencing either a Sweet or Bitter Taste Among Aldoximes. *Journal of pharmaceutical sciences*, 69(4), 416-419.
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2015). PubChem Substance and Compound databases. *Nucleic acids research*, 44(D1), D1202-D1213.
- Kode srl. (2016). Dragon version 7. Software for Molecular Descriptor Calculation. <http://chm.kode-solutions.net/>.
- Leardi, R. (2009). Genetic Algorithms. In R. Tauler, B. Walczak & S. D. Brown (Eds.), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, (pp. 631-653): Elsevier.
- Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., & Adler, E. (2002). Human Receptors for Sweet and Umami Taste. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), 4692-4696.
- Linstrom, P. J., & Mallard, W. G. (2001). NIST Chemistry WebBook, NIST Standard Reference Database Number 69. In). Gaithersburg MD: National Institute of Standards and Technology.
- Liu, X., & Korenaga, T. (2001). Dynamics Analysis for the Distribution of Polycyclic Aromatic Hydrocarbons in Rice. *Journal of health science*, 47(5), 446-451.
- Lohninger, H. (1993). Evaluation of Neural Networks Based on Radial Basis Functions and Their Application to the Prediction of Boiling Points from Structural Parameters. *Journal of chemical information and computer sciences*, 33(5), 736-744.
- Magnuson, V., Harriss, D., & Basak, S. (1983). Topological Indices Based on Neighborhood Symmetry: Chemical and Biological Applications. In R. King (Ed.), *Chemical Applications of Topology and Graph Theory*, (pp. 178-191): Elsevier.
- Miyashita, Y., Takahashi, Y., Takayama, C., Ohkubo, T., Funatsu, K., & Sasaki, S.-I. (1986a). Computer-Assisted Structure/Taste Studies on Sulfamates by Pattern Recognition Methods. *Analytica Chimica Acta*, 184, 143-149.

- Miyashita, Y., Takahashi, Y., Takayama, C., Sumi, K., Nakatsuka, K., Ohkubo, T., Abe, H., & Sasaki, S. (1986b). Structure-Taste Correlation of L-Aspartyl Dipeptides Using the SIMCA Method. *Journal of Medicinal Chemistry*, 29(6), 906-912.
- Moran, P. A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17-23.
- Morini, G., Bassoli, A., & Borgonovo, G. (2011). Molecular Modelling and Models in the Study of Sweet and Umami Taste Receptors. A Review. *Flavour and fragrance journal*, 26(4), 254-259.
- NCI/CADD Group. (2013). Chemical Identifier Resolver. In).
- Nofre, C., & Tinti, J.-M. (1996). Sweetness Reception in Man: The Multipoint Attachment Theory. *Food chemistry*, 56(3), 263-274.
- Oertly, E., & Myers, R. G. (1919). A New Theory Relating Constitution to Taste. Simple Relations Between the Constitution of Aliphatic Compounds and Their Sweet Taste. *Journal of the American Chemical Society*, 41(6), 855-867.
- Okuyama, T., Miyashita, Y., Kanaya, S., Katsumi, H., Sasaki, S.-i., & Randić, M. (1988). Computer Assisted Structure-Taste Studies on Sulfamates by Pattern Recognition Method Using Graph Theoretical Invariants. *Journal of Computational Chemistry*, 9(6), 636-646.
- Pence, H. E., & Williams, A. (2010). ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education*, 87(11), 1123-1124.
- Phillips, D. H. (1999). Polycyclic Aromatic Hydrocarbons in the Diet. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 443(1), 139-147.
- Randić, M. (1975). Characterization of Molecular Branching. *Journal of the American Chemical Society*, 97(23), 6609-6615.
- Randić, M. (1997). On Characterization of Cyclic Structures. *Journal of chemical information and modeling*, 37(6), 1063-1071.
- Riahi, S., Ganjali, M. R., Pourbasheer, E., & Norouzi, P. (2008). QSRR Study of GC Retention Indices of Essential-Oil Compounds by Multiple Linear Regression with a Genetic Algorithm. *Chromatographia*, 67(11-12), 917-922.
- Rogers, D. J., & Tanimoto, T. T. (1960). A Computer Program for Classifying Plants. *Science*, 132(3434), 1115-1118.
- Rojas, C., Duchowicz, P. R., Tripaldi, P., & Pis Diez, R. (2015). Quantitative Structure–Property Relationship Analysis for the Retention Index of Fragrance-Like Compounds on a Polar Stationary Phase. *Journal of Chromatography A*, 1422, 277-288.
- Rojas, C., Duchowicz, P. R., Pis Diez, R., & Tripaldi, P. (2016a). Applications of Quantitative Structure-Relative Sweetness Relationships in Food Chemistry. In A. G. Mercader, P. R. Duchowicz & P. M. Sivakumar (Eds.), *Chemometrics Applications*

- and Research: QSAR in Medicinal Chemistry, (pp. 317-339): Apple Academic Press.
- Rojas, C., Tripaldi, P., & Duchowicz, P. R. (2016b). A New QSPR Study on Relative Sweetness. *International Journal of Quantitative Structure-Property Relationships*, 1(1), 78-92.
- Rojas, C., Ballabio, D., Consonni, V., Tripaldi, P., Mauri, A., & Todeschini, R. (2016c). Quantitative Structure–Activity Relationships to Predict Sweet and Non-Sweet Tastes. *Theoretical Chemistry Accounts*, 135:66, 1-13.
- Rojas, C., Todeschini, R., Ballabio, D., Mauri, A., Consonni, V., Tripaldi, P., & Grisoni, F. (2017). A QSTR-Based Expert System to Predict Sweetness of Molecules. *Frontiers in Chemistry*, 5(53), 1-12.
- Sahigara, F., Ballabio, D., Todeschini, R., & Consonni, V. (2013). Defining a Novel k-Nearest Neighbours Approach to Assess the Applicability Domain of a QSAR Model for Reliable Predictions. *Journal of cheminformatics*, 5(1), 27.
- Shallenberger, R. S., & Acree, T. E. (1967). Molecular Theory of Sweet Taste. *Nature*, 216, 480-482.
- Shallenberger, R. S. (1993). *Taste Chemistry*: Springer Science & Business Media.
- Shamil, S., Birch, G., Mathlouthi, M., & Clifford, M. (1987). Apparent Molar Volumes and Tastes of Molecules with More than One Sapophore. *Chemical senses*, 12(2), 397-409.
- Singh, R. K., Khan, M. A., & Singh, P. P. (2014). Rating of Sweetness by Molar Refractivity and Ionization Potential: QSAR Study of Sucrose and Guanidine Derivatives. *South African Journal of Chemistry*, 67, 12-20.
- Spillane, W. J., & McGlinchey, G. (1981). Structure-Activity Studies on Sulfamate Sweeteners II: Semiquantitative Structure-Taste Relationship for Sulfamate (RNHSO<sub>3</sub><sup>-</sup>) Sweeteners-The Role of R. *Journal of pharmaceutical sciences*, 70(8), 933-935.
- Spillane, W. J., McGlinchey, G., Muircheartaigh, I. Ó., & Benson, G. A. (1983). Structure-Activity Studies on Sulfamate Sweeteners III: Structure-Taste Relationships for Heterosulfamates. *Journal of pharmaceutical sciences*, 72(8), 852-856.
- Spillane, W. J., & Sheahan, M. B. (1989). Semi-Quantitative and Quantitative Structure-Taste Relationships for Carboand Hetero-Sulphamate (RNHSO<sub>3</sub><sup>-</sup>) Sweeteners. *Journal of the Chemical Society, Perkin Transactions 2*(7), 741-746.
- Spillane, W. J., & Sheahan, M. (1991). Structure-Taste Relationships for Sulfamate Sweeteners (RNHSO<sub>3</sub><sup>-</sup>). *Phosphorus, Sulfur, and Silicon and the Related Elements*, 59(1-4), 255-258.
- Spillane, W. J., Sheahan, M. B., & Ryder, C. A. (1993). Synthesis and Taste Properties of Sodium Disubstituted Phenylsulfamates. Structure-Taste Relationships for Sweet and Bitter/Sweet Sulfamates. *Food chemistry*, 47(4), 363-369.

- Spillane, W. J., Ryder, C. A., Walsh, M. R., Curran, P. J., Concagh, D. G., & Wall, S. N. (1996). Sulfamate Sweeteners. *Food chemistry*, 56(3), 255-261.
- Spillane, W. J., Ryder, C. A., Curran, P. J., Wall, S. N., Kelly, L. M., Feeney, B. G., & Newell, J. (2000). Development of Structure-Taste Relationships for Sweet and Non-Sweet Heterosulfamates. *Journal of the Chemical Society, Perkin Transactions 2*(7), 1369-1374.
- Spillane, W. J., Feeney, B. G., & Coyle, C. M. (2002). Further Studies on the Synthesis and Tastes of Monosubstituted Benzenesulfamates. A Semi-Quantitative Structure-Taste Relationship for the Meta-Compounds. *Food chemistry*, 79(1), 15-22.
- Spillane, W. J., Kelly, L. M., Feeney, B. G., Drew, M. G., & Hattotuwagama, C. K. (2003). Synthesis of Heterosulfamates. Search for Structure-Taste Relationships. *Arkivoc*, 7, 297-309.
- Spillane, W. J., Coyle, C. M., Feeney, B. G., & Thompson, E. F. (2009). Development of Structure-Taste Relationships for Thiazolyl-, Benzothiazolyl-, and Thiadiazolylsulfamates. *Journal of Agricultural and Food Chemistry*, 57(12), 5486-5493.
- Stewart, J. J. P. (2016). MOPAC2016, <http://OpenMOPAC.net>. USA: Stewart Computational Chemistry.
- Takahashi, Y., Miyashita, Y., Tanaka, Y., Abe, H., & Sasaki, S. (1982). A Consideration for Structure-Taste Correlations of Perillartines Using Pattern-Recognition Techniques. *Journal of Medicinal Chemistry*, 25(10), 1245-1248.
- Takahashi, Y., Abe, H., Miyashita, Y., Tanaka, Y., Hayasaka, H., & Sasaki, S. I. (1984). Discriminative Structural Analysis Using Pattern Recognition Techniques in the Structure-Taste Problem of Perillartines. *Journal of pharmaceutical sciences*, 73(6), 737-741.
- Talete srl. (2015). Dragon version 6. Software for Molecular Descriptor Calculation. <http://www.talete.mi.it/>.
- Tao, S., Jiao, X., Chen, S., Liu, W., Coveney, R., Zhu, L., & Luo, Y. (2006). Accumulation and Distribution of Polycyclic Aromatic Hydrocarbons in Rice (*Oryza sativa*). *Environmental Pollution*, 140(3), 406-415.
- Tarko, L., Lupescu, I., & Groposila-Constantinescu, D. (2005). Sweetness Power QSARs by PRECLAV Software. *Arkivoc*, 10, 254-271.
- Tarko, L., Lupescu, I., & Constantinescu-Groposila, D. (2006). QSAR Studies on Amino-Succinamic Acid Derivatives Sweeteners. *Arkivoc*, 13, 22-40.
- Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*: Wiley-VCH.
- Todeschini, R., Ballabio, D., Cassotti, M., & Consonni, V. (2015). N3 and BNN: Two New Similarity Based Classification Methods in

- Comparison with Other Classifiers. *Journal of chemical information and modeling*, 55(11), 2365-2374.
- van Rhee, A. M. (2003). Use of Recursion Forests in the Sequential Screening Process: Consensus Selection by Multiple Recursion Trees. *Journal of chemical information and computer sciences*, 43(3), 941-948.
- Vepuri, S. B., Tawari, N. R., & Degani, M. S. (2007). Quantitative Structure-Activity Relationship Study of Some Aspartic Acid Analogues to Correlate and Predict Their Sweetness Potency. *QSAR & Combinatorial Science*, 26(2), 204-214.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130.
- Yan, J., Huang, J.-H., He, M., Lu, H.-B., Yang, R., Kong, B., Xu, Q.-S., & Liang, Y.-Z. (2013). Prediction of Retention Indices for Frequently Reported Compounds of Plant Essential Oils Using Multiple Linear Regression, Partial Least Squares, and Support Vector Machine. *Journal of Separation Science*, 36(15), 2464-2471.
- Yang, X., Chong, Y., Yan, A., & Chen, J. (2011). In-Silico Prediction of Sweetness of Sugars and Sweeteners. *Food chemistry*, 128(3), 653-658.
- Yuasa, Y., Nagakura, A., & Tsuruta, H. (1994). The Sweetness and Stereochemistry of L-Aspartyl-Fenchylaminoalcohol Derivatives. *Tetrahedron Letters*, 35(37), 6891-6894.
- Zhong, M., Chong, Y., Nie, X., Yan, A., & Yuan, Q. (2013). Prediction of Sweetness by Multilinear Regression Analysis and Support Vector Machine. *Journal of Food Science*, 78(9), S1445-S1450.



---

# CONCLUSIONES GENERALES Y PROYECCIONES FUTURAS

---

En la presente tesis doctoral se ha estudiado y aplicado la teoría QSAR/QSPR para generar relaciones predictivas para actividades/propiedades de interés en Química Analítica y Química de los Alimentos.

Para garantizar que las predicciones de los modelos desarrollados a lo largo de esta tesis sean confiables, se ha seguido el protocolo recomendado por la Organización para la Cooperación Económica y el Desarrollo (OECD), el cual estipula que se deben considerar 5 principios básicos: actividad/propiedad definida, uso de un algoritmo inequívoco, definición del dominio de aplicabilidad del modelo, medida apropiada de la calidad del modelo (ajuste, robustez y predictividad) e interpretación del mecanismo de acción de los descriptores involucrados en el modelo.

De esta manera, se han utilizado descriptores moleculares proporcionados por el programa DRAGON como representación de la estructura molecular, los cuales fueron seleccionados mediante el Método de Reemplazo (RM) para los modelos continuos y Algoritmos Genéticos (GAs) para los modelos discretos. En cada modelo de regresión se ha definido el dominio de aplicabilidad (AD) mediante el criterio del valor de influencia, mientras que en los modelos de clasificación se definió el AD mediante el análisis de similitudes de los  $k$ -vecinos más cercanos ( $k$ NN). Finalmente, el mecanismo de acción de los descriptores se realizó mediante el análisis de los coeficientes de regresión y el análisis de componentes principales.

La preparación y curado de los datos ha sido un aspecto fundamental para garantizar la confiabilidad en la construcción de los modelos QSAR/QSPR. Por ello se ha automatizado esta etapa mediante el uso del programa de minería de datos KNIME, tanto para el curado de bases de datos de índices de retención como la discriminación del dulzor y no dulzor, así como la predicción del dulzor relativo. Esto ha permitido definir bases de datos confiables y que cubren un amplio espectro de diversidad molecular, lo que permite una mayor generalización de los modelos desarrollados.

Los modelos QSPR independientes de la conformación son un enfoque útil cuando se trabaja con bases de datos de índices de retención cromatográficos de compuestos orgánicos volátiles. De hecho, la consideración de los descriptores moleculares 3D para el desarrollo de los modelos no refleja mejoras significativas que justifiquen su inclusión. Así, para esta propiedad, los descriptores moleculares que

brindan información de la conectividad molecular brindan una buena correlación tanto en fases apolares como en fases polares.

Por otra parte, diferentes métodos de clasificación (lineal y no lineal basado en similitudes locales) han sido usados para establecer las relaciones cuantitativas para la discriminación de moléculas dulces y no dulces (amargas e insípidas), y la predicción del dulzor relativo de diversos edulcorantes. Particularmente, los modelos basados en similitudes locales *k*NN usando descriptores moleculares independientes de la conformación han mostrado buenos resultados predictivos para la discriminación del dulzor. Adicionalmente, se ha desarrollado por primera vez un modelo para la predicción del dulzor que contempla una representación más general de la estructura molecular, la cual está basada en el uso de las huellas digitales moleculares de conectividad ampliada (ECFPs) y descriptores moleculares independientes de la conformación. Para analizar dichos datos, se ha recurrido a diversas técnicas quimiométricas, tales como el escalado multidimensional, los métodos de clasificación de los *N*-vecinos más cercanos (N3), el análisis discriminante de mínimos cuadrados parciales (PLSDA) y el análisis de componentes principales (PCA).

Los modelos QSAR/QSPR independientes de la conformación son un enfoque válido y útil para desarrollar relaciones cuantitativas estructura-actividad/propiedad predictivas. Para su uso, es importante un curado minucioso de las bases de datos, de tal forma que se evite trabajar con moléculas que presenten la misma estructura molecular no conformacional y diferente valor de actividad/propiedad. El proceso de curado permite su identificación y el aplicar tratamientos específicos para dichas estructuras, según se trabaje en regresión o clasificación. Cuando se trata de una respuesta continua (regresión) es una práctica común el uso del valor medio de la actividad/propiedad para todas las moléculas que tienen la misma representación independiente de la conformación. Por otro lado, cuando se trata de una respuesta discreta (clasificación), se distinguen dos casos: 1) retener una sola estructura cuando existen diversos compuestos con la misma representación molecular y 2) excluir las estructuras moleculares ambiguas, es decir, aquellas representaciones moleculares no conformacionales que pertenecen a clases diferentes.

Por otra parte, el modelado QSAR/QSPR independiente de la conformación no requiere información adicional experimental de la forma en la que interactúa el ligando con el receptor (docking molecular), de la forma en que se optimizan las geometrías moleculares mediante diversos métodos computacionales disponibles (QSAR/QSPR 3D), de las energías de interacción en el espacio 3D entre la molécula y grupos específicos (QSAR/QSPR 3D-CoMFA) ni de las regiones de la molécula con afinidad/no afinidad en presencia de dicho grupo (QSAR/QSPR 3D-CoMSIA).

Existe aún mucho por explorar dentro del desarrollo de este trabajo de tesis doctoral, particularmente el desarrollo de nuevos modelos que contemplen: 1) uso de nuevos descriptores moleculares, tales como descriptores flexibles implementados en el programa CORAL, huellas digitales moleculares implementadas en el programa PADEL, así como nuevos descriptores topológicos propuestos en el programa QuBiLS-MAS; 2) utilización de otros métodos quimiométricos, por ejemplo, regresión de mínimos cuadrados parciales (PLS), redes neuronales de contrapropagación (CP-ANN), método de árboles de clasificación y regresión (CART), máquinas de soporte vectorial (SVM), funciones potenciales (estimadores de densidad de Kernel), y otros métodos basados en máquinas de aprendizaje; y 3) uso de nuevas bases de datos con las mismas actividades/propiedades que se han estudiado para fusionarlas y poder generar modelos QSAR/QSPR que cubran un mayor espectro del dominio químico.

Finalmente, existen también otras propiedades/actividades que se podrían estudiar dentro del campo de la Química Analítica: índices de retención medidos en otras fases estacionarias apolares (SE-30, DB-1, OV-1, CP-Sil-5CB), ligeramente polares (DB-5, HP-5, SE-52, SE-54, BPX-5) y polares (DB-Wax), propiedades espectroscópicas, propiedades de transición vítrea, propiedades calorimétricas, estabilidad térmica y solubilidad, entre otras. De igual manera en el campo de la Química de los Alimentos se podrían estudiar: péptidos bioactivos, capacidad antioxidante, actividad bactericida y fungicida de aditivos alimentarios, mejoradores de sabor/olor para alimentos, estudio de compuestos activos/nutraceuticos, toxicidad y carcinogenicidad de aditivos, QSAR para suplementos alimenticios y productos naturales, entre los más importantes.



---

## **PUBLICACIONES Y TRABAJOS PRESENTADOS EN EVENTOS CIENTÍFICOS**

---

El presente trabajo de tesis doctoral dio lugar a las siguientes publicaciones científicas:

1. C. Rojas, PR. Duchowicz y R. Pis Diez. Análisis QSPR de índices de retención de aromas medidos en cromatografía de gases. *Investigación Joven*, Vol. 1 (2) (2014): 61
2. F. Cárdenas, P. Tripaldi y C. Rojas. Estudio de la relación cuantitativa estructura–actividad de pesticidas mediante técnicas de clasificación. *Avances en Ciencias e Ingenierías*, Vol. 6 (2) (2014): B26–B37
3. C. Rojas, PR. Duchowicz, P. Tripaldi and R. Pis Diez. QSPR analysis for the retention index of flavors and fragrances on a OV–101 column. *Chemometrics and Intelligent Laboratory Systems*, Vol. 140 (2015): 126–132
4. C. Rojas, PR. Duchowicz, P. Tripaldi, R. Pis Diez. Quantitative structure–property relationship Analysis for the Retention Index of Fragrance–Like Compounds on a Polar Stationary Phase. *Journal of Chromatography A*. Vol. 1422 (2015): 277–288
5. C. Rojas, PR. Duchowicz, R. Pis Diez, P. Tripaldi. Applications of quantitative structure–relative sweetness relationships in food chemistry. In *Handbook of Research on Chemometrics: QSAR in Medicinal Chemistry (Vol I)*. Edited by Mercader, AG., Duchowicz, PR. and Sivakumar, PM. Apple Academic Press (2016)
6. C. Rojas, P. Tripaldi, PR. Duchowicz. A new QSPR study on relative sweetness. *International Journal of QSPR*. Vol. 1 (1) (2016): 76–90
7. C. Rojas, D. Ballabio, P. Tripaldi, V. Consonni, A. Mauri, R. Todeschini. Quantitative structure–activity relationships to predict sweet and non–sweet tastes. *Theoretical Chemistry Accounts*. Vol. 135 (3) (2016): 1–13
8. C. Rojas, R. Todeschini, D. Ballabio, A. Mauri, V. Consonni, P. Tripaldi, F. Grisoni. A QSTR–based Expert System to Predict

Sweetness of Molecules. *Frontiers in Chemistry*. 5(53) (2017): 1–12

9. C. Rojas, P.R. Duchowicz, P. Tripaldi, P. Diez, R. Quantitative Structure–Property Relationships for Predicting the Retention Indices of Fragrances on Stationary Phases of Different Polarity. *Journal of the Argentine Chemical Society*, In Press (2017)

así como a los siguientes trabajos presentados en congresos nacionales e internacionales:

1. F. Cárdenas, P. Tripaldi, D. Matamoros y C. Rojas. Estudio cuantitativo de la relación estructura–actividad de pesticidas empleando el algoritmo de clasificación  $k$ -NN. Sexta Semana Internacional de la Estadística y la Probabilidad. Puebla, México. Junio 2013
2. F. Cárdenas, P. Tripaldi y C. Rojas. Modelización de la estructura–actividad de pesticidas utilizando los métodos CP-ANN y  $k$ -NN. III Congreso Binacional AUSENP. Lambayeque, Perú. Octubre 2013
3. C. Rojas. Análisis QSPR sobre índices de retención de aromas y fragancias medidos por cromatografía de gases. Seminario de la División Química Teórica del INIFTA. La Plata, Argentina. Abril 2014.
4. F. Cárdenas, P. Tripaldi y C. Rojas. Uso de técnicas estadísticas multivariantes aplicadas a un estudio de calidad del agua. VII Congreso Iberoamericano de Física y Química Ambiental y XII Encuentro de Química Analítica y Ambiental. Viña del Mar, Chile. Octubre 2014
5. C. Rojas. Desarrollo y aplicación de técnicas matemáticas de la teoría QSAR–QSPR en química analítica y química de los alimentos. IX Jornadas de Becarios del INIFTA–2014. La Plata, Argentina. Octubre 2014
6. P. Tripaldi y C. Rojas. Funciones de deseabilidad como función de puntuación (fitness function) en la utilización de los algoritmos genéticos para la selección de variables. 30° Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2014
7. C. Rojas, P.R. Duchowicz, P. Tripaldi y R. Pis Diez. Análisis QSPR para índices de retención medidos en la columna OV–101. 30° Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2014
8. C. Rojas, P.R. Duchowicz, P. Tripaldi y R. Pis Diez. Relación cuantitativa estructura–dulzor relativo. V Congreso de Ingeniería

- de Alimentos y XIV Jornadas de Ciencia y Tecnología de Alimentos. Quito, Ecuador. Noviembre 2014
9. P. Tripaldi, C. Rojas y A. Pérez. Clasificación del poder edulcorante de moléculas naturales y sintéticas mediante aplicación de métodos QSAR. V Congreso de Ingeniería de Alimentos y XIV Jornadas de Ciencia y Tecnología de Alimentos. Quito, Ecuador. Noviembre 2014
  10. C. Rojas, P. Duchowicz, P. Tripaldi y R. Pis Diez. Análisis QSPR para índices de retención medidos en la columna Carbowax 20M. XIX Congreso Argentino de Físicoquímica y Química Inorgánica. Buenos Aires, Argentina. Abril 2015
  11. C. Rojas, P. Duchowicz, R. Pis Diez, P. Tripaldi, A. Pérez. A preliminary quantitative structure–relative sweetness relationship. 13th Latin American Conference on Physical Organic Chemistry. Carlos Paz, Argentina. Mayo 2015
  12. C. Rojas, D. Ballabio, P. Duchowicz, R. Todeschini, P. Tripaldi. QSAR study for modelling the sweetness and bitterness tastes. Congress of Theoretical Chemists of Latin Expression. Torino, Italia. Julio 2015
  13. C. Rojas. Relaciones cuantitativas estructura–actividad para predecir los gustos dulce y no dulce. Seminario de la División Química Teórica del INIFTA. La Plata, Argentina. Abril 2016
  14. C. Rojas, P. Tripaldi, P. Duchowicz, R. Pis Diez, A. Pérez. Modelado QSPR de índices de retención de aromas de arroz. Primer Workshop Latinoamericano de Modelado Molecular & Simulación Computacional. Buenos Aires, Argentina. Julio 2016
  15. P. Tripaldi, C. Rojas, A. Pérez, Davide Ballabio, R. Todeschini. Desarrollo de un modelo QSAR–N3 para péptidos inhibidores de acetilcolinesterasa (AChE). XXXI Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2016
  16. M. Valdez, P. Tripaldi, A. Pérez, C. Rojas. Caracterización reológica de mayonesas comerciales mediante análisis de componentes principales. XXXI Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2016
  17. A. Pérez, P. Tripaldi, C. Morales, C. Rojas. Estudio de la calidad físicoquímica de mieles de abeja comerciales mediante métodos quimiométricos. XXXI Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2016
  18. C. Rojas, R. Todeschini, D. Ballabio, A. Mauri, V. Consonni, P. Tripaldi, F. Grisoni. Relaciones cuantitativas estructura–dulzor

basadas en los métodos de clasificación N3 y PLSDA. XXXI Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2016

19. C. Rojas, P. Duchowicz, P. Tripaldi, P. Pis Diez. Desarrollo de modelos QSPR para índices de retención de aromas medidos en columnas de diferente polaridad. XXXI Congreso Argentino de Química. Buenos Aires, Argentina. Octubre 2016
20. C. Rojas, R. Todeschini, D. Ballabio, A. Mauri, V. Consonni, P. Tripaldi, F. Grisoni Quantitative structure–taste relationship to predict the sweetness of molecules. International Workshop on Drug Design and Neglected Tropical Diseases. La Plata, Argentina. November 2016

---

## TABLAS ANEXAS (en CD)

---

**Tabla 1A.** Nombres químicos e índices de retención experimentales de 1208 sabores y aromas medidos en la columna apolar OV-101

| No.             | Nombre                                   | I      |
|-----------------|--|--------|
| 1               | Acetaldehyde di-(cis-3-hexenil) acetal   | 1465   |
| 2 <sup>^</sup>  | Acetaldehyde ethyl-cis-3-hexenyl-acetal  | 1095   |
| 3               | Acetaldehyde linalyl ethyl acetal        | 1361.5 |
| 4 <sup>*</sup>  | Acetaldehyde phenylethyl n-propyl acetal | 1424   |
| 5 <sup>^</sup>  | Acetaldehyde styleneglycol acetal        | 1279.5 |
| 6               | Acetone                                  | 530    |
| 7               | Acetone-1-phenyl 1,2-ethandiol ketal     | 1260   |
| 8 <sup>*</sup>  | Acetonyl acetone                         | 894    |
| 9               | Acetophenone                             | 1048   |
| 10              | 4-Acetyl-6-tert-butyl-1,1-dimethyl indan | 1706   |
| 11 <sup>*</sup> | Acetyl eugenol eugenyl acetate           | 1541   |
| 12              | 2-Acetylfuran                            | 892    |
| 13              | Acetyl methyl carbinol acetoin           | 697    |
| 14 <sup>^</sup> | 2-Acetylpyridine                         | 1014   |
| 15              | 2-Acetylthiazole                         | 995    |
| 16 <sup>*</sup> | 2-Acetylthiophene                        | 1069   |
| 17 <sup>^</sup> | Allyl acetate                            | 675    |
| 18 <sup>^</sup> | Allyl anthranilate                       | 1411   |
| 19 <sup>*</sup> | Allyl benzoate                           | 1239   |
| 20 <sup>*</sup> | Allyl n-butyrate                         | 867    |
| 21 <sup>^</sup> | Allyl cyclohexane propionate             | 1405   |
| 22 <sup>*</sup> | Allyl n-decanoate                        | 1462   |
| 23 <sup>^</sup> | Allyl 2-ethyl butyrate                   | 995    |
| 24 <sup>*</sup> | Allyl formate                            | 586    |
| 25 <sup>*</sup> | Allyl 2-furoate                          | 1113   |
| 26 <sup>^</sup> | Allyl n-heptanoate                       | 1164   |
| 27 <sup>^</sup> | Allyl n-hexanoate                        | 1065   |
| 28 <sup>*</sup> | Allyl ionone                             | 1676.8 |
| 29 <sup>^</sup> | Allyl levulinate                         | 1114   |
| 30 <sup>*</sup> | Allyl n-nonanoate                        | 1365   |
| 31 <sup>^</sup> | Allyl n-octanoate                        | 1265   |

---

---

|     |                                      |      |
|-----|--------------------------------------|------|
| 32^ | Allyl phenoxy acetate                | 1429 |
| 33^ | Allyl phenyl acetate                 | 1422 |
| 34^ | Allyl n-propionate                   | 777  |
| 35  | Allyl salicylate                     | 1339 |
| 36* | Allyl tiglate                        | 1002 |
| 37^ | Allyl iso-valerate                   | 920  |
| 38^ | o-Amino acetophenone                 | 1288 |
| 39  | p-Amino acetophenone                 | 1506 |
| 40* | n-Amyl acetate                       | 985  |
| 41  | iso-Amyl acetate                     | 860  |
| 42* | 2-n-Amyl-3-acetonyl-1-cyclopentanone | 1600 |
| 43  | n-Amyl alcohol                       | 756  |
| 44^ | iso-Amyl alcohol                     | 719  |
| 45  | sec-Amyl alcohol                     | 685  |
| 46  | tert-Amyl alcohol                    | 631  |
| 47* | iso-Amyl benzoate                    | 1421 |
| 48  | iso-Amyl benzyl ether                | 1297 |
| 49^ | n-Amyl n-butyrate                    | 1078 |
| 50^ | n-Amyl iso-butyrate                  | 1035 |
| 51  | iso-Amyl n-butyrate                  | 1042 |
| 52  | iso-Amyl iso-butyrate                | 997  |
| 53^ | iso-Amyl cinnamate                   | 1719 |
| 54* | alpha-Amyl cinnamic aldehyde         | 1631 |
| 55* | alpha-n-Amyl cinnamyl acetate        | 1757 |
| 56^ | n-Amyl cyclopentenone                | 1180 |
| 57  | p-tert-amyl cyclohexanone            | 1334 |
| 58^ | iso-Amyl n-decanoate                 | 1633 |
| 59* | iso-Amyl n-dodecanoate               | 1829 |
| 60  | n-Amyl formate                       | 810  |
| 61* | iso-Amyl formate                     | 775  |
| 62^ | iso-Amyl 2-furoate                   | 1287 |
| 63  | n-amyl furylpropionate               | 1453 |
| 64* | iso-amyl furylpropionate             | 1415 |
| 65* | iso-amyl n-pentanoate                | 1334 |
| 66^ | n-amyl n-hexanoate                   | 1279 |
| 67^ | iso-amyl n-hexanoate                 | 1238 |
| 68* | n-amyl levulinate                    | 1325 |
| 69^ | iso-amyl levulinate                  | 1284 |
| 70^ | n-amyl 2-methyl butyrate             | 1126 |
| 71  | iso-amyl 2-methyl butyrate           | 1087 |

---

---

|      |                                      |      |
|------|--------------------------------------|------|
| 72*  | n–amyl n–octanoate                   | 1471 |
| 73*  | iso–amyl n–octanoate                 | 1433 |
| 74^  | n–amyl phenylacetate                 | 1506 |
| 75^  | iso–amyl phenylacetate               | 1468 |
| 76*  | n–amyl n–propionate                  | 955  |
| 77^  | iso–amyl n–propionate                | 954  |
| 78^  | 6–amyl alpha–pyrone                  | 1434 |
| 79   | iso–amyl pyruvate                    | 1027 |
| 80*  | n–amyl salicylate                    | 1557 |
| 81*  | iso–amyl salicylate                  | 1528 |
| 82*  | iso–amyl tiglate                     | 1178 |
| 83   | iso–amyl 10–undecenoate              | 1717 |
| 84^  | n–amyl n–valerate                    | 1185 |
| 85*  | n–amyl iso–valerate                  | 1135 |
| 86^  | iso–amyl n–valerate                  | 1138 |
| 87^  | Isoamyl isovalerate                  | 1092 |
| 88*  | anethole                             | 1270 |
| 89   | anis alcohol                         | 1267 |
| 90*  | anis aldehyde                        | 1234 |
| 91^  | anis aldehyde–propyleneglycol acetal | 1510 |
| 92*  | anisole                              | 900  |
| 93^  | benzal acetone                       | 1337 |
| 94   | benzaldehyde                         | 947  |
| 95*  | benzotrile                           | 965  |
| 96*  | benzophenone                         | 1604 |
| 97   | Benzyl acetate                       | 1144 |
| 98*  | Benzyl acetone                       | 1218 |
| 99^  | Benzyl alcohol                       | 1033 |
| 100* | Benzyl n–butyrate                    | 1322 |
| 101* | Benzyl iso–butirate                  | 1277 |
| 102* | Benzyl ethyl ether                   | 1046 |
| 103^ | Benzyl formate                       | 1058 |
| 104^ | Benzyl n–heptanoate                  | 1620 |
| 105^ | Benzyl n–hexanoate                   | 1521 |
| 106^ | Benzyl n–nonanoate                   | 1823 |
| 107  | Benzyl n–octanoate                   | 1720 |
| 108^ | Benzyl n–propionate                  | 1234 |
| 109^ | Benzyl tiglate                       | 1474 |
| 110  | Benzyl n–valerate                    | 1421 |
| 111  | Benzyl iso–valerate                  | 1374 |

---

---

|      |   |       |
|------|---|-------|
| 112* | iso-bornyl acetate                                | 1279  |
| 113* | iso-bornyl n-propionate                           | 1376  |
| 114  | bromostryrol                                      | 1197  |
| 115* | 2,3 butadienone diacetyl                          | 606   |
| 116  | n butanol   | 655   |
| 117^ | n-butanol-2-ol                                    | 586   |
| 118^ | 2-butanoylfuran                                   | 1078  |
| 119^ | 2-butanoyl-5-methylfuran                          | 1192  |
| 120  | 2-butanoylthiophene                               | 1252  |
| 121^ | n-butyl acetate                                   | 793   |
| 122  | iso-butyl acetate                                 | 758   |
| 123* | iso-butyl aldehyde                                | 500   |
| 124  | iso-butyl aldehyde-1-phenyl,1,2-ethanediol acetal | 1384  |
| 125* | iso-butyl aldehyde-propylene glycol acetal        | 835.5 |
| 126^ | n-butyl anisate                                   | 1632  |
| 127^ | iso-butyl anisate                                 | 1586  |
| 128  | n-butyl antranilate                               | 1600  |
| 129* | iso-butyl antranilate                             | 1556  |
| 130^ | n-butyl benzoate                                  | 1354  |
| 131  | iso-butyl benzoate                                | 1310  |
| 132* | iso-butyl benzyl carbinol                         | 1366  |
| 133^ | n-butyl n-butyrate                                | 979   |
| 134* | n-butyl iso-butyrate                              | 939   |
| 135  | iso-butyl n-butyrate                              | 941   |
| 136* | iso-butyl iso-butyrate                            | 900   |
| 137^ | n butyl n-butyru lactate                          | 1331  |
| 138* | iso-butyl cinnamate                               | 1605  |
| 139* | o-ter-butyl cyclohexyl acetate                    | 1295  |
| 140* | p-tert-butyl cyclohexyl-acetate trans             | 1322  |
| 141^ | p-tert-butyl cyclohexyl-acetate cis               | 1360  |
| 142  | p-tert-butyl cyclohexanone                        | 1208  |
| 143* | n-Butyl n-decanoate                               | 1575  |
| 144^ | 2-iso-Butyl-4,5-dimethyloxazole                   | 1044  |
| 145^ | 2-n-Butyl-4-5-dimethylthiazole                    | 1251  |
| 146* | 2-iso-Butyl-4,5-dimethylthizole                   | 1193  |
| 147^ | n-Butyl n-dodecanoate                             | 1772  |
| 148  | iso-Butyl 2-ethyl n-hexanoate                     | 1248  |
| 149* | 2-n-Butyl-4-ethyl-5-methyloxazole                 | 1159  |
| 150^ | 2-iso-Butyl-4-ethyl-5-methyloxazole               | 1106  |
| 151^ | n-Butyl formate                                   | 696   |

---

---

|      |  |        |
|------|--|--------|
| 152  | iso-Butyl formate  | 673    |
| 153^ | 2-n-Butylfuran   | 883    |
| 154* | iso-Butyl beta-2-furyl acrylate                                  | 1435   |
| 155^ | n-Butyl furyl n-propionate                                       | 1354   |
| 156^ | n-Butyl n-heptanoate   | 1276   |
| 157* | iso-Butyl n-heptanoate   | 1239   |
| 158^ | n-Butyl n-hexanoate  | 1177   |
| 159  | iso-Butyl n-hexanoate  | 1140   |
| 160* | tert-Butyl-4-hydroxy anisole                                     | 1480   |
| 161  | n-Butyl levulinate   | 1225   |
| 162* | iso-Butyl levulinate   | 1183   |
| 163^ | iso-Butyl n-methyl anthranilate                                  | 1617   |
| 164* | n-Butyl 2-methyl butyrate  | 1029   |
| 165^ | iso-Butyl 2-methyl butyrate                                      | 991    |
| 166* | p-tert-Butyl-alpha-methyl-hydrocinnamic aldehyde (lily aldehyde) | 1506   |
| 167  | n-Butyl methyl phenyl-glycidate                                  | 1645.5 |
| 168* | 2-n-Butyl-4-methylthiazole                                       | 1141   |
| 169  | 2-iso-Butyl-4-methylthiazole                                     | 1086   |
| 170  | n-Butyl n-octanoate  | 1373   |
| 171^ | iso-Butyl n-octanoate  | 1338   |
| 172^ | n-Butyl phenyl acetate   | 1408   |
| 173^ | iso-Butyl phenyl acetate   | 1371   |
| 174  | p-tert-Butyl phenylpropyl-aldehyde                               | 1212   |
| 175* | iso-Butyl pivalate   | 933    |
| 176  | n-Butyl n-propionate   | 889    |
| 177  | iso-Butyl n-propionate   | 852    |
| 178  | 6-sec-Butyl quinoline  | 1592   |
| 179^ | n-Butyl salicylate   | 1457   |
| 180* | iso-Butyl salicylate   | 1410   |
| 181^ | 2-iso-butylthiazole  | 1020   |
| 182* | 2-n-butylthiophene   | 1052   |
| 183  | n-Butyl tiglata  | 1116   |
| 184^ | iso-Butyl tiglata  | 1076   |
| 185  | n-Butyl n-undecanoate  | 1674   |
| 186* | n-Butyl 10-undecenoate   | 1660   |
| 187^ | n-Butyl iso-valerate   | 1021   |
| 188* | iso-Butyl iso-valerate   | 992    |
| 189  | iso-Butyrophenone  | 1185   |
| 190^ | Camphor  | 1136   |
| 191^ | Carvacrol  | 1297   |

---

---

|      |   |        |
|------|---|--------|
| 192* | Carveol                                     | 1215.5 |
| 193^ | Carvone                                     | 1228   |
| 194  | Caryophyllene                               | 1428   |
| 195^ | Cedrol                                      | 1609   |
| 196* | Cinnamic aldehyde                           | 1250   |
| 197* | Cinnamyl acetate                            | 1419   |
| 198* | Cinnamyl alcohol                            | 1300   |
| 199^ | Cinnamyl n-butyrate                         | 1604   |
| 200  | Cinnamyl iso-butyrate                       | 1555   |
| 201^ | Cinnamyl n-hexanoate                        | 1805   |
| 202^ | Cinnamyl n-propionate                       | 1515   |
| 203^ | Cinnamyl n-valerate                         | 1705   |
| 204* | Cinnamyl iso-valerate                       | 1655   |
| 205  | Citral                                      | 1236   |
| 206^ | Citronellal                                 | 1137   |
| 207^ | Citronellol                                 | 1215   |
| 208^ | Citronellyl acetate                         | 1335   |
| 209^ | Citronellyl n-butyrate                      | 1511   |
| 210  | Citronellyl iso-butyrate                    | 1469   |
| 211* | Citronellyl formate                         | 1261   |
| 212^ | Citronellyl n-propionate                    | 1427   |
| 213^ | Citronellyl iso-valerate                    | 1563   |
| 214  | Coumarin                                    | 1418   |
| 215  | p-Cresol methyl ether                       | 1005   |
| 216* | p-Cresyl acetate                            | 1150   |
| 217^ | p-Cresyl iso-butyrate                       | 1291   |
| 218^ | p-Cresyl n-octanoate                        | 1755   |
| 219* | Cumin alcohol                               | 1283   |
| 220  | Cumin aldehyde                              | 1227   |
| 221  | Cyclohexane glycidate                       | 1344   |
| 222  | Cyclohexanol                                | 880    |
| 223* | Cyclohexanone                               | 875    |
| 224  | Cyclohexanone-1,3-butylene ketal            | 1186   |
| 225  | Cyclohexanone-1-phenyl-1,2-ethanediol ketal | 1665   |
| 226^ | Cyclohexil acetate                          | 1027   |
| 227* | Cyclohexyl n-butyrate                       | 1209   |
| 228^ | Cyclohexyl iso-butyrate                     | 1164   |
| 229^ | 2-Cyclohexylethyl acetate                   | 1233   |
| 230  | Cyclohexyl formate                          | 951    |
| 231* | Cyclohexyl n-hexanoate                      | 1411   |

---

---

|      |                                   |        |
|------|-----------------------------------|--------|
| 232^ | Cyclohexyl n-propionate           | 1120   |
| 233^ | Cyclohexyl iso-valerate           | 1264   |
| 234^ | trans-2-trans-4-Decadienol        | 1310   |
| 235* | Decahydro beta-naphthol           | 1301.5 |
| 236* | gamma-Decalactone                 | 1437   |
| 237  | delta-Decalactone                 | 1463   |
| 238  | n-Decanal                         | 1188   |
| 239  | n-Decanal diethyl acetal          | 1473   |
| 240* | n-Decanal dimethyl acetal         | 1366   |
| 241  | n-Decane                          | 1000   |
| 242* | n-Decanol                         | 1263   |
| 243* | cis-4-Decenal                     | 1177   |
| 244^ | 1-Decene                          | 991    |
| 245^ | 9-Decen-1-al                      | 1251   |
| 246  | 9-Decen-1-yl acetate              | 1383   |
| 247^ | n-Decyl acetate                   | 1393   |
| 248^ | n-Decyl n-propionate              | 1486   |
| 249  | Di-iso-amyl ether                 | 1000   |
| 250^ | Di-n-amyl ketone                  | 1258   |
| 251^ | Di-n-Butyl ether                  | 876    |
| 252* | Di-n-butyl fumarate               | 1558   |
| 253  | Di-n-butyl ketone                 | 1058   |
| 254  | Di-iso-butyl ketone               | 983    |
| 255^ | Diethyl adipate                   | 1358   |
| 256  | Diethyleneglycol diethyl ether    | 1058   |
| 257* | Diethyleneglycol dimethyl ether   | 924    |
| 258^ | Diethyleneglycol monoethyl ether  | 986    |
| 259  | Diethyleneglycol monomethyl ether | 920    |
| 260* | Diethyl ether                     | 572    |
| 261  | Diethyl ketone                    | 681    |
| 262  | Diethyl malonate                  | 1043   |
| 263^ | Diethyl-5-methyloxazole           | 983    |
| 264* | Diethyl phthalate                 | 1565   |
| 265* | Diethyl sebacate                  | 1752   |
| 266  | Diethyl suberate                  | 1553   |
| 267* | Diethyl succinate                 | 1153   |
| 268* | Di-n-hexyl ketone                 | 1458   |
| 269^ | Dihydro anethole                  | 1193   |
| 270^ | Dihydrocarveol                    | 1188   |
| 271  | Dihydro-nor-dicyclopentadienyl    | 1406   |

---

---

|                  |                                    |      |
|------------------|------------------------------------|------|
| 272 <sup>^</sup> | Dihydrio iso-jasmone               | 1374 |
| 273              | Dihydro myrcenol                   | 1063 |
| 274              | Dihydro terpinil acetate           | 1282 |
| 275 <sup>*</sup> | 1,4-Dimethoxy benzene              | 1145 |
| 276              | 1,4-Dimethoxy-2-tert-butylbenzene  | 1398 |
| 277 <sup>*</sup> | 2,4-Dimethyl-5-acetylthiazole      | 1217 |
| 278              | Dimethyl adipate                   | 1212 |
| 279 <sup>*</sup> | Dimethyl benzyl carbinol           | 1147 |
| 280 <sup>*</sup> | Dimethyl benzyl carbiny acetate    | 1302 |
| 281              | Dimethyl benzyl carbiny n-butyrate | 1476 |
| 282              | Dimethyl benzyl carbiny propionate | 1391 |
| 283 <sup>^</sup> | 2,5-Dimethyl-4-ethyloxazole        | 900  |
| 284 <sup>*</sup> | Dimethyl fumarate                  | 997  |
| 285 <sup>^</sup> | 2,5-Dimethylfuran                  | 697  |
| 286 <sup>^</sup> | 2,6-Dimethyl n-heptan-2-ol         | 983  |
| 287              | 2,6-Dimethyl hept-5-en-1-al        | 1039 |
| 288              | 2,6-Dimethylheptyl-4-acetate       | 1092 |
| 289              | Dimethyl malonate                  | 896  |
| 290 <sup>*</sup> | Dimethyl phenyl ethyl carbinol     | 1282 |
| 291 <sup>*</sup> | 2,3-Dimethyl pyrazine              | 900  |
| 292              | 2,5-Dimethyl pyrazine              | 893  |
| 293 <sup>^</sup> | Dimethyl sebacate                  | 1616 |
| 294              | Dimethyl succinate                 | 1002 |
| 295              | 2,4-Dimethylthiazole               | 869  |
| 296              | 4,5-Dimethylthiazole               | 917  |
| 297              | Diphenyl oxide                     | 1386 |
| 298 <sup>*</sup> | Di-n-propyl ketone                 | 857  |
| 299              | Di-iso-propyl ketone               | 783  |
| 300              | Di-n-propyl malonate               | 1227 |
| 301 <sup>^</sup> | gamma-Dodecalactone                | 1647 |
| 302              | delta-Dodecalactone                | 1675 |
| 303 <sup>*</sup> | n-Dodecanal                        | 1392 |
| 304 <sup>*</sup> | n-Dodecanal dimethyl acetal        | 1565 |
| 305 <sup>*</sup> | n-Dodecane                         | 1200 |
| 306 <sup>*</sup> | n-Dodecanol                        | 1468 |
| 307              | n-Eicosane                         | 2000 |
| 308              | Estragole                          | 1183 |
| 309 <sup>*</sup> | p-Ethoxy benzaldehyde              | 1308 |
| 310 <sup>*</sup> | Ethyl acetate                      | 595  |
| 311              | Ethyl acetyl acetate               | 907  |

---

---

|      |                                 |      |
|------|---------------------------------|------|
| 312^ | Ethyl acrylate                  | 681  |
| 313  | Ethyl alcohol                   | 500  |
| 314^ | Ethyl n-amy l ketone            | 928  |
| 315  | Ethyl anisate                   | 1426 |
| 316* | Ethyl anthranilate              | 1396 |
| 317^ | Ethyl benzoate                  | 1154 |
| 318  | Ethyl n-butyl ketone            | 869  |
| 319  | Ethyl n-butyrate                | 784  |
| 320* | Ethyl iso-butyr ate             | 746  |
| 321  | Ethyl cloro acetate             | 810  |
| 322^ | Ethyl cinnamate                 | 1443 |
| 323* | Ethyl crotonate                 | 823  |
| 324* | Ethyl n-decanoate               | 1379 |
| 325  | 2-Ethyl-4,5-dimethyloxazole     | 914  |
| 326  | 2-Ethyl-4,5-dimethylthiazole    | 1065 |
| 327* | Ethyl n-deodecanoate            | 1579 |
| 328^ | Ethyleneglycol monophenyl ether | 1213 |
| 329  | Ethyl formate                   | 530  |
| 330* | 2-Ethylfuran                    | 694  |
| 331  | Ethyl 2-furoate                 | 1029 |
| 332  | Ethyl n-heptanoate              | 1082 |
| 333  | Ethyl n-hexanoate               | 983  |
| 334* | Ethyl iso-hexanoate             | 951  |
| 335* | Ethyl levulinate                | 1029 |
| 336^ | Ethyl 2-methyl butyr ate        | 837  |
| 337^ | 2-Ethyl-4-methylthiazole        | 955  |
| 338* | 4-Ethyl-5-methylthiazole        | 991  |
| 339^ | Ethyl n-nonanoate               | 1280 |
| 340  | Ethyl n-octanoate               | 1180 |
| 341  | Ethyl phenyl acetate            | 1219 |
| 342  | Ethyl pivalate                  | 776  |
| 343^ | Ethyl n-propionate              | 691  |
| 344  | Ethyl n-propyl ketone           | 767  |
| 345^ | Ethyl salicylate                | 1257 |
| 346  | Ethyl n-tetradecanoate          | 1780 |
| 347  | 2-Ethylthiazole                 | 879  |
| 348^ | 2-Ethylthiophene                | 861  |
| 349* | Ethyl tiglate                   | 922  |
| 350* | Ethyl n-undecanoate             | 1479 |
| 351  | Ethyl 10-undecenoate            | 1469 |

---

---

|      |                                  |        |
|------|----------------------------------|--------|
| 352* | Ethyl n–valerate                 | 884    |
| 353* | Ethyl iso–valerate               | 840    |
| 354  | Ethyl vanillin                   | 1448   |
| 355^ | Eugenol                          | 1351   |
| 356  | iso–Eugenol                      | 1416   |
| 357  | Furfural                         | 815    |
| 358* | Furfuryl acetate                 | 969    |
| 359* | Furfuryl n–butyrate              | 1148   |
| 360  | Furfuryl n–hexanoate             | 1343   |
| 361* | Galaxolide                       | 1837   |
| 362* | Geraniol                         | 1243   |
| 363* | Gerano nitrile                   | 1217.5 |
| 364  | Geranyl acetate                  | 1364   |
| 365  | Geranyl n–butyrate               | 1532   |
| 366  | Geranyl iso–butyrate             | 1493   |
| 367^ | Geranyl ethyl ether              | 1265   |
| 368^ | Geranyl formate                  | 1282   |
| 369* | Geranyl n–hexanoate              | 1731   |
| 370  | Geranyl n–propionate             | 1424   |
| 371* | Geranyl iso–valerate             | 1593   |
| 372  | Helional                         | 1543   |
| 373* | Heliotropine                     | 1318   |
| 374  | n–Heneicosane                    | 2100   |
| 375* | n–Heptadecane                    | 1700   |
| 376* | n–Heptanal                       | 883    |
| 377^ | n–Heptanal diethyl acetal        | 1179   |
| 378* | n–Heptanal dimethyl acetal       | 1069   |
| 379* | n–Heptanal ethyleneglycol acetal | 1144   |
| 380  | n–Heptane                        | 700    |
| 381  | 2,3–Heptanedione                 | 816    |
| 382^ | n–Heptanol                       | 957    |
| 383^ | n–Heptan–2–ol                    | 888    |
| 384* | n–Heptan–4–ol                    | 879    |
| 385^ | 2–Heptanoylfuran                 | 1384   |
| 386  | 2–Heptanoylthiophene             | 1565   |
| 387* | 1–Hepten–3–ol                    | 868    |
| 388^ | n–Heptyl acetate                 | 1095   |
| 389* | n–Heptyl n–butyrate              | 1275   |
| 390^ | n–Heptyl iso butyrate            | 1233   |
| 391* | n–Heptyl formate                 | 1012   |

---

---

|      |                                       |      |
|------|---------------------------------------|------|
| 392  | 2-n-Heptylfuran                       | 1184 |
| 393^ | n-Heptyl n-octanoate                  | 1666 |
| 394^ | 2-n-Heptylthiophene                   | 1359 |
| 395^ | n-Hexadecane                          | 1600 |
| 396  | n-Hexanal                             | 780  |
| 397  | n-Hexanal diethyl acetal              | 1082 |
| 398* | n-Hexane                              | 600  |
| 399  | 3,4-Hexanedione                       | 777  |
| 400^ | n-Hexanol                             | 858  |
| 401^ | n-Hexan-2-ol                          | 786  |
| 402  | 2-Hexanoylfuran                       | 1281 |
| 403* | 2-Hexanoylthiophene                   | 1459 |
| 404* | trans-2-Hexenal                       | 832  |
| 405* | 1-Hexen-3-ol                          | 770  |
| 406  | cis-3-Hexenol                         | 847  |
| 407* | trans-2-Hexenol                       | 854  |
| 408^ | cis-3-Hexenyl acetate                 | 987  |
| 409* | trans-2-Hexenyl acetate               | 997  |
| 410  | cis-3-Hexenyl n-butyrate              | 1170 |
| 411^ | trans-2-Hexenyl n-butyrate            | 1180 |
| 412* | cis-3-Hexenyl iso-butyrate            | 1129 |
| 413  | cis-3-Hexenyl n-decanoate             | 1760 |
| 414* | cis-3-Hexenyl formate                 | 902  |
| 415  | cis-3-Hexenyl n-hexanoate             | 1370 |
| 416^ | trans-2-hexenyl n-hexanoate           | 1375 |
| 417  | cis-3-Hexenyl lactate                 | 1187 |
| 418* | cis-3-Hexenyl methoxy formate         | 1073 |
| 419* | cis-3-Hexanyl n-octanoate             | 1564 |
| 420^ | trans-2-Hexenyl n-octanoate           | 1573 |
| 421* | cis-3-Hexenyl n-propionate            | 1083 |
| 422  | trans-2-Hexenyl n-propionate          | 1085 |
| 423* | cis-3-Hexenyl salicylate              | 1654 |
| 424  | cis-3-Hexenyl n-valerate              | 1270 |
| 425* | cis-3-Hexenyl iso valerate            | 1223 |
| 426* | 2-Hexoxy acetaldehyde dimethyl acetal | 1234 |
| 427* | n-Hexyl acetate                       | 1012 |
| 428* | n-Hexyl n-butyrate                    | 1176 |
| 429^ | n-Hexyl iso-butyrate                  | 1135 |
| 430^ | alpha-n-Hexyl cinnamic aldehyde       | 1727 |
| 431^ | n-Hexyl formate                       | 994  |

---

---

|      |                            |      |
|------|----------------------------|------|
| 432  | 2-Hexylfuran               | 1083 |
| 433^ | n-Hexyl n-hexanoate        | 1371 |
| 434  | n-Hexyl 2-methyl butyrate  | 1224 |
| 435^ | n-Hexyl pivalate           | 1163 |
| 436^ | n-Hexyl n-propionate       | 1088 |
| 437  | n-Hexyl salicylate         | 1664 |
| 438* | 2-n-Hexylthiophene         | 1256 |
| 439* | n-Hexyl iso-valerate       | 1228 |
| 440  | Hydroxy citronellal        | 1269 |
| 441  | Indole                     | 1304 |
| 442  | alpha-Ionone               | 1416 |
| 443^ | Beta-Ionone                | 1474 |
| 444^ | cis-Jasmone                | 1378 |
| 445  | Lavandulyl acetate         | 1274 |
| 446^ | Limonene                   | 1030 |
| 447* | Linalool                   | 1092 |
| 448  | cis-Linalool oxide         | 1068 |
| 449  | trans-Linalool oxide       | 1082 |
| 450* | Linalyl acetate            | 1246 |
| 451  | Linalyl iso-butyrate       | 1366 |
| 452  | Linalyl formate            | 1206 |
| 453* | Linalyl n-hexanoate        | 1582 |
| 454^ | Linalyl n-propionate       | 1324 |
| 455^ | Linalyl n-valerate         | 1500 |
| 456* | Linalyl iso-valerate       | 1461 |
| 457^ | Menthol                    | 1171 |
| 458  | m-Methoxy acetophenone     | 1279 |
| 459* | o-Methoxy acetophenone     | 1269 |
| 460^ | p-Methoxy acetophenone     | 1327 |
| 461  | m-Methyl acetophenone      | 1156 |
| 462^ | o-Methyl acetophenone      | 1118 |
| 463  | p-Methyl acetophenone      | 1166 |
| 464  | 2-Methyl-5-acetylthiophene | 1185 |
| 465^ | Methyl n-amyl ketone       | 872  |
| 466* | Methyl iso-amyl ketone     | 825  |
| 467* | Methyl anisate             | 1354 |
| 468  | Methyl anthranilate        | 1332 |
| 469  | Methyl benzoate            | 1078 |
| 470^ | 3-Methylbutan-2-ol         | 671  |
| 471^ | Methyl n-butyl ketone      | 772  |

---

---

|      |   |        |
|------|---|--------|
| 472^ | Methyl iso-butyl ketone   | 725    |
| 473* | Methyl n-butyrate   | 705    |
| 474* | Methyl iso-butyrate   | 673    |
| 475^ | Methyl cinnamate  | 1365   |
| 476* | Methyl crotonate  | 745    |
| 477^ | Methyl n-decanoate  | 1307   |
| 478* | Methyl n-dodecanoate  | 1507   |
| 479  | 2-Methyl-5-ethylfuran   | 791    |
| 480^ | 2-Methyl-3-ethyl pyrazine   | 987    |
| 481^ | Methyl iso-eugenol  | 1447.5 |
| 482^ | 2-Methylfuran   | 614    |
| 483* | 5-Methyl furfural   | 942    |
| 484* | Methyl furoate  | 956    |
| 485^ | Methyl n-heptanoate   | 1006   |
| 486* | 6-Methyl hept-5-en-2-one  | 968    |
| 487  | Methyl n-heptyl ketone  | 1074   |
| 488* | Methyl n-hexanoate  | 906    |
| 489* | Methyl n-hexyl ketone   | 972    |
| 490  | Methyl n-hexyl ketone-1-phenyl-1,2-ethanediol ketal                 | 1748.5 |
| 491  | alpha-n-Methyl ionone   | 1530   |
| 492* | beta-n-Methyl ionone  | 1564   |
| 493  | alpha-iso-Methyl ionone   | 1471   |
| 494^ | beta-iso-Methyl ionone  | 1506   |
| 495* | Methyl levulinate   | 956    |
| 496^ | Methyl n-methyl anthranilate  | 1389   |
| 497  | Methyl 2-methyl butyrate  | 765    |
| 498* | beta-Methyl naphthyl ketone   | 1592   |
| 499* | Methyl n-nonanoate  | 1207   |
| 500  | Methyl n-nonyl ketone   | 1276   |
| 501* | Methyl n-octanoate  | 1107   |
| 502* | Methyl n-octyl ketone   | 1176   |
| 503^ | Methyl phenylacetate  | 1154   |
| 504  | Methyl phenyl carbinol  | 1051   |
| 505^ | 2-Methyl-1-propanol   | 616    |
| 506^ | 2-Methyl n-propan-2-ol  | 500    |
| 507^ | 2-Methyl-5-propionylthiophene                                       | 1280   |
| 508* | Methyl n-propyl ketone  | 672    |
| 509^ | beta-Methyl-p-iso-propyl phenyl propionaldehyde (cyclamen aldehyde) | 1444   |
| 510* | 2-Methyl-3-iso-propyl pyrazine                                      | 1028   |
| 511  | 2-Methyl pyrazine   | 805    |

---

---

|                  |                                 |        |
|------------------|---------------------------------|--------|
| 512 <sup>^</sup> | Methyl salicylate (wintergreen) | 1181   |
| 513 <sup>^</sup> | Methyl n-tetradecanoate         | 1707   |
| 514 <sup>^</sup> | 4-Methylthiazole                | 800    |
| 515              | 4-Methyl-5-thiazole ethanol     | 1283   |
| 516 <sup>*</sup> | Methyl p-toluate                | 1194   |
| 517              | 2-Methyl undecanal              | 1353   |
| 518              | Methyl 10-undecenoate           | 1396   |
| 519              | Methyl n-undecyl ketone         | 1479   |
| 520 <sup>^</sup> | Methyl n-valerate               | 806    |
| 521              | Methyl iso-valerate             | 764    |
| 522 <sup>*</sup> | Muscone                         | 1831   |
| 523              | Musk xylol                      | 1506   |
| 524              | (S)-Myrac aldehyde              | 1506   |
| 525 <sup>*</sup> | (R)-Myrac aldehyde              | 1509   |
| 526 <sup>^</sup> | Myrcene                         | 986    |
| 527              | Nerol                           | 1218   |
| 528              | Nerolidol                       | 1538.5 |
| 529 <sup>*</sup> | Neryl acetate                   | 1345   |
| 530 <sup>^</sup> | Neryl n-butyrate                | 1519   |
| 531              | Neryl iso-butyrate              | 1474   |
| 532 <sup>^</sup> | Neryl n-propionate              | 1436   |
| 533 <sup>*</sup> | n-Nonadecane                    | 1900   |
| 534              | gamma-Nonalactone               | 1328   |
| 535 <sup>*</sup> | delta-Nonalactone               | 1356   |
| 536              | n-Nonanal                       | 1087   |
| 537 <sup>*</sup> | n-Nonanal diethyl acetal        | 1374   |
| 538              | n-Nonanal dimethyl acetal       | 1267   |
| 539 <sup>*</sup> | n-Nonane                        | 900    |
| 540 <sup>^</sup> | n-Nonanol                       | 1161   |
| 541 <sup>*</sup> | n-Nonan-2-ol                    | 1089   |
| 542 <sup>^</sup> | 1-Nonen-3-ol                    | 1068   |
| 543              | Nopyl acetate                   | 1412   |
| 544              | n-Octadecane                    | 1800   |
| 545 <sup>^</sup> | gamma-Octalactone               | 1225   |
| 546              | delta-Octalactone               | 1252   |
| 547 <sup>*</sup> | n-Octanal                       | 985    |
| 548 <sup>*</sup> | n-Octanal diethyl acetal        | 1276   |
| 549 <sup>^</sup> | n-Octanal dimethyl acetal       | 1167   |
| 550 <sup>^</sup> | n-Octane                        | 800    |
| 551              | n-Octanol                       | 1061   |

---

---

|      |  |      |
|------|--|------|
| 552^ | n-Octan-2-ol                             | 988  |
| 553* | 2-Octanoylfuran                          | 1487 |
| 554* | 2-Octanoylthiophene                      | 1667 |
| 555  | 1-Octen-3-ol                             | 968  |
| 556* | 1-Octen-3-yl-n-propionate                | 1180 |
| 557^ | n-Octyl-acetate                          | 1193 |
| 558^ | n-Octyl n-butyrate                       | 1373 |
| 559^ | n-Octyl iso-butyrate                     | 1332 |
| 560^ | 2-n-Octylfuran                           | 1285 |
| 561  | n-Octyl 2-methyl butyrate                | 1422 |
| 562  | 2-n-Octylthiophene                       | 1463 |
| 563  | Paraldehyde                              | 763  |
| 564  | n-Pentadecane                            | 1500 |
| 565* | n-Pentane                                | 500  |
| 566* | 2,3-Pentanedione                         | 681  |
| 567^ | 2-Pentanoylthiophene                     | 1355 |
| 568^ | 1-Penten-3-ol                            | 673  |
| 569* | 2-n-Pentylfuran                          | 983  |
| 570  | 2-n-Pentyl-3-methyl-2-cyclopenten-1-one  | 1400 |
| 571^ | 2-n-Pentylthiophene                      | 1153 |
| 572  | beta-Phenoxyethyl iso-butyrate           | 1493 |
| 573^ | Phenol                                   | 1002 |
| 574^ | Phenylacetaldehyde dimethyl acetal       | 1200 |
| 575* | Phenyl iso-butyl methyl-carbinyl acetate | 1459 |
| 576^ | Phenylethyl acetate                      | 1233 |
| 577* | Phenylethyl alcohol                      | 1104 |
| 578  | Phenylethyl iso-butyrate                 | 1374 |
| 579^ | Phenylethyl formate                      | 1156 |
| 580^ | Phenylethyl n-heptanoate                 | 1718 |
| 581^ | Phenylethyl n-hexanoate                  | 1618 |
| 582* | Phenylethyl 2-methyl butyrate            | 1472 |
| 583  | Phenylethyl n-nonanoate                  | 1921 |
| 584* | Phenylethyl n-octanoate                  | 1819 |
| 585* | Phenylethyl pivalate                     | 1400 |
| 586^ | Phenylethyl n-propionate                 | 1328 |
| 587^ | Phenylethyl tiglate                      | 1562 |
| 588  | Phenylethyl n-valerate                   | 1517 |
| 589^ | Phenylethyl iso-valerate                 | 1474 |
| 590^ | Phenylpropyl acetate                     | 1347 |
| 591  | 3-Phenylpropyl alcohol                   | 1218 |

---

---

|      |  |      |
|------|--|------|
| 592^ | Phenylpropyl n-butyrate                        | 1535 |
| 593^ | Phenylpropyl iso-butyrate                      | 1490 |
| 594^ | Phenylpropyl n-propionate                      | 1445 |
| 595* | alpha-Pinene                                   | 942  |
| 596  | n-Propanol                                     | 535  |
| 597* | 2-Propanol                                     | 500  |
| 598* | Propionaldehyde-1-phenyl-1,2-ethenediol acetal | 1330 |
| 599  | 2-Propionylfuran                               | 988  |
| 600* | 2-Propionyl 5-methylfuran                      | 1106 |
| 601^ | 2-Propionylthiophene                           | 1164 |
| 602  | n-Propyl acetate                               | 694  |
| 603^ | iso-Propyl acetate                             | 645  |
| 604* | iso-Propyl benzoate                            | 1189 |
| 605^ | n-Propyl n-butyrate                            | 881  |
| 606^ | n-Propyl iso-butyrate                          | 842  |
| 607  | iso-Propyl n-butyrate                          | 825  |
| 608^ | iso-Propyl cinnamate                           | 1485 |
| 609  | 2-n-Propyl-4,5-dimethyloxazole                 | 996  |
| 610* | 2-iso-Propyl-4,5-dimethyloxazole               | 960  |
| 611  | 2-n-Propyl-4,5-dimethylthiazole                | 1151 |
| 612* | 2-iso-Propyl-4,5-dimethyloxazole               | 1109 |
| 613  | 2-n-Propyl-4-ethyl-5-methyloxazole             | 1064 |
| 614* | 2-iso-Propyl-4-ethyl-5-methyloxazole           | 1021 |
| 615  | n-Propyl formate                               | 606  |
| 616^ | iso-Propyl formate                             | 573  |
| 617^ | 2-n-Propylfuran                                | 782  |
| 618^ | n-Propyl n-heptanoate                          | 1177 |
| 619  | iso-Propyl n-heptanoate                        | 1120 |
| 620^ | n-Propyl n-hexanoate                           | 1079 |
| 621^ | iso-Propyl n-hexanoate                         | 1021 |
| 622  | n-Propyl levulinate                            | 1125 |
| 623^ | iso-Propyl levulinate                          | 1068 |
| 624  | n-Propyl n-methyl anthranilate                 | 1560 |
| 625^ | iso-Propyl n-methyl anthranilate               | 1491 |
| 626  | n-Propyl 2-methyl butyrate                     | 933  |
| 627^ | 2-n-Propyl-4-methylthiazole                    | 1040 |
| 628^ | n-Propyl n-octanoate                           | 1277 |
| 629  | iso-propyl n-octanoate                         | 1219 |
| 630* | iso-Propyl pivalate                            | 810  |
| 631* | n-Propyl n-propionate                          | 785  |

---

---

|      |  |      |
|------|--|------|
| 632^ | iso-Propyl n-propionate                              | 738  |
| 633* | n-Propyl salicylate                                  | 1357 |
| 634^ | iso-Propyl n-tetradecanoate                          | 1811 |
| 635  | 2-n-Propylthiazole                                   | 970  |
| 636* | 2-n-Propylthiophene                                  | 951  |
| 637* | iso-Propyl tiglate                                   | 959  |
| 638^ | n-Propyl 10-undecenoate                              | 1565 |
| 639* | n-Propyl n-valerate                                  | 981  |
| 640* | n-Propyl iso-valerate                                | 924  |
| 641^ | iso-Propyl iso-valerate                              | 883  |
| 642  | Pyrazine   | 739  |
| 643* | cis-Rose oxide                                       | 1087 |
| 644^ | trans-Rose oxide                                     | 1100 |
| 645  | Rosephenone trichloro methyl phenyl-carbinyl acetate | 1538 |
| 646  | Sabinete   | 976  |
| 647  | Salicylic aldehyde                                   | 1029 |
| 648^ | Styrallyl acetate                                    | 1173 |
| 649  | alpha-Terpineol                                      | 1185 |
| 650* | Terpinyl acetate                                     | 1333 |
| 651  | Terpinyl n-butyrate                                  | 1514 |
| 652^ | Terpinyl iso-butyrate                                | 1467 |
| 653* | Terpinyl n-propionate                                | 1426 |
| 654^ | Terpinyl iso-valerate                                | 1565 |
| 655* | n-Tetradecane  | 1400 |
| 656* | Tetrahydro geraniol                                  | 1185 |
| 657  | Tetrahydro geranyl acetate                           | 1306 |
| 658^ | Tetrahydro linalool                                  | 1087 |
| 659* | Tetrahydro myrcenol                                  | 1090 |
| 660* | 2,3,5,6-Tetramethyl pyrazine                         | 1069 |
| 661  | Thiophene  | 650  |
| 662* | Thymol   | 1287 |
| 663  | Tonalid  | 1849 |
| 664  | n-Tridecane  | 1300 |
| 665  | Triethyl citrate                                     | 1627 |
| 666  | 3,5,5-Trimethyl n-hexanal                            | 963  |
| 667* | 3,5,5-trimethyl n-hexanol                            | 1041 |
| 668  | 2,4,5-Trimethyloxazole                               | 829  |
| 669  | 2,3,5-Trimethyl pyrazine                             | 985  |
| 670^ | Trimethylthiazole                                    | 981  |
| 671^ | delta-Undecalactone                                  | 1579 |

---

---

|      |  |       |
|------|--|-------|
| 672^ | n–Undecanal                              | 1290  |
| 673^ | n–Undecane                               | 1100  |
| 674  | n–Undecanol                              | 1364  |
| 675^ | 6–Undecanol                              | 1281  |
| 676  | iso–Valeraldehyde–propyleneglycol acetal | 941.5 |
| 677* | Vanillin                                 | 1392  |
| 678  | Methyl ether                             | 350   |
| 679* | Acetaldehyde                             | 363   |
| 680* | Propionaldehyde                          | 481   |
| 681  | Furan                                    | 500   |
| 682  | tert–Butanol                             | 512   |
| 683  | Propyl methyl ether                      | 512   |
| 684* | Methyl acetate                           | 513   |
| 685  | Methyl vinyl ketone                      | 550   |
| 686  | Vinyl acetate                            | 562   |
| 687* | tert–Butyl methyl ether                  | 563   |
| 688  | 2,3 Butanedione                          | 575   |
| 689^ | 2–Butanone                               | 579   |
| 690* | Allyl ethyl ether                        | 586   |
| 691  | Isopropyl ether                          | 590   |
| 692  | sec–Butanol                              | 591   |
| 693  | Methyl propionate                        | 611   |
| 694^ | Butyl methyl ether                       | 615   |
| 695* | Tetrahydrofuran                          | 636   |
| 696* | 3–Buten–1–ol                             | 638   |
| 697  | Isovaleraldehyde                         | 649   |
| 698^ | 2–Butene–1–ol                            | 650   |
| 699* | Vinyl propionate                         | 650   |
| 700  | Isobutyl mercaptan                       | 660   |
| 701* | 2–Methyltetrahydrofuran                  | 674   |
| 702  | tert–Butyl acetate                       | 676   |
| 703* | Propyl ether                             | 676   |
| 704  | Cyclohexane                              | 677   |
| 705* | Methyl pyruvate                          | 680   |
| 706  | n–Butyl ethyl ether                      | 684   |
| 707* | Diethyl sulfide                          | 690   |
| 708  | Tetrahydropyran                          | 690   |
| 709^ | Valeraldehyde                            | 694   |
| 710  | Pinacolone                               | 695   |
| 711  | Pyridine                                 | 695   |

---

---

|                  |                            |     |
|------------------|----------------------------|-----|
| 712              | Methyl methacrylate        | 699 |
| 713 <sup>^</sup> | 3,4-Dihydropyran           | 705 |
| 714 <sup>*</sup> | 1,1-Diethoxyethane         | 710 |
| 715 <sup>*</sup> | N-Methylpyrrole            | 715 |
| 716 <sup>*</sup> | Thiazole                   | 715 |
| 717              | Methyl disulfide           | 730 |
| 718 <sup>^</sup> | Vinyl butyrate             | 745 |
| 719              | sec-Butyl acetate          | 746 |
| 720 <sup>^</sup> | Ethyl isobutyrate          | 746 |
| 721 <sup>*</sup> | Methyl isobutyl carbinol   | 748 |
| 722 <sup>*</sup> | Piperidine                 | 750 |
| 723 <sup>*</sup> | Propylene glycol           | 750 |
| 724              | 3-Methyl-2-pentanol        | 755 |
| 725 <sup>^</sup> | 3-Methyl-3-pentanol        | 757 |
| 726              | 4-Methyl-2-pentanol        | 758 |
| 727 <sup>*</sup> | 2-Methylthiophene          | 775 |
| 728 <sup>^</sup> | 2-Methyl-3-pentanol        | 777 |
| 729              | 2-Methylthiazole           | 780 |
| 730 <sup>^</sup> | Isobutyl acrylate          | 781 |
| 731 <sup>*</sup> | Ethyl pyruvate             | 785 |
| 732 <sup>^</sup> | 1,3-Butylene glycol acetal | 786 |
| 733 <sup>^</sup> | 2-Methyl-2-pentanol        | 797 |
| 734 <sup>*</sup> | 3-Methyl-2-pentanol        | 797 |
| 735              | 1-Octene                   | 797 |
| 736              | Tetrahydrothiophene        | 801 |
| 737 <sup>^</sup> | Cyclopentanol              | 802 |
| 738 <sup>*</sup> | Methyl mercaptoacetate     | 802 |
| 739              | Cyclopentanone             | 805 |
| 740 <sup>*</sup> | Elemol                     | 805 |
| 741 <sup>*</sup> | 3-Buten-2-ol               | 811 |
| 742 <sup>*</sup> | 2-Octene                   | 811 |
| 743 <sup>*</sup> | n-Butyl methyl sulfide     | 813 |
| 744 <sup>*</sup> | Allyl isobutyrate          | 820 |
| 745 <sup>^</sup> | Methyl hexyl ether         | 824 |
| 746              | 2-Methyl-1-pentanol        | 837 |
| 747 <sup>*</sup> | 4-Methyl-1-pentanol        | 838 |
| 748 <sup>*</sup> | Methyl sulfoxide           | 840 |
| 749 <sup>*</sup> | Methyl tiglate             | 850 |
| 750 <sup>^</sup> | 3-Methyl-1-pentanol        | 852 |
| 751              | Diallyl sulfide            | 854 |

---

---

|      |                            |      |
|------|----------------------------|------|
| 752  | Allyl isothiocyanate       | 860  |
| 753* | p-Xylene                   | 860  |
| 754  | n-Propyl pivalate          | 863  |
| 755  | m-Xylene                   | 863  |
| 756* | 2,5-Dimethylthiophene      | 866  |
| 757  | 3-Ethyl-3-pentanol         | 866  |
| 758* | Cyclohexenyl cyclohexanone | 1553 |
| 759^ | Tetrahydrofurfuryl alcohol | 884  |
| 760* | o-Xylene                   | 884  |
| 761* | gamma-Butyrolactone        | 885  |
| 762^ | Butyl acrylate             | 892  |
| 763  | n-Butyl mercaptan          | 892  |
| 764  | n-Butyl ethyl sulfide      | 893  |
| 765^ | 2,6-Dimethylpyrazine       | 895  |
| 766^ | Prenyl acetate             | 900  |
| 767* | Ethyl disulfide            | 910  |
| 768^ | n-Hexyl mercaptan          | 910  |
| 769  | Ethyl angelate             | 920  |
| 770  | gamma-Valerolactone        | 921  |
| 771^ | alpha-Thujene              | 938  |
| 772^ | 1,3-Butanediol             | 941  |
| 773* | 2-Ethoxythiazole           | 943  |
| 774  | Dimethyl trisulfide        | 952  |
| 775  | Camphene                   | 954  |
| 776^ | 2-Methyl-5-ethylthiophene  | 957  |
| 777  | 2-Ethylbutyl acetate       | 959  |
| 778  | 2-Ethylhexanal             | 959  |
| 779  | n-Butyl pivalate           | 963  |
| 780* | 1,1-Dimethoxy-n-hexane     | 964  |
| 781* | Allyl n-valerate           | 965  |
| 782* | Amyl vinyl carbinol        | 968  |
| 783  | Diisobutyl sulfide         | 969  |
| 784^ | Methyl benzyl ether        | 981  |
| 785  | trans-Pinane               | 981  |
| 786^ | beta-Pinene                | 981  |
| 787^ | cis-Dihydroocimene         | 982  |
| 788  | 2,6-Dimethylcyclohexanone  | 985  |
| 789^ | cis-Pinane                 | 987  |
| 790  | tert-Butyl benzene         | 988  |
| 791^ | 2-Octanone                 | 991  |

---

---

|      |                                     |      |
|------|-------------------------------------|------|
| 792* | Diethyl malate                      | 994  |
| 793* | trans-Dihydroocimene                | 995  |
| 794  | n-Butyl lactate                     | 997  |
| 795  | Methyl sorbate                      | 998  |
| 796  | Diisobutyl acetal                   | 999  |
| 797* | Isopentyl ether                     | 999  |
| 798* | Angelica lactone                    | 1000 |
| 799* | o-Methyl anisole                    | 1000 |
| 800* | p-Methyl anisole                    | 1000 |
| 801* | 2-Isopropyl-4-methylthiazole        | 1000 |
| 802* | Dimethyl succinate                  | 1002 |
| 803  | alpha-Phellandrene                  | 1002 |
| 804^ | Ethyl 2-methylpentanoate            | 1005 |
| 805* | Pyrrrol-2-carboxaldehyde            | 1005 |
| 806  | 1,4-Cineole                         | 1010 |
| 807* | n-Heptyl mercaptan                  | 1011 |
| 808  | 4-Methyl-5-vinylthiazole            | 1011 |
| 809  | Isopropyl disulfide                 | 1014 |
| 810* | Methyl octyl ether                  | 1016 |
| 811^ | p-Cimene                            | 1020 |
| 812^ | 2-Isopropyl-4-ethyl-5-methyloxazole | 1021 |
| 813^ | Carvomenthene                       | 1022 |
| 814* | Phenylacetaldehyde                  | 1024 |
| 815  | cis-Ocimene                         | 1025 |
| 816^ | beta-Phellandrene                   | 1025 |
| 817* | 1,8-Cineole                         | 1027 |
| 818  | Isobutyl angelate                   | 1033 |
| 819  | 2-Ethyl-1-hexyl vinyl ether         | 1038 |
| 820* | trans-Ocimene                       | 1038 |
| 821^ | Isobutyl n-valerate                 | 1040 |
| 822  | trans-2-Octenal                     | 1045 |
| 823  | 1,4-Butanediol                      | 1046 |
| 824^ | Methyl hexyl acetaldehyde           | 1048 |
| 825* | 2-Acetylpyrrole                     | 1050 |
| 826^ | 2,5-Dimethyl-4-ethylthiazole        | 1050 |
| 827* | p-Cresol                            | 1051 |
| 828^ | m-Tolualdehyde                      | 1053 |
| 829* | o-Tolualdehyde                      | 1054 |
| 830^ | n-Methyl-2-acetylpyrrole            | 1055 |
| 831  | Tetrahydrofurfuryl acetate          | 1055 |

---

---

|      |                             |      |
|------|-----------------------------|------|
| 832* | gamma-Terpinene             | 1057 |
| 833^ | n-Pentyl ether              | 1065 |
| 834  | p-Tolualdehyde              | 1067 |
| 835  | m-Diethylbenzene            | 1070 |
| 836* | Guaiacol                    | 1071 |
| 837^ | Di-n-butyl sulfide          | 1073 |
| 838^ | 2-Methyl-3-n-propylpyrazine | 1074 |
| 839^ | 5-Nonanone                  | 1074 |
| 840* | Ethyl sorbate               | 1075 |
| 841* | p-Diethylbenzene            | 1080 |
| 842^ | alpha-p-dimethylstyrene     | 1080 |
| 843  | Fenchone                    | 1080 |
| 844* | trans-3-Heptenyl acetate    | 1080 |
| 845^ | cis-3-Heptenyl acetate      | 1084 |
| 846^ | 2,6-Dimethyl phenol         | 1087 |
| 847  | o-Diethylbenzene            | 1088 |
| 848* | 2-Nonanone                  | 1093 |
| 849^ | Amyl vinyl carbonyl acetate | 1094 |
| 850* | cis-3-Hexenyl ethyl acetal  | 1094 |
| 851* | Propyl disulfide            | 1096 |
| 852  | 2-Cyclohexyl ethanol        | 1098 |
| 853^ | Myrcenol                    | 1103 |
| 854* | 2-Buten-1,4-diol            | 1104 |
| 855* | Butyl acetoacetate          | 1104 |
| 856* | Maltol                      | 1105 |
| 857^ | alpha-Fenchyl alcohol       | 1110 |
| 858* | Norbornyl acetate           | 1112 |
| 859* | p-tert-Amylcyclohexanone    | 1113 |
| 860  | 2-Ethyl phenol              | 1113 |
| 861* | Methyl 2-hydroxyisobutyrate | 1116 |
| 862* | Methyl nicotinate           | 1116 |
| 863* | Octyl formate               | 1117 |
| 864^ | 2-methyl-1-octanol          | 1119 |
| 865^ | Dihydrolinalool             | 1122 |
| 866^ | p-Methyl benzyl alcohol     | 1122 |
| 867  | Diethyl trisulfide          | 1125 |
| 868  | Propyl furoate              | 1125 |
| 869  | Cyclooctanol                | 1133 |
| 870* | Sabinol                     | 1135 |
| 871  | Dihydrocinnamic aldehyde    | 1139 |

---

---

|      |                               |      |
|------|-------------------------------|------|
| 872^ | 1,3-Dimethoxy benzene         | 1143 |
| 873* | Menthone                      | 1143 |
| 874^ | p-Hydroxyacetophenone         | 1144 |
| 875  | Isopulegol                    | 1145 |
| 876^ | 3-Ethyl phenol                | 1146 |
| 877^ | trans-2-Nonenal               | 1146 |
| 878* | Tetrahydrofurfuryl propionate | 1153 |
| 879* | Lavandulol                    | 1154 |
| 880* | Isooctyl acetate              | 1154 |
| 881^ | Isobutyl 2-methylpentanoate   | 1155 |
| 882  | Isoborneol                    | 1157 |
| 883* | trans-2-nonenol               | 1157 |
| 884  | Diethyl fumarate              | 1160 |
| 885* | Borneol                       | 1164 |
| 886^ | Methyl 2-methylbenzoate       | 1165 |
| 887^ | cis-3-Heptenyl propionate     | 1171 |
| 888  | Amyl butyl carbinol           | 1175 |
| 889^ | Terpinene-4-ol                | 1175 |
| 890^ | Isobutyl 2-hexenoate          | 1180 |
| 891  | 2-Pentanoylfuran              | 1180 |
| 892^ | Methyl chavicol               | 1182 |
| 893^ | Dihydrocarvone                | 1183 |
| 894* | Methyl 3-methylbenzoate       | 1190 |
| 895  | Verbenone                     | 1195 |
| 896  | Methyl 4-methylbenzoate       | 1199 |
| 897* | p-isopropylphenol             | 1200 |
| 898^ | Dihydromyrcenyl acetate       | 1202 |
| 899* | Isobutyl disulfide            | 1205 |
| 900^ | Fenchyl acetate               | 1220 |
| 901^ | Benzyl ethyl carbinol         | 1222 |
| 902  | 2-Ethyl-1-hexyl acrylate      | 1224 |
| 903* | 2,4-Dimethyl benzyl alcohol   | 1226 |
| 904* | Neral                         | 1227 |
| 905^ | Isobornyl formate             | 1228 |
| 906^ | Pulegone                      | 1230 |
| 907^ | 2,4-Dimethylacetophenone      | 1233 |
| 908  | Citronellyl vinyl ether       | 1235 |
| 909* | n-Butyl benzyl ether          | 1238 |
| 910^ | Chavicol                      | 1238 |
| 911  | Bornyl formate                | 1239 |

---

---

|                  |                               |      |
|------------------|-------------------------------|------|
| 912              | Myrcenyl acetate              | 1247 |
| 913              | Piperitone                    | 1247 |
| 914 <sup>^</sup> | Geranial                      | 1252 |
| 915 <sup>^</sup> | Methyl octyl acetaldehyde     | 1254 |
| 916              | n-Propyl benzoate             | 1254 |
| 917 <sup>^</sup> | cis-3-Heptenyl butyrate       | 1255 |
| 918              | 2-decen-1-ol                  | 1257 |
| 919 <sup>*</sup> | trans-2-Decenol               | 1257 |
| 920 <sup>*</sup> | Diisoamyl ketone              | 1258 |
| 921 <sup>*</sup> | Methyl dihydrocinnamate       | 1258 |
| 922 <sup>^</sup> | Isopulegyl acetate            | 1258 |
| 923              | Carvone oxide                 | 1261 |
| 924 <sup>*</sup> | Sabinyl acetate               | 1262 |
| 925 <sup>*</sup> | 4-Ethylguaiacol               | 1265 |
| 926              | Neryl formate                 | 1267 |
| 927 <sup>^</sup> | Safrole                       | 1278 |
| 928              | 10-Undecen-1-al               | 1280 |
| 929              | Octyl propionate              | 1283 |
| 930 <sup>*</sup> | Dihydrosafrole                | 1286 |
| 931 <sup>*</sup> | n-Nonanoic acid               | 1286 |
| 932 <sup>*</sup> | n-Nonyl acetate               | 1292 |
| 933 <sup>*</sup> | Isopropyl salicylate          | 1292 |
| 934              | Butyl disulfide               | 1295 |
| 935              | 2-Nonyl-1-al dimethyl acetal  | 1300 |
| 936              | n-Propyl phenylacetate        | 1300 |
| 937 <sup>^</sup> | Anisyl formate                | 1307 |
| 938 <sup>^</sup> | alpha-Methylcinnamic aldehyde | 1309 |
| 939              | Cyclohexyl n-valerate         | 1310 |
| 940 <sup>*</sup> | n-Hexyltiglate                | 1310 |
| 941 <sup>^</sup> | Methyl 3-hydroxybutyrate      | 1318 |
| 942 <sup>*</sup> | Methyl undecyl ether          | 1318 |
| 943              | Dihydrocarvyl acetate         | 1319 |
| 944              | Myrcenyl propionate           | 1327 |
| 945              | Cinnamyl formate              | 1330 |
| 946 <sup>^</sup> | p-Isopropylacetophenone       | 1332 |
| 947 <sup>^</sup> | Terpinyl formate              | 1333 |
| 948 <sup>^</sup> | Methyllavender ketone         | 1341 |
| 949 <sup>*</sup> | alpha-Methylcinnamyl alcohol  | 1343 |
| 950 <sup>*</sup> | Hydroxycitronellol            | 1347 |
| 951 <sup>*</sup> | gamma-Ionone                  | 1347 |

---

---

|      |  |      |
|------|--|------|
| 952^ | 2,4-Dimethyl benzylacetate             | 1348 |
| 953^ | Undecenol                              | 1350 |
| 954* | Methyl nonyl acetaldehyde              | 1352 |
| 955* | 2-Methyl-5-isopropylacetophenone       | 1358 |
| 956* | Dihydrocoumarin                        | 1359 |
| 957  | Isosafrole                             | 1360 |
| 958^ | beta-Methylcinnamyl alcohol            | 1365 |
| 959  | 4-Methyl-5-thiazoleethanol acetate     | 1368 |
| 960^ | Diphenyl                               | 1369 |
| 961  | n-Pentadecanol                         | 1373 |
| 962* | Acetaldehyde citronellyl methyl acetal | 1374 |
| 963^ | cis-3-Decenyl acetate                  | 1376 |
| 964  | Acetylcymene                           | 1381 |
| 965^ | Isoamyl phenylethyl ether              | 1384 |
| 966^ | p-Cresyl isovalerate                   | 1389 |
| 967* | Anisyl acetate                         | 1390 |
| 968  | alpha-Copaene                          | 1398 |
| 969* | Dihydrojasmane                         | 1400 |
| 970^ | Cuminyl acetate                        | 1401 |
| 971  | beta-Bourbonene                        | 1406 |
| 972* | m-Aminoacetophenone                    | 1409 |
| 973^ | Dimethyl suberate                      | 1414 |
| 974^ | Isopropyl n-decanoate                  | 1417 |
| 975* | Dihydrocumyl acetate                   | 1418 |
| 976^ | Linalyl n-butyrate                     | 1420 |
| 977^ | 2-Methyl-1-undecanol                   | 1422 |
| 978* | Phenylethyl butyrate                   | 1422 |
| 979^ | Dimethyl phenyl ethyl carbonyl acetate | 1428 |
| 980* | alpha-Cedrene                          | 1436 |
| 981^ | Carvyl propionate                      | 1440 |
| 982  | beta-Copaene                           | 1445 |
| 983  | beta-Cedrene                           | 1446 |
| 984* | Phenoxyethyl propionate                | 1447 |
| 985  | 2-Decenal                              | 1449 |
| 986  | Linalyl 2-methylbutyrate               | 1450 |
| 987* | Thujopsene                             | 1451 |
| 988^ | n-Amyl benzoate                        | 1454 |
| 989* | Di-n-heptyl ether                      | 1458 |
| 990* | alpha-Humulene                         | 1465 |
| 991^ | Citronellyl isobutyrate                | 1469 |

---

---

|                   |                              |        |
|-------------------|------------------------------|--------|
| 992               | Bornyl butyrate              | 1473   |
| 993 <sup>^</sup>  | Dihydropseudo ionone         | 1473   |
| 994               | p-Methoxybenzylacetone       | 1473   |
| 995               | Alloaromadendrene            | 1478   |
| 996               | p-Cresyl isotiglate          | 1482   |
| 997*              | alpha-Methylcinnamyl acetate | 1484   |
| 998*              | Heliotropyl acetate          | 1485   |
| 999*              | Phenoxyethyl isobutyrate     | 1486   |
| 1000*             | n-Undecyl acetate            | 1487   |
| 1001 <sup>^</sup> | Valencene                    | 1487   |
| 1002 <sup>^</sup> | 3-Methylcoumarin             | 1490   |
| 1003              | 2-Cyclohexylcyclohexanone    | 1496   |
| 1004*             | alpha-Murolene               | 1500   |
| 1005              | o-Methoxy cinnamic aldehyde  | 1504   |
| 1006 <sup>^</sup> | Methyl isovalerate           | 1505   |
| 1007 <sup>^</sup> | Bornyl isovalerate           | 1512   |
| 1008              | gamma-Cadinene               | 1518   |
| 1009 <sup>^</sup> | Calamenene                   | 1518   |
| 1010 <sup>^</sup> | delta-Cadinene               | 1524   |
| 1011*             | alpha-Nerolidol              | 1524   |
| 1012*             | Citronellyl isocrotonate     | 1526   |
| 1013*             | 3,4-Dimethoxyacetophenone    | 1532   |
| 1014              | Di-n-butyl succinate         | 1534   |
| 1015 <sup>^</sup> | gamma-Undecalactone          | 1542   |
| 1016 <sup>^</sup> | Tangerinal                   | 1554.5 |
| 1017              | 6-Methylcoumarin             | 1545   |
| 1018 <sup>^</sup> | cis-3-Hexenyl benzoate       | 1553   |
| 1019              | 7-Methylcoumarin             | 1553   |
| 1020 <sup>^</sup> | beta-Nerolidol               | 1553   |
| 1021 <sup>^</sup> | Citronellyl crotonate        | 1558   |
| 1022*             | Di-n-amyl fumarate           | 1558   |
| 1023*             | n-Hexyl benzoate             | 1558   |
| 1024              | Triacetin                    | 1563   |
| 1025 <sup>^</sup> | n-Hexyl octanoate            | 1564   |
| 1026 <sup>^</sup> | Anisyl n-butyrate            | 1569   |
| 1027              | Geranyl 2-methylbutyrate     | 1574   |
| 1028 <sup>^</sup> | Neryl isovalerate            | 1574   |
| 1029              | alpha-Cedrene epoxide        | 1585   |
| 1030 <sup>^</sup> | Phenylpropyl isovalerate     | 1590   |
| 1031*             | n-Hexyl phenylacetate        | 1607   |

---

---

|                   |   |        |
|-------------------|---|--------|
| 1032              | cis-3-hexenyl phenylacetate                         | 1610   |
| 1033 <sup>^</sup> | Methyl n-tridecanoate                               | 1612   |
| 1034 <sup>^</sup> | Isobutyl 10-undecenoate                             | 1617   |
| 1035              | Dibenzyl ether                                      | 1631   |
| 1036 <sup>^</sup> | Geranyl tiglate                                     | 1650   |
| 1037 <sup>^</sup> | Isoamyl anthranilate                                | 1656   |
| 1038              | Di-n-butyl adipate                                  | 1658   |
| 1039              | Furfuryl disulfide                                  | 1660   |
| 1040 <sup>*</sup> | Methyl zingerone                                    | 1660   |
| 1041              | Patchouli alcohol                                   | 1667   |
| 1042 <sup>^</sup> | Methyl N-propylanthranilate                         | 1678   |
| 1043 <sup>^</sup> | Guaiacyl n-caproate                                 | 1681   |
| 1044 <sup>*</sup> | Isoamyl pyruvate                                    | 1712   |
| 1045              | Cyclohexadecanone                                   | 1731   |
| 1046              | Acetyl cedrene                                      | 1768   |
| 1047 <sup>^</sup> | n-heptyl salicylate                                 | 1790   |
| 1048              | Neryl n-Heptanoate                                  | 1808   |
| 1049 <sup>^</sup> | Pentalide   | 1823   |
| 1050 <sup>^</sup> | Geranyl n-heptanoate                                | 1831   |
| 1051              | Nerolidyl ethanol                                   | 1851   |
| 1052 <sup>*</sup> | Octyl salicylate                                    | 1895   |
| 1053 <sup>^</sup> | Methyl hexadecanoate                                | 1911   |
| 1054 <sup>*</sup> | Phenylethyl n-decanoate                             | 2022   |
| 1055 <sup>*</sup> | Methyl octadecanoate                                | 2101   |
| 1056              | Ethyl oleate  | 2155   |
| 1057              | n-Butyl n-hexanoate                                 | 2180   |
| 1058              | Ethyl n-octanoate                                   | 2180   |
| 1059 <sup>^</sup> | beta-Terpinenol                                     | 1137   |
| 1060 <sup>*</sup> | delta-Terpinenol                                    | 1160   |
| 1061 <sup>^</sup> | sec-butylcyclohexanone                              | 1195   |
| 1062 <sup>^</sup> | 4-(2,4,6-Trimethyl-3-cyclohexen-1-yl)-3-buten-2-one | 1434   |
| 1063              | Allyl cinnamate                                     | 1450   |
| 1064 <sup>*</sup> | Cedryl acetate                                      | 1596.5 |
| 1065              | 1,1-Dimethoxy-n-propane                             | 650    |
| 1066 <sup>*</sup> | n-Pentan-2-ol                                       | 685    |
| 1067              | Nootketone  | 1802   |
| 1068              | Dibutyl butyrolactone                               | 1531   |
| 1069 <sup>*</sup> | 2-Propionylpyrrol                                   | 1145   |
| 1070 <sup>^</sup> | Octyl-n-valerate                                    | 1474   |
| 1071              | n-Propyl n-decanoate                                | 1476   |

---

---

|                   |                              |      |
|-------------------|------------------------------|------|
| 1072              | Isoamyl n-nonanoate          | 1533 |
| 1073 <sup>^</sup> | n-Butyl n-nonanoate          | 1475 |
| 1074*             | 1,1-Dimethoxy-n-butane       | 770  |
| 1075              | 1,1-Dimethoxy-n-pentane      | 868  |
| 1076              | 2-n-Butylthiazole            | 1070 |
| 1077 <sup>^</sup> | n-gamma-heptalactone         | 1126 |
| 1078*             | n-Heptyl pivalate            | 1263 |
| 1079*             | Isopropyl n-dodecanoate      | 1814 |
| 1080              | n-Butyl n-tridecanoate       | 1880 |
| 1081 <sup>^</sup> | 2,2-Dimethyl-n-pentanol      | 874  |
| 1082 <sup>^</sup> | Isopropyl n-valerate         | 924  |
| 1083*             | Salicaldehyde                | 1027 |
| 1084              | n-Butyl propyl sulfide       | 972  |
| 1085 <sup>^</sup> | Furfuryl n-propionate        | 1059 |
| 1086              | n-Propyl tiglate             | 1020 |
| 1087*             | 2,4-Dimethyl-n-pentan-3-ol   | 828  |
| 1088*             | 2-Ethyl-2-n-hexanal          | 1007 |
| 1089              | 1,1-Dimethoxy-n-heptane      | 1063 |
| 1090*             | n-Decan-2-ol                 | 1190 |
| 1091 <sup>^</sup> | D-n-hexyl ether              | 1269 |
| 1092*             | n-Butyl angelate             | 1116 |
| 1093              | n-Amyl tiglate               | 1216 |
| 1094*             | Benzyl n-decanoate           | 1923 |
| 1095              | n-Propyl anisate             | 1527 |
| 1096*             | Anisyl n-propionate          | 1482 |
| 1097*             | n-Propyl anthranilate        | 1500 |
| 1098 <sup>^</sup> | Neryl n-valerate             | 1610 |
| 1099 <sup>^</sup> | Furfuryl n-heptanoate        | 1443 |
| 1100*             | Isopropyl n-nonanoate        | 1318 |
| 1101*             | n-Heptyl valerate            | 1372 |
| 1102*             | n-Amyl anthranilate          | 1700 |
| 1103              | Cinnamyl n-heptanoate        | 1905 |
| 1104              | trans-2-Hexenyl n-heptanoate | 1474 |
| 1105*             | Citronellyl n-valerate       | 1608 |
| 1106 <sup>^</sup> | n-hexyl nonanoate            | 1668 |
| 1107              | Linalyl n-heptanoate         | 1670 |
| 1108*             | Terpinyl n-valerate          | 1614 |
| 1109*             | n-Heptyl 2-methylbutyrate    | 1324 |
| 1110*             | n-Amyl furfurylacrylate      | 1544 |
| 1111              | Geranyl n-valerate           | 1632 |

---

---

|       |                                   |      |
|-------|-----------------------------------|------|
| 1112  | n–Amyl anisate                    | 1732 |
| 1113* | n–Butyl n–valerate                | 1078 |
| 1114* | Isopropyl n–undecanoate           | 1516 |
| 1115  | Methyl n–dodecyl ketone           | 1580 |
| 1116^ | Neryl n–hexanoate                 | 1709 |
| 1117* | n–Heptyl phenylacetate            | 1717 |
| 1118  | n–Undecanal diethyl acetal        | 1572 |
| 1119  | Methyl n–decyl ketone             | 1377 |
| 1120^ | n–Hexyl n–heptanoate              | 1470 |
| 1121* | n–Amyl n–heptanoate               | 1375 |
| 1122^ | n–Undecanal dimethyl acetal       | 1466 |
| 1123* | Isoamyl n–undecanoate             | 1733 |
| 1124^ | n–Amyl 10–undecenoate             | 1760 |
| 1125  | n–Butyl n–pentadecanoate          | 2080 |
| 1126* | n–Butyl n–tetradecanoate          | 1977 |
| 1127* | n–Propyl n–dodecanoate            | 1676 |
| 1128^ | n–Propyl n–nonanoate              | 1377 |
| 1129^ | Furfuryl n–valerate               | 1245 |
| 1130  | n–Hexyl n–valerate                | 1275 |
| 1131  | alpha–n–Butyl cinnamic aldehyde   | 1535 |
| 1132* | n–Dodecanal diethyl acetal        | 1671 |
| 1133* | n–Butyl N–methylantranilate       | 1660 |
| 1134  | 2–Nonanoylfuran                   | 1588 |
| 1135^ | 2–Decanoylfuran                   | 1689 |
| 1136  | alpha–n–Heptylcinnamic aldehyde   | 1827 |
| 1137^ | 2–Ethyl–2–n–hexene–1–ol           | 1051 |
| 1138^ | Benzyl n–propyl carbinol          | 1310 |
| 1139  | Benzyl isopropyl carbinol         | 1292 |
| 1140* | n–Propyl cinnamic aldehyde        | 1531 |
| 1141^ | cis–3–Hexenyl n–heptanoate        | 1465 |
| 1142^ | trans–2–Hexenyl n–valerate        | 1275 |
| 1143  | Cyclooctanyl acetate              | 1280 |
| 1144^ | cis–3–Decenol                     | 1245 |
| 1145* | p–cresyl ethylcarbonate           | 1304 |
| 1146^ | 2–Ethyl–1–hexyl propionate        | 1231 |
| 1147* | 2–Methyl–2–phenylhexan–4–one      | 1405 |
| 1148* | 4–Methyl–5–hydroxy phenyl acetate | 1263 |
| 1149^ | Jasmal                            | 1459 |
| 1150* | cis–3–Octenol                     | 1041 |
| 1151^ | trans–3–Octenol                   | 1036 |

---

---

|                   |  |      |
|-------------------|--|------|
| 1152 <sup>^</sup> | cis-3-Decenyl butyrate                       | 1563 |
| 1153              | trans-2-Decenyl butyrate                     | 1563 |
| 1154 <sup>^</sup> | n-Hexyl angelate                             | 788  |
| 1155              | n-Propyl n-undecanoate                       | 1576 |
| 1156 <sup>^</sup> | cis-3-Hexenyl n-nonanoate                    | 1664 |
| 1157              | trans-2-Hexenyl n-nonanoate                  | 1673 |
| 1158 <sup>^</sup> | cis-3-Hexenyl methyl acetal                  | 1035 |
| 1159              | trans-2-Heptenyl butyrate                    | 1275 |
| 1160 <sup>*</sup> | trans-2-Heptenyl propionate                  | 1182 |
| 1161              | Citronellyl ethyl acetal                     | 1423 |
| 1162 <sup>^</sup> | o-Toluyl thiol                               | 1067 |
| 1163 <sup>^</sup> | Isoamyl anisate                              | 1686 |
| 1164 <sup>*</sup> | n-Amyl pivalate                              | 1063 |
| 1165 <sup>^</sup> | Isolongiforanone                             | 1616 |
| 1166 <sup>*</sup> | Di-n-Propyl fumarate                         | 1360 |
| 1167 <sup>^</sup> | N-Methyl-2-pyrrolaldehyde                    | 986  |
| 1168 <sup>^</sup> | 1-Methyl 4-hepten-1-ol                       | 982  |
| 1169 <sup>*</sup> | alpha-Methyl citronellol                     | 1220 |
| 1170 <sup>^</sup> | Phenylpropyl n-valerate                      | 1635 |
| 1171 <sup>*</sup> | n-Propionyl methylantranilate                | 1673 |
| 1172 <sup>*</sup> | 1,1-Diethoxyhexane                           | 1080 |
| 1173              | o-Methyl phenyl ethyl alcohol                | 1216 |
| 1174 <sup>*</sup> | omega-Decenyl butyrate                       | 1558 |
| 1175 <sup>^</sup> | omega-Methyl undecylenate                    | 1400 |
| 1176 <sup>^</sup> | 4-Ethyl isohexanoate                         | 951  |
| 1177 <sup>*</sup> | Dihydrolinalyl acetaldehyde                  | 1315 |
| 1178 <sup>*</sup> | Methyl isogeranylacetate                     | 1362 |
| 1179              | Dihydrocumyl alcohol                         | 1286 |
| 1180 <sup>^</sup> | Methyl isohexyl carbonyl acetate             | 1080 |
| 1181              | Ethyl phenyl ethyl acetal                    | 1332 |
| 1182              | 2-Methyl-2-(4-methyl n-amyl) tetrahydrofuran | 1159 |
| 1183 <sup>*</sup> | sec-Butyl ethyl formal                       | 826  |
| 1184              | 1,3-Butylene glycol formal                   | 777  |
| 1185 <sup>*</sup> | 1,4-Butylene glycol formal                   | 829  |
| 1186 <sup>*</sup> | 2,3-Butylene glycol formal                   | 740  |
| 1187              | Pinacol                                      | 843  |
| 1188 <sup>^</sup> | sec-Butyl benzene                            | 1011 |
| 1189 <sup>^</sup> | trans-Sabinene hydrate                       | 1078 |
| 1190 <sup>^</sup> | gamma-Muurolene                              | 1475 |
| 1191 <sup>*</sup> | Anisyl isobutyrate                           | 1520 |

---

---

|       |                            |      |
|-------|----------------------------|------|
| 1192^ | Zingerone                  | 1625 |
| 1193* | Anisyl n-valerate          | 1665 |
| 1194^ | Benzyl benzoate            | 1741 |
| 1195^ | Phenyl salicylate          | 1742 |
| 1196  | Bornyl benzoate            | 1749 |
| 1197  | p-Cresyl benzoate          | 1764 |
| 1198^ | cis-3-hexenyl anthranilate | 1807 |
| 1199^ | p-Cresyl phenylacetate     | 1827 |
| 1200  | Phenylethyl benzoate       | 1841 |
| 1201* | p-Cresyl salicylate        | 1850 |
| 1202  | 2-Mercaptobenzothiazole    | 1944 |
| 1203* | Geranyl benzoate           | 1949 |
| 1204  | Cinnamyl cinnamate         | 2055 |
| 1205* | Phenylethyl anthranilate   | 2091 |
| 1206* | Phenylethyl cinnamate      | 2147 |
| 1207* | Anisyl n-heptanoate        | 1862 |
| 1208^ | Anisyl n-hexanoate         | 1763 |

---

\* validación, ^ predicción

**Tabla 2A.** Nombres químicos e índices de retención experimentales de 1184 sabores y aromas medidos en la columna polar Carbowax 20M

| No. | Nombre                                   | I      |
|-----|--|--------|
| 1   | Acetaldehyde di-(cis-3-hexenil) acetal   | 1700   |
| 2*  | Acetaldehyde ethyl-cis-3-hexenyl-acetal  | 1297   |
| 3^  | Acetaldehyde linalyl ethyl acetal        | 1561   |
| 4^  | Acetaldehyde phenylethyl n-propyl acetal | 1836   |
| 5^  | Acetaldehyde styleneglycol acetal        | 1786.5 |
| 6   | Acetone                                  | 810    |
| 7*  | Acetone-1-phenyl 1,2-ethandiol ketal     | 1766   |
| 8*  | Acetonyl acetone                         | 1500   |
| 9*  | Acetophenone                             | 1627   |
| 10  | 4-Acetyl-6-tert-butyl-1,1-dimethyl indan | 2145   |
| 11* | Acetyl eugenol eugenyl acetate           | 2277   |
| 12  | 2-Acetylfuran                            | 1491   |
| 13  | Acetyl methyl carbinol acetoin           | 1276   |
| 14^ | 2-Acetylpyridine                         | 1600   |
| 15  | 2-Acetylthiazole                         | 1639   |
| 16^ | 2-Acetylthiophene                        | 1760   |
| 17* | Allyl acetate                            | 1010   |
| 18  | Allyl anthranilate                       | 2289   |
| 19^ | Allyl benzoate                           | 1800   |
| 20^ | Allyl n-butyrate                         | 1161   |
| 21* | Allyl cyclohexane propionate             | 1800   |
| 22* | Allyl n-decanoate                        | 1755   |
| 23* | Allyl 2-ethyl butyrate                   | 1254   |
| 24* | Allyl formate                            | 957    |
| 25* | Allyl 2-furoate                          | 1748   |
| 26^ | Allyl n-heptanoate                       | 1454   |
| 27^ | Allyl n-hexanoate                        | 1356   |
| 28  | Allyl ionone                             | 2088   |
| 29^ | Allyl levulinate                         | 1743   |
| 30* | Allyl n-nonanoate                        | 1655   |
| 31^ | Allyl n-octanoate                        | 1554   |
| 32* | Allyl phenoxy acetate                    | 2176   |
| 33^ | Allyl phenyl acetate                     | 2175   |
| 34* | Allyl n-propionate                       | 1090   |
| 35^ | Allyl salicylate                         | 1946   |
| 36^ | Allyl tiglate                            | 1370   |
| 37* | Allyl iso-valerate                       | 1190   |

---

|     |                                      |      |
|-----|--------------------------------------|------|
| 38* | o-Amino acetophenone                 | 2181 |
| 39  | p-Amino acetophenone                 | 2181 |
| 40* | n-Amyl acetate                       | 1161 |
| 41  | iso-Amyl acetate                     | 1110 |
| 42  | 2-n-Amyl-3-acetonyl-1-cyclopentanone | 2259 |
| 43  | n-Amyl alcohol                       | 1213 |
| 44* | iso-Amyl alcohol                     | 1184 |
| 45  | sec-Amyl alcohol                     | 1091 |
| 46  | tert-Amyl alcohol                    | 987  |
| 47^ | iso-Amyl benzoate                    | 1894 |
| 48  | iso-Amyl benzyl ether                | 1668 |
| 49  | n-Amyl n-butyrate                    | 1305 |
| 50* | n-Amyl iso-butyrate                  | 1237 |
| 51  | iso-Amyl n-butyrate                  | 1259 |
| 52  | iso-Amyl iso-butyrate                | 1187 |
| 53* | iso-Amyl cinnamate                   | 2355 |
| 54* | alpha-Amyl cinnamic aldehyde         | 2211 |
| 55* | alpha-n-Amyl cinnamyl acetate        | 2318 |
| 56^ | n-Amyl cyclopentenone                | 1719 |
| 57  | p-tert-amyl cyclohexanone            | 1800 |
| 58^ | iso-Amyl n-decanoate                 | 1848 |
| 59^ | iso-Amyl n-dodecanoate               | 2048 |
| 60  | n-Amyl formate                       | 1107 |
| 61  | iso-Amyl formate                     | 1058 |
| 62* | iso-Amyl 2-furoate                   | 1840 |
| 63  | n-amyl furylpropionate               | 1947 |
| 64^ | iso-amyl furylpropionate             | 1894 |
| 65  | iso-amyl n-pentanoate                | 1548 |
| 66* | n-amyl n-hexanoate                   | 1500 |
| 67* | iso-amyl n-hexanoate                 | 1451 |
| 68  | n-amyl levulinate                    | 1860 |
| 69* | iso-amyl levulinate                  | 1807 |
| 70^ | n-amyl 2-methyl butyrate             | 1324 |
| 71* | iso-amyl 2-methyl butyrate           | 1273 |
| 72  | n-amyl n-octanoate                   | 1700 |
| 73^ | iso-amyl n-octanoate                 | 1648 |
| 74* | n-amyl phenylacetate                 | 2047 |
| 75^ | iso-amyl phenylacetate               | 1991 |
| 76* | n-amyl n-propionate                  | 1180 |
| 77^ | iso-amyl n-propionate                | 1180 |

---

---

|      |                                      |        |
|------|--------------------------------------|--------|
| 78^  | 6-amyl alpha-pyrone                  | 2166   |
| 79   | iso-amyl pyruvate                    | 1464   |
| 80   | n-amyl salicylate                    | 2077   |
| 81*  | iso-amyl salicylate                  | 2021   |
| 82^  | iso-amyl tiglate                     | 1469   |
| 83   | iso-amyl 10-undecenoate              | 2000   |
| 84^  | n-amyl n-valerate                    | 1401   |
| 85*  | n-amyl iso-valerate                  | 1337   |
| 86^  | iso-amyl n-valerate                  | 1354   |
| 87*  | Isoamyl isovalerate                  | 1287   |
| 88*  | anethole                             | 1809   |
| 89   | anis alcohol                         | 2210   |
| 90*  | anis aldehyde                        | 1982   |
| 91^  | anis aldehyde-propyleneglycol acetal | 2232.5 |
| 92*  | anisole                              | 1327   |
| 93^  | benzal acetone                       | 2065   |
| 94^  | benzaldehyde                         | 1502   |
| 95   | benzotrile                           | 1583   |
| 96^  | benzophenone                         | 2410   |
| 97   | Benzyl acetate                       | 1697   |
| 98*  | Benzyl acetone                       | 1849   |
| 99   | Benzyl alcohol                       | 1822   |
| 100^ | Benzyl n-butyrate                    | 1856   |
| 101* | Benzyl iso-butyrate                  | 1771   |
| 102^ | Benzyl ethyl ether                   | 1439   |
| 103* | Benzyl formate                       | 1675   |
| 104^ | Benzyl n-heptanoate                  | 2158   |
| 105^ | Benzyl n-hexanoate                   | 2057   |
| 106* | Benzyl n-nonanoate                   | 2362   |
| 107  | Benzyl n-octanoate                   | 2260   |
| 108^ | Benzyl n-propionate                  | 1769   |
| 109* | Benzyl tiglate                       | 2075   |
| 110  | Benzyl n-valerate                    | 1956   |
| 111  | Benzyl iso-valerate                  | 1880   |
| 112* | iso-bornyl acetate                   | 1584   |
| 113^ | iso-bornyl n-propionate              | 1676   |
| 114  | bromostryrol                         | 1778   |
| 115  | 2,3 butadienone diacetyl             | 963    |
| 116  | n butanol                            | 1113   |
| 117* | n-butanol-2-ol                       | 1000   |

---

---

|      |   |        |
|------|---|--------|
| 118  | 2-butanoylfuran                                   | 1644   |
| 119^ | 2-butanoyl-5-methylfuran                          | 1748   |
| 120  | 2-butanoylthiophene                               | 1894   |
| 121* | n-butyl acetate                                   | 1059   |
| 122  | iso-butyl acetate                                 | 1000   |
| 123  | iso-butyl aldehyde                                | 800    |
| 124* | iso-butyl aldehyde-1-phenyl,1,2-ethanediol acetal | 1900   |
| 125  | iso-butyl aldehyde-propylene glycol acetal        | 1067   |
| 126^ | n-butyl anisate                                   | 2305   |
| 127* | iso-butyl anisate                                 | 2233   |
| 128  | n-butyl antranilate                               | 2419   |
| 129^ | iso-butyl antranilate                             | 2347   |
| 130* | n-butyl benzoate                                  | 1841   |
| 131* | iso-butyl benzoate                                | 1771   |
| 132^ | iso-butyl benzyl carbinol                         | 1983   |
| 133^ | n-butyl n-butyrate                                | 1207   |
| 134* | n-butyl iso-butyrate                              | 1139   |
| 135^ | iso-butyl n-butyrate                              | 1152   |
| 136* | iso-butyl iso-butyrate                            | 1084   |
| 137^ | n butyl n-butyru lactate                          | 1733   |
| 138^ | iso-butyl cinnamate                               | 2228   |
| 139^ | o-ter-butyl cyclohexyl acetate                    | 1580.5 |
| 140  | p-tert-butyl cyclohexyl-acetate trans             | 1628   |
| 141* | p-tert-butyl cyclohexyl-acetate cis               | 1675   |
| 142  | p-tert-butyl cyclohexanone                        | 1645   |
| 143^ | n-Butyl n-decanoate                               | 1798   |
| 144* | 2-iso-Butyl-4,5-dimethyloxazole                   | 1330   |
| 145  | 2-n-Butyl-4-5-dimethylthiazole                    | 1600   |
| 146  | 2-iso-Butyl-4,5-dimethylthizole                   | 1517   |
| 147* | n-Butyl n-dodecanoate                             | 2000   |
| 148* | iso-Butyl 2-ethyl n-hexanoate                     | 1400   |
| 149* | 2-n-Butyl-4-ethyl-5-methyloxazole                 | 1441   |
| 150  | 2-iso-Butyl-4-ethyl-5-methyloxazole               | 1359   |
| 151^ | n-Butyl formate                                   | 996    |
| 152* | iso-Butyl formate                                 | 955    |
| 153^ | 2-n-Butylfuran                                    | 1130   |
| 154* | iso-Butyl beta-2-furyl acrylate                   | 2006   |
| 155^ | n-Butyl furyl n-propionate                        | 1728   |
| 156^ | n-Butyl n-heptanoate                              | 1500   |
| 157* | iso-Butyl n-heptanoate                            | 1448   |

---

---

|      |  |      |
|------|--|------|
| 158* | n-Butyl n-hexanoate  | 1402 |
| 159  | iso-Butyl n-hexanoate  | 1350 |
| 160* | tert-Butyl-4-hydroxy anisole                                     | 2424 |
| 161^ | n-Butyl levulinate   | 1760 |
| 162* | iso-Butyl levulinate   | 1696 |
| 163  | iso-Butyl n-methyl anthranilate                                  | 2174 |
| 164^ | n-Butyl 2-methyl butyrate  | 1226 |
| 165^ | iso-Butyl 2-methyl butyrate                                      | 1171 |
| 166^ | p-tert-Butyl-alpha-methyl-hydrocinnamic aldehyde (lily aldehyde) | 2039 |
| 167  | n-Butyl methyl phenyl-glycidate                                  | 2273 |
| 168^ | 2-n-Butyl-4-methylthiazole                                       | 1500 |
| 169* | 2-iso-Butyl-4-methylthiazole                                     | 1420 |
| 170^ | n-Butyl n-octanoate  | 1600 |
| 171* | iso-Butyl n-octanoate  | 1543 |
| 172^ | n-Butyl phenyl acetate   | 1932 |
| 173* | iso-Butyl phenyl acetate   | 1864 |
| 174  | p-tert-Butyl phenylpropyl-aldehyde                               | 2030 |
| 175^ | iso-Butyl pivalate   | 1085 |
| 176  | n-Butyl n-propionate   | 1130 |
| 177  | iso-Butyl n-propionate   | 1071 |
| 178  | 6-sec-Butyl quinoline  | 2235 |
| 179  | n-Butyl salicylate   | 1976 |
| 180* | iso-Butyl salicylate   | 1896 |
| 181  | 2-iso-butylthiazole  | 1404 |
| 182* | 2-n-butylthiophene   | 1353 |
| 183^ | n-Butyl tiglate  | 1419 |
| 184* | iso-Butyl tiglate  | 1357 |
| 185  | n-Butyl n-undecanoate  | 1900 |
| 186* | n-Butyl 10-undecenoate   | 1954 |
| 187^ | n-Butyl iso-valerate   | 1242 |
| 188* | iso-Butyl iso-valerate   | 1184 |
| 189  | iso-Butyrophenone  | 1685 |
| 190* | Camphor  | 1518 |
| 191^ | Carvacrol  | 2159 |
| 192^ | Carveol  | 1805 |
| 193* | Carvone  | 1715 |
| 194  | Caryophyllene  | 1617 |
| 195* | Cedrol   | 2100 |
| 196^ | Cinnamic aldehyde  | 1996 |
| 197^ | Cinnamyl acetate   | 2103 |

---

---

|      |   |      |
|------|---|------|
| 198^ | Cinnamyl alcohol                            | 2207 |
| 199^ | Cinnamyl n-butyrate                         | 2247 |
| 200  | Cinnamyl iso-butyrate                       | 2171 |
| 201^ | Cinnamyl n-hexanoate                        | 2445 |
| 202^ | Cinnamyl n-propionate                       | 2169 |
| 203^ | Cinnamyl n-valerate                         | 2347 |
| 204^ | Cinnamyl iso-valerate                       | 2271 |
| 205* | Citral                                      | 1685 |
| 206^ | Citronellal                                 | 1465 |
| 207* | Citronellol                                 | 1722 |
| 208  | Citronellyl acetate                         | 1645 |
| 209  | Citronellyl n-butyrate                      | 1786 |
| 210^ | Citronellyl iso-butryrate                   | 1705 |
| 211  | Citronellyl formate                         | 1600 |
| 212^ | Citronellyl n-propionate                    | 1700 |
| 213^ | Citronellyl iso-valerate                    | 1800 |
| 214  | Coumarin                                    | 2361 |
| 215* | p-Cresol methyl ether                       | 1415 |
| 216* | p-Cresyl acetate                            | 1700 |
| 217  | p-Cresyl iso-butyrate                       | 1763 |
| 218^ | p-Cresyl n-octanoate                        | 2264 |
| 219* | Cumin alcohol                               | 2045 |
| 220  | Cumin aldehyde                              | 1766 |
| 221  | Cyclohexane glycidate                       | 1875 |
| 222  | Cyclohexanol                                | 1375 |
| 223^ | Cyclohexanone                               | 1306 |
| 224  | Cyclohexanone-1,3-butylene ketal            | 1500 |
| 225* | Cyclohexanone-1-phenyl-1,2-ethanodiol ketal | 2268 |
| 226^ | Cyclohexil acetate                          | 1343 |
| 227* | Cyclohexyl n-butyrate                       | 1492 |
| 228^ | Cyclohexyl iso-butyrate                     | 1427 |
| 229* | 2-Cyclohexylethyl acetate                   | 1591 |
| 230  | Cyclohexyl formate                          | 1305 |
| 231  | Cyclohexyl n-hexanoate                      | 1695 |
| 232  | Cyclohexyl n-propionate                     | 1408 |
| 233^ | Cyclohexyl iso-valerate                     | 1527 |
| 234* | trans-2-trans-4-Decadienol                  | 1938 |
| 235* | Decahydro beta-naphthol                     | 1918 |
| 236  | gamma-Decalactone                           | 2101 |
| 237  | delta-Decalactone                           | 2144 |

---

---

|      |                                   |      |
|------|-----------------------------------|------|
| 238* | n-Decanal                         | 1485 |
| 239^ | n-Decanal diethyl acetal          | 1613 |
| 240^ | n-Decanal dimethyl acetal         | 1567 |
| 241  | n-Decane                          | 1000 |
| 242  | n-Decanol                         | 1723 |
| 243  | cis-4-Decenal                     | 1523 |
| 244* | 1-Decene                          | 1043 |
| 245^ | 9-Decen-1-al                      | 1775 |
| 246^ | 9-Decen-1-yl acetate              | 1722 |
| 247* | n-Decyl acetate                   | 1662 |
| 248  | n-Decyl n-propionate              | 1729 |
| 249  | Di-iso-amyl ether                 | 1064 |
| 250* | Di-n-amyl ketone                  | 1528 |
| 251^ | Di-n-Butyl ether                  | 965  |
| 252* | Di-n-butyl fumarate               | 2006 |
| 253^ | Di-n-butyl ketone                 | 1330 |
| 254* | Di-iso-butyl ketone               | 1207 |
| 255^ | Diethyl adipate                   | 1858 |
| 256  | Diethyleneglycol diethyl ether    | 1572 |
| 257* | Diethyleneglycol dimethyl ether   | 1396 |
| 258^ | Diethyleneglycol monoethyl ether  | 1583 |
| 259  | Diethyleneglycol monomethyl ether | 1321 |
| 260* | Diethyl ether                     | 590  |
| 261  | Diethyl ketone                    | 980  |
| 262* | Diethyl malonate                  | 1542 |
| 263* | Diethyl-5-methyloxazole           | 1274 |
| 264  | Diethyl phthalate                 | 2303 |
| 265  | Diethyl sebacate                  | 2272 |
| 266  | Diethyl suberate                  | 2065 |
| 267* | Diethyl succinate                 | 1642 |
| 268* | Di-n-hexyl ketone                 | 1728 |
| 269^ | Dihydro anethole                  | 1600 |
| 270* | Dihydrocarveol                    | 1713 |
| 271  | Dihydro-nor-dicyclopentadienyl    | 1881 |
| 272^ | Dihydro iso-jasmone               | 1842 |
| 273  | Dihydro myrcenol                  | 1438 |
| 274^ | Dihydro terpinyl acetate          | 1561 |
| 275* | 1,4-Dimethoxy benzene             | 1705 |
| 276* | 1,4-Dimethoxy-2-tert-butylbenzene | 1870 |
| 277* | 2,4-Dimethyl-5-acetylthiazole     | 1835 |

---

---

|      |                                     |      |
|------|-------------------------------------|------|
| 278  | Dimethyl adipate                    | 1779 |
| 279* | Dimethyl benzyl carbinol            | 1715 |
| 280  | Dimethyl benzyl carbonyl acetate    | 1755 |
| 281  | Dimethyl benzyl carbonyl n-butylate | 1889 |
| 282  | Dimethyl benzyl carbonyl propionate | 1810 |
| 283  | 2,5-Dimethyl-4-ethyloxazole         | 1231 |
| 284* | Dimethyl fumarate                   | 1530 |
| 285* | 2,5-Dimethylfuran                   | 951  |
| 286^ | 2,6-Dimethyl n-heptan-2-ol          | 1300 |
| 287^ | 2,6-Dimethyl hept-5-en-1-ol         | 1358 |
| 288  | 2,6-Dimethylheptyl-4-acetate        | 1265 |
| 289  | Dimethyl malonate                   | 1472 |
| 290^ | Dimethyl phenyl ethyl carbinol      | 1916 |
| 291  | 2,3-Dimethyl pyrazine               | 1330 |
| 292  | 2,5-Dimethyl pyrazine               | 1306 |
| 293^ | Dimethyl sebacate                   | 2195 |
| 294  | Dimethyl succinate                  | 1558 |
| 295  | 2,4-Dimethylthiazole                | 1271 |
| 296* | 4,5-Dimethylthiazole                | 1359 |
| 297  | Diphenyl oxide                      | 1991 |
| 298^ | Di-n-propyl ketone                  | 1131 |
| 299  | Di-iso-propyl ketone                | 1007 |
| 300* | Di-n-propyl malonate                | 1700 |
| 301  | gamma-Dodecalactone                 | 2317 |
| 302^ | delta-Dodecalactone                 | 2358 |
| 303* | n-Dodecanal                         | 1695 |
| 304* | n-Dodecanal dimethyl acetal         | 1769 |
| 305^ | n-Dodecane                          | 1200 |
| 306* | n-Dodecanol                         | 1925 |
| 307  | n-Eicosane                          | 2000 |
| 308^ | Estragole                           | 1652 |
| 309  | p-Ethoxy benzaldehyde               | 2017 |
| 310  | Ethyl acetate                       | 872  |
| 311* | Ethyl acetyl acetate                | 1427 |
| 312^ | Ethyl acrylate                      | 980  |
| 313  | Ethyl alcohol                       | 900  |
| 314^ | Ethyl n-amyl ketone                 | 1190 |
| 315  | Ethyl anisate                       | 2110 |
| 316^ | Ethyl anthranilate                  | 2232 |
| 317^ | Ethyl benzoate                      | 1647 |

---

---

|      |                                 |        |
|------|---------------------------------|--------|
| 318  | Ethyl n-butyl ketone            | 1155   |
| 319^ | Ethyl n-butyrate                | 1025   |
| 320* | Ethyl iso-butyrate              | 956    |
| 321  | Ethyl cloro acetate             | 1281   |
| 322* | Ethyl cinnamate                 | 2095   |
| 323^ | Ethyl crotonate                 | 1161   |
| 324* | Ethyl n-decanoate               | 1624   |
| 325* | 2-Ethyl-4,5-dimethyloxazole     | 1243   |
| 326  | 2-Ethyl-4,5-dimethylthiazole    | 1429   |
| 327* | Ethyl n-deodecanoate            | 1826   |
| 328* | Ethyleneglycol monophenyl ether | 2100   |
| 329  | Ethyl formate                   | 806    |
| 330* | 2-Ethylfuran                    | 951    |
| 331* | Ethyl 2-furoate                 | 1599   |
| 332^ | Ethyl n-heptanoate              | 1321   |
| 333  | Ethyl n-hexanoate               | 1223   |
| 334  | Ethyl iso-hexanoate             | 1181   |
| 335* | Ethyl levulinate                | 1567   |
| 336* | Ethyl 2-methyl butyrate         | 1049   |
| 337* | 2-Ethyl-4-methylthiazole        | 1331   |
| 338^ | 4-Ethyl-5-methylthiazole        | 1400   |
| 339* | Ethyl n-nonanoate               | 1523   |
| 340  | Ethyl n-octanoate               | 1423   |
| 341  | Ethyl phenyl acetate            | 1773   |
| 342  | Ethyl pivalate                  | 947    |
| 343^ | Ethyl n-propionate              | 944    |
| 344  | Ethyl n-propyl ketone           | 1055   |
| 345  | Ethyl salicylate                | 1787   |
| 346^ | Ethyl n-tetradecanoate          | 2027   |
| 347  | 2-Ethylthiazole                 | 1300   |
| 348^ | 2-Ethylthiophene                | 1179   |
| 349^ | Ethyl tiglate                   | 1234   |
| 350* | Ethyl n-undecanoate             | 1725   |
| 351^ | Ethyl 10-undecenoate            | 1775   |
| 352^ | Ethyl n-valerate                | 1124   |
| 353* | Ethyl iso-valerate              | 1060   |
| 354^ | Ethyl vanillin                  | 2414   |
| 355  | Eugenol                         | 2103   |
| 356* | iso-Eugenol                     | 2227.5 |
| 357  | Furfural                        | 1449   |

---

---

|      |                                  |        |
|------|----------------------------------|--------|
| 358  | Furfuryl acetate                 | 1518   |
| 359^ | Furfuryl n-butyrate              | 1655   |
| 360* | Furfuryl n-hexanoate             | 1850   |
| 361* | Galaxolide                       | 2303   |
| 362^ | Geraniol                         | 1797   |
| 363^ | Gerano nitrile                   | 1701.5 |
| 364^ | Geranyl acetate                  | 1735   |
| 365  | Geranyl n-butyrate               | 1872   |
| 366  | Geranyl iso-butyrate             | 1795   |
| 367^ | Geranyl ethyl ether              | 1491   |
| 368* | Geranyl formate                  | 1684   |
| 369^ | Geranyl n-hexanoate              | 2057   |
| 370^ | Geranyl n-propionate             | 1799   |
| 371^ | Geranyl iso-valerate             | 1895   |
| 372  | Helional                         | 2383   |
| 373* | Heliotropine                     | 2171   |
| 374* | n-Heneicosane                    | 2100   |
| 375^ | n-Heptadecane                    | 1700   |
| 376* | n-Heptanal                       | 1186   |
| 377^ | n-Heptanal diethyl acetal        | 1319   |
| 378* | n-Heptanal dimethyl acetal       | 1265   |
| 379* | n-Heptanal ethyleneglycol acetal | 1460   |
| 380  | n-Heptane                        | 700    |
| 381* | 2,3-Heptanedione                 | 1138   |
| 382^ | n-Heptanol                       | 1419   |
| 383* | n-Heptan-2-ol                    | 1284   |
| 384  | n-Heptan-4-ol                    | 1250   |
| 385  | 2-Heptanoylfuran                 | 1956   |
| 386* | 2-Heptanoylthiophene             | 2209   |
| 387^ | 1-Hepten-3-ol                    | 1322   |
| 388^ | n-Heptyl acetate                 | 1361   |
| 389^ | n-Heptyl n-butyrate              | 1503   |
| 390  | n-Heptyl iso butyrate            | 1433   |
| 391  | n-Heptyl formate                 | 1310   |
| 392* | 2-n-Heptylfuran                  | 1429   |
| 393^ | n-Heptyl n-octanoate             | 1892   |
| 394* | 2-n-Heptylthiophene              | 1670   |
| 395* | n-Hexadecane                     | 1600   |
| 396* | n-Hexanal                        | 1084   |
| 397^ | n-Hexanal diethyl acetal         | 1223   |

---

---

|      |                                       |      |
|------|---------------------------------------|------|
| 398* | n-Hexane                              | 600  |
| 399  | 3,4-Hexanedione                       | 1123 |
| 400* | n-Hexanol                             | 1316 |
| 401^ | n-Hexan-2-ol                          | 1192 |
| 402* | 2-Hexanoylfuran                       | 1850 |
| 403* | 2-Hexanoylthiophene                   | 2104 |
| 404^ | trans-2-Hexenal                       | 1207 |
| 405* | 1-Hexen-3-ol                          | 1225 |
| 406^ | cis-3-Hexenol                         | 1351 |
| 407  | trans-2-Hexenol                       | 1368 |
| 408  | cis-3-Hexenyl acetate                 | 1300 |
| 409* | trans-2-Hexenyl acetate               | 1315 |
| 410  | cis-3-Hexenyl n-butyrate              | 1450 |
| 411^ | trans-2-Hexenyl n-butyrate            | 1461 |
| 412^ | cis-3-Hexenyl iso-butyrate            | 1377 |
| 413^ | cis-3-Hexenyl n-decanoate             | 2038 |
| 414  | cis-3-Hexenyl formate                 | 1252 |
| 415  | cis-3-Hexenyl n-hexanoate             | 1654 |
| 416* | trans-2-hexenyl n-hexanoate           | 1656 |
| 417  | cis-3-Hexenyl lactate                 | 1727 |
| 418* | cis-3-Hexenyl methoxy formate         | 1475 |
| 419^ | cis-3-Hexenyl n-octanoate             | 1838 |
| 420^ | trans-2-Hexenyl n-octanoate           | 1853 |
| 421^ | cis-3-Hexenyl n-propionate            | 1371 |
| 422  | trans-2-Hexenyl n-propionate          | 1370 |
| 423^ | cis-3-Hexenyl salicylate              | 2227 |
| 424  | cis-3-Hexenyl n-valerate              | 1584 |
| 425^ | cis-3-Hexenyl iso valerate            | 1480 |
| 426* | 2-Hexoxy acetaldehyde dimethyl acetal | 1528 |
| 427^ | n-Hexyl acetate                       | 1307 |
| 428  | n-Hexyl n-butyrate                    | 1398 |
| 429^ | n-Hexyl iso-butyrate                  | 1333 |
| 430* | alpha-n-Hexyl cinnamic aldehyde       | 2309 |
| 431^ | n-Hexyl formate                       | 1258 |
| 432^ | 2-Hexylfuran                          | 1326 |
| 433* | n-Hexyl n-hexanoate                   | 1599 |
| 434^ | n-Hexyl 2-methyl butyrate             | 1418 |
| 435^ | n-Hexyl pivalate                      | 1328 |
| 436^ | n-Hexyl n-propionate                  | 1326 |
| 437* | n-Hexyl salicylate                    | 2175 |

---

---

|      |                            |      |
|------|----------------------------|------|
| 438* | 2-n-Hexylthiophene         | 1564 |
| 439* | n-Hexyl iso-valerate       | 1433 |
| 440  | Hydroxy citronellal        | 1882 |
| 441  | Indole                     | 2351 |
| 442  | alpha-Ionone               | 1833 |
| 443* | Beta-Ionone                | 1918 |
| 444^ | cis-Jasmone                | 1914 |
| 445^ | Lavandulyl acetate         | 1597 |
| 446* | Limonene                   | 1206 |
| 447* | Linalool                   | 1506 |
| 448* | cis-Linalool oxide         | 1423 |
| 449  | trans-Linalool oxide       | 1451 |
| 450  | Linalyl acetate            | 1538 |
| 451^ | Linalyl iso-butyrate       | 1597 |
| 452  | Linalyl formate            | 1570 |
| 453* | Linalyl n-hexanoate        | 1843 |
| 454* | Linalyl n-propionate       | 1596 |
| 455  | Linalyl n-valerate         | 1765 |
| 456* | Linalyl iso-valerate       | 1698 |
| 457^ | Menthol                    | 1612 |
| 458  | m-Methoxy acetophenone     | 2011 |
| 459* | o-Methoxy acetophenone     | 1975 |
| 460* | p-Methoxy acetophenone     | 2115 |
| 461  | m-Methyl acetophenone      | 1749 |
| 462^ | o-Methyl acetophenone      | 1679 |
| 463* | p-Methyl acetophenone      | 1750 |
| 464  | 2-Methyl-5-acetylthiophene | 1836 |
| 465* | Methyl n-amyl ketone       | 1172 |
| 466* | Methyl iso-amyl ketone     | 1100 |
| 467* | Methyl anisate             | 2071 |
| 468* | Methyl anthranilate        | 2181 |
| 469  | Methyl benzoate            | 1600 |
| 470* | 3-Methylbutan-2-ol         | 1052 |
| 471^ | Methyl n-butyl ketone      | 1070 |
| 472^ | Methyl iso-butyl ketone    | 1000 |
| 473  | Methyl n-butyrate          | 975  |
| 474* | Methyl iso-butyrate        | 913  |
| 475^ | Methyl cinnamate           | 2051 |
| 476* | Methyl crotonate           | 1100 |
| 477^ | Methyl n-decanoate         | 1581 |

---

---

|                  |   |      |
|------------------|---|------|
| 478 <sup>^</sup> | Methyl n-dodecanoate  | 1785 |
| 479              | 2-Methyl-5-ethylfuran   | 1024 |
| 480 <sup>*</sup> | 2-Methyl-3-ethyl pyrazine   | 1381 |
| 481              | Methyl iso-eugenol  | 2085 |
| 482 <sup>^</sup> | 2-Methylfuran   | 866  |
| 483 <sup>^</sup> | 5-Methyl furfural   | 1563 |
| 484              | Methyl furoate  | 1561 |
| 485 <sup>^</sup> | Methyl n-heptanoate   | 1276 |
| 486 <sup>*</sup> | 6-Methyl hept-5-en-2-one  | 1335 |
| 487              | Methyl n-heptyl ketone  | 1377 |
| 488 <sup>^</sup> | Methyl n-hexanoate  | 1177 |
| 489              | Methyl n-hexyl ketone   | 1275 |
| 490              | Methyl n-hexyl ketone-1-phenyl-1,2-ethanediol ketal                 | 2236 |
| 491 <sup>^</sup> | alpha-n-Methyl ionone   | 1930 |
| 492              | beta-n-Methyl ionone  | 1981 |
| 493              | alpha-iso-Methyl ionone   | 1836 |
| 494 <sup>^</sup> | beta-iso-Methyl ionone  | 1897 |
| 495              | Methyl levulinate   | 1534 |
| 496 <sup>^</sup> | Methyl n-methyl anthranilate  | 2042 |
| 497              | Methyl 2-methyl butyrate  | 1000 |
| 498              | beta-Methyl naphthyl ketone   | 2471 |
| 499 <sup>*</sup> | Methyl n-nonanoate  | 1479 |
| 500              | Methyl n-nonyl ketone   | 1585 |
| 501 <sup>*</sup> | Methyl n-octanoate  | 1378 |
| 502 <sup>^</sup> | Methyl n-octyl ketone   | 1480 |
| 503 <sup>^</sup> | Methyl phenylacetate  | 1747 |
| 504 <sup>^</sup> | Methyl phenyl carbinol  | 1765 |
| 505 <sup>^</sup> | 2-Methyl-1-propanol   | 1054 |
| 506 <sup>^</sup> | 2-Methyl n-propan-2-ol  | 871  |
| 507 <sup>^</sup> | 2-Methyl-5-propionylthiophene                                       | 1900 |
| 508 <sup>*</sup> | Methyl n-propyl ketone  | 969  |
| 509 <sup>*</sup> | beta-Methyl-p-iso-propyl phenyl propionaldehyde (cyclamen aldehyde) | 1954 |
| 510              | 2-Methyl-3-iso-propyl pyrazine                                      | 1387 |
| 511 <sup>^</sup> | 2-Methyl pyrazine   | 1251 |
| 512 <sup>*</sup> | Methyl salicylate (wintergreen)                                     | 1754 |
| 513 <sup>^</sup> | Methyl n-tetradecanoate   | 1990 |
| 514 <sup>*</sup> | 4-Methylthiazole  | 1263 |
| 515 <sup>^</sup> | 4-Methyl-5-thiazole ethanol   | 2216 |
| 516              | Methyl p-toluate  | 1725 |
| 517 <sup>*</sup> | 2-Methyl undecanal  | 1609 |

---

---

|      |                           |        |
|------|---------------------------|--------|
| 518  | Methyl 10–undecenoate     | 1733   |
| 519  | Methyl n–undecyl ketone   | 1792   |
| 520^ | Methyl n–valerate         | 1076   |
| 521* | Methyl iso–valerate       | 1008   |
| 522* | Muscone                   | 2281   |
| 523  | Musk xylol                | 2475   |
| 524^ | (S)–Myrac aldehyde        | 1959   |
| 525* | (R)–Myrac aldehyde        | 1978   |
| 526* | Myrcene                   | 1156   |
| 527  | Nerol                     | 1757   |
| 528  | Nerolidol                 | 1980.5 |
| 529* | Neryl acetate             | 1699   |
| 530* | Neryl n–butyrate          | 1838   |
| 531* | Neryl iso–butyrate        | 1764   |
| 532  | Neryl n–propionate        | 1771   |
| 533^ | n–Nonadecane              | 1900   |
| 534  | gamma–Nonalactone         | 1991   |
| 535^ | delta–Nonalactone         | 2038   |
| 536  | n–Nonanal                 | 1382   |
| 537  | n–Nonanal diethyl acetal  | 1514   |
| 538  | n–Nonanal dimethyl acetal | 1465   |
| 539* | n–Nonane                  | 900    |
| 540* | n–Nonanol                 | 1624   |
| 541  | n–Nonan–2–ol              | 1484   |
| 542* | 1–Nonen–3–ol              | 1520   |
| 543  | Nopyl acetate             | 1777   |
| 544  | n–Octadecane              | 1800   |
| 545* | gamma–Octalactone         | 1883   |
| 546  | delta–Octalactone         | 1929   |
| 547* | n–Octanal                 | 1278   |
| 548* | n–Octanal diethyl acetal  | 1417   |
| 549^ | n–Octanal dimethyl acetal | 1366   |
| 550^ | n–Octane                  | 800    |
| 551* | n–Octanol                 | 1519   |
| 552^ | n–Octan–2–ol              | 1385   |
| 553^ | 2–Octanoylfuran           | 2062   |
| 554^ | 2–Octanoylthiophene       | 2313   |
| 555  | 1–Octen–3–ol              | 1420   |
| 556* | 1–Octen–3–yl–n–propionate | 1432   |
| 557^ | n–Octyl–acetate           | 1459   |

---

---

|      |  |      |
|------|--|------|
| 558  | n-Octyl n-butyrate                       | 1597 |
| 559^ | n-Octyl iso-butyrate                     | 1529 |
| 560  | 2-n-Octylfuran                           | 1530 |
| 561* | n-Octyl 2-methyl butyrate                | 1615 |
| 562* | 2-n-Octylthiophene                       | 1780 |
| 563  | Paraldehyde                              | 1069 |
| 564  | n-Pentadecane                            | 1500 |
| 565  | n-Pentane                                | 500  |
| 566* | 2,3-Pentanedione                         | 1044 |
| 567^ | 2-Pentanoylthiophene                     | 1993 |
| 568* | 1-Penten-3-ol                            | 1130 |
| 569  | 2-n-Pentylfuran                          | 1229 |
| 570  | 2-n-Pentyl-3-methyl-2-cyclopenten-1-one  | 1892 |
| 571  | 2-n-Pentylthiophene                      | 1462 |
| 572  | beta-Phenoxyethyl iso-butyrate           | 2100 |
| 573  | Phenol                                   | 1932 |
| 574* | Phenylacetaldehyde dimethyl acetal       | 1665 |
| 575^ | Phenyl iso-butyl methyl-carbinyl acetate | 1943 |
| 576* | Phenylethyl acetate                      | 1785 |
| 577* | Phenylethyl alcohol                      | 1859 |
| 578^ | Phenylethyl iso-butyrate                 | 1855 |
| 579^ | Phenylethyl formate                      | 1752 |
| 580^ | Phenylethyl n-heptanoate                 | 2233 |
| 581^ | Phenylethyl n-hexanoate                  | 2134 |
| 582^ | Phenylethyl 2-methyl butyrate            | 1945 |
| 583  | Phenylethyl n-nonanoate                  | 2439 |
| 584  | Phenylethyl n-octanoate                  | 2337 |
| 585* | Phenylethyl pivalate                     | 1832 |
| 586^ | Phenylethyl n-propionate                 | 1855 |
| 587^ | Phenylethyl tiglate                      | 2154 |
| 588  | Phenylethyl n-valerate                   | 2034 |
| 589  | Phenylethyl iso-valerate                 | 1955 |
| 590* | Phenylpropyl acetate                     | 1926 |
| 591  | 3-Phenylpropyl alcohol                   | 1993 |
| 592^ | Phenylpropyl n-butyrate                  | 2083 |
| 593* | Phenylpropyl iso-butyrate                | 1996 |
| 594^ | Phenylpropyl n-propionate                | 1994 |
| 595^ | alpha-Pinene                             | 1039 |
| 596^ | n-Propanol                               | 1002 |
| 597* | 2-Propanol                               | 884  |

---

---

|      |  |        |
|------|--|--------|
| 598  | Propionaldehyde-1-phenyl-1,2-ethenodiol acetal | 1875.5 |
| 599  | 2-Propionylfuran                               | 1563   |
| 600* | 2-Propionyl 5-methylfuran                      | 1672   |
| 601* | 2-Propionylthiophene                           | 1821   |
| 602  | n-Propyl acetate                               | 962    |
| 603^ | iso-Propyl acetate                             | 883    |
| 604^ | iso-Propyl benzoate                            | 1639   |
| 605* | n-Propyl n-butyrate                            | 1110   |
| 606^ | n-Propyl iso-butyrate                          | 1044   |
| 607^ | iso-Propyl n-butyrate                          | 1030   |
| 608* | iso-Propyl cinnamate                           | 2097   |
| 609^ | 2-n-Propyl-4,5-dimethyloxazole                 | 1310   |
| 610* | 2-iso-Propyl-4,5-dimethyloxazole               | 1249   |
| 611* | 2-n-Propyl-4,5-dimethylthiazole                | 1500   |
| 612* | 2-iso-Propyl-4,5-dimethyloxazole               | 1439   |
| 613^ | 2-n-Propyl-4-ethyl-5-methyloxazole             | 1345   |
| 614  | 2-iso-Propyl-4-ethyl-5-methyloxazole           | 1279   |
| 615  | n-Propyl formate                               | 907    |
| 616  | iso-Propyl formate                             | 838    |
| 617^ | 2-n-Propylfuran                                | 1083   |
| 618^ | n-Propyl n-heptanoate                          | 1398   |
| 619  | iso-Propyl n-heptanoate                        | 1317   |
| 620* | n-Propyl n-hexanoate                           | 1298   |
| 621* | iso-Propyl n-hexanoate                         | 1223   |
| 622  | n-Propyl levulinate                            | 1663   |
| 623^ | iso-Propyl levulinate                          | 1575   |
| 624* | n-Propyl n-methyl anthranilate                 | 2166   |
| 625^ | iso-Propyl n-methyl anthranilate               | 2058   |
| 626  | n-Propyl 2-methyl butyrate                     | 1134   |
| 627^ | 2-n-Propyl-4-methylthiazole                    | 1400   |
| 628  | n-Propyl n-octanoate                           | 1498   |
| 629^ | iso-propyl n-octanoate                         | 1419   |
| 630* | iso-Propyl pivalate                            | 956    |
| 631  | n-Propyl n-propionate                          | 1010   |
| 632  | iso-Propyl n-propionate                        | 950    |
| 633* | n-Propyl salicylate                            | 1878   |
| 634* | iso-Propyl n-tetradacanoate                    | 2017   |
| 635  | 2-n-Propylthiazole                             | 1380   |
| 636  | 2-n-Propylthiophene                            | 1259   |
| 637* | iso-Propyl tiglate                             | 1238   |

---

---

|      |  |      |
|------|--|------|
| 638^ | n-Propyl 10-undecenoate                              | 1860 |
| 639  | n-Propyl n-valerate                                  | 1200 |
| 640  | n-Propyl iso-valerate                                | 1144 |
| 641  | iso-Propyl iso-valerate                              | 1034 |
| 642  | Pyrazine   | 1194 |
| 643  | cis-Rose oxide                                       | 1354 |
| 644^ | trans-Rose oxide                                     | 1370 |
| 645  | Rosephenone trichloro methyl phenyl-carbinyl acetate | 2172 |
| 646  | Sabinete   | 1130 |
| 647^ | Salicylic aldehyde                                   | 1668 |
| 648^ | Styrallyl acetate                                    | 1673 |
| 649  | alpha-Terpineol                                      | 1661 |
| 650  | Terpinyl acetate                                     | 1687 |
| 651  | Terpinyl n-butyrate                                  | 1828 |
| 652* | Terpinyl iso-butyrate                                | 1748 |
| 653^ | Terpinyl n-propionate                                | 1747 |
| 654* | Terpinyl iso-valerate                                | 1858 |
| 655^ | n-Tetradecane  | 1400 |
| 656^ | Tetrahydro geraniol                                  | 1626 |
| 657* | Tetrahydro geranyl acetate                           | 1562 |
| 658^ | Tetrahydro linalool                                  | 1397 |
| 659* | Tetrahydro myrcenol                                  | 1414 |
| 660* | 2,3,5,6-Tetramethyl pyrazine                         | 1458 |
| 661* | Thiophene  | 1035 |
| 662  | Thymol   | 2100 |
| 663  | Tonalid  | 2373 |
| 664  | n-Tridecane  | 1300 |
| 665  | Triethyl citrate                                     | 2386 |
| 666* | 3,5,5-Trimethyl n-hexanal                            | 1200 |
| 667  | 3,5,5-trimethyl n-hexanol                            | 1480 |
| 668* | 2,4,5-Trimethyloxazole                               | 1179 |
| 669  | 2,3,5-Trimethyl pyrazine                             | 1387 |
| 670^ | Trimethylthiazole                                    | 1367 |
| 671* | delta-Undecalactone                                  | 2251 |
| 672  | n-Undecanal  | 1589 |
| 673* | n-Undecane   | 1100 |
| 674^ | n-Undecanol  | 1822 |
| 675^ | 6-Undecanol  | 1640 |
| 676  | iso-Valeraldehyde-propyleneglycol acetal             | 1174 |
| 677^ | Vanillin   | 2449 |

---

---

|      |                         |      |
|------|-------------------------|------|
| 678  | Methyl ether            | 524  |
| 679* | Acetaldehyde            | 690  |
| 680  | Propionaldehyde         | 784  |
| 681  | Furan                   | 786  |
| 682  | tert-Butanol            | 830  |
| 683  | Propyl methyl ether     | 644  |
| 684* | Methyl acetate          | 813  |
| 685  | Methyl vinyl ketone     | 995  |
| 686* | Vinyl acetate           | 878  |
| 687  | tert-Butyl methyl ether | 688  |
| 688* | 2,3 Butanedione         | 963  |
| 689^ | 2-Butanone              | 908  |
| 690* | Allyl ethyl ether       | 767  |
| 691  | Isopropyl ether         | 649  |
| 692  | sec-Butanol             | 975  |
| 693* | Methyl propionate       | 895  |
| 694^ | Butyl methyl ether      | 755  |
| 695  | Tetrahydrofuran         | 898  |
| 696* | 3-Buten-1-ol            | 1137 |
| 697* | Isovaleraldehyde        | 937  |
| 698^ | 2-Butene-1-ol           | 1193 |
| 699* | Vinyl propionate        | 960  |
| 700* | Isobutyl mercaptan      | 889  |
| 701* | 2-Methyltetrahydrofuran | 901  |
| 702  | tert-Butyl acetate      | 893  |
| 703^ | Propyl ether            | 766  |
| 704  | Cyclohexane             | 765  |
| 705^ | Methyl pyruvate         | 1217 |
| 706  | n-Butyl ethyl ether     | 788  |
| 707  | Diethyl sulfide         | 904  |
| 708  | Tetrahydropyran         | 930  |
| 709* | Valeraldehyde           | 1002 |
| 710  | Pinacolone              | 960  |
| 711* | Pyridine                | 1180 |
| 712  | Methyl methacrylate     | 1008 |
| 713* | 3,4-Dihydropyran        | 982  |
| 714* | 1,1-Diethoxyethane      | 880  |
| 715  | N-Methylpyrrole         | 1139 |
| 716  | Thiazole                | 1246 |
| 717  | Methyl disulfide        | 1081 |

---

---

|      |                            |      |
|------|----------------------------|------|
| 718  | Vinyl butyrate             | 1045 |
| 719^ | sec-Butyl acetate          | 982  |
| 720^ | Ethyl isobutyrate          | 956  |
| 721^ | Methyl isobutyl carbinol   | 1142 |
| 722^ | Piperidine                 | 1042 |
| 723  | Propylene glycol           | 1561 |
| 724* | 3-Methyl-2-pentanol        | 1054 |
| 725  | 3-Methyl-3-pentanol        | 1080 |
| 726* | 4-Methyl-2-pentanol        | 1124 |
| 727^ | 2-Methylthiophene          | 1123 |
| 728  | 2-Methyl-3-pentanol        | 1121 |
| 729  | 2-Methylthiazole           | 1256 |
| 730^ | Isobutyl acrylate          | 1107 |
| 731* | Ethyl pyruvate             | 1253 |
| 732* | 1,3-Butylene glycol acetal | 1072 |
| 733* | 2-Methyl-2-pentanol        | 1160 |
| 734* | 3-Methyl-2-pentanol        | 1160 |
| 735  | 1-Octene                   | 830  |
| 736  | Tetrahydrothiophene        | 1130 |
| 737^ | Cyclopentanol              | 1283 |
| 738^ | Methyl mercaptoacetate     | 1346 |
| 739  | Cyclopentanone             | 1238 |
| 740* | Elemol                     | 1318 |
| 741^ | 3-Buten-2-ol               | 1022 |
| 742* | 2-Octene                   | 880  |
| 743* | n-Butyl methyl sulfide     | 1043 |
| 744* | Allyl isobutyrate          | 1090 |
| 745^ | Methyl hexyl ether         | 960  |
| 746  | 2-Methyl-1-pentanol        | 1268 |
| 747* | 4-Methyl-1-pentanol        | 1282 |
| 748  | Dimethyl sulfoxide         | 1554 |
| 749* | Methyl tiglate             | 1188 |
| 750^ | 3-Methyl-1-pentanol        | 1297 |
| 751* | Diallyl sulfide            | 1150 |
| 752  | Allyl isothiocyanate       | 1352 |
| 753* | p-Xylene                   | 1140 |
| 754  | n-Propyl pivalate          | 1028 |
| 755^ | m-Xylene                   | 1147 |
| 756  | 2,5-Dimethylthiophene      | 1161 |
| 757* | 3-Ethyl-3-pentanol         | 1183 |

---

---

|      |                            |      |
|------|----------------------------|------|
| 758* | Cyclohexenyl cyclohexanone | 2128 |
| 759^ | Tetrahydrofurfuryl alcohol | 1494 |
| 760  | o-Xylene                   | 1191 |
| 761* | gamma-Butyrolactone        | 1632 |
| 762  | Butyl acrylate             | 1189 |
| 763^ | n-Butyl mercaptan          | 944  |
| 764  | n-Butyl ethyl sulfide      | 1090 |
| 765* | 2,6-Dimethylpyrazine       | 1325 |
| 766^ | Prenyl acetate             | 1243 |
| 767* | Ethyl disulfide            | 1232 |
| 768^ | n-Hexyl mercaptan          | 1145 |
| 769  | Ethyl angelate             | 1228 |
| 770* | gamma-Valerolactone        | 1617 |
| 771^ | alpha-Thujene              | 1036 |
| 772* | 1,3-Butanediol             | 1692 |
| 773* | 2-Ethoxythiazole           | 1380 |
| 774* | Dimethyl trisulfide        | 1400 |
| 775  | Camphene                   | 1083 |
| 776^ | 2-Methyl-5-ethylthiophene  | 1245 |
| 777^ | 2-Ethylbutyl acetate       | 1205 |
| 778  | 2-Ethylhexanal             | 1216 |
| 779  | n-Butyl pivalate           | 1128 |
| 780^ | 1,1-Dimethoxy-n-hexane     | 1156 |
| 781^ | Allyl n-valerate           | 1256 |
| 782* | Amyl vinyl carbinol        | 1426 |
| 783  | Diisobutyl sulfide         | 1149 |
| 784^ | Methyl benzyl ether        | 1391 |
| 785* | trans-Pinane               | 1062 |
| 786  | beta-Pinene                | 1124 |
| 787* | cis-Dihydroocimene         | 1088 |
| 788  | 2,6-Dimethylcyclohexanone  | 1322 |
| 789^ | cis-Pinane                 | 1075 |
| 790* | tert-Butyl benzene         | 1247 |
| 791* | 2-Octanone                 | 1304 |
| 792  | Diethyl malate             | 1638 |
| 793^ | trans-Dihydroocimene       | 1110 |
| 794* | n-Butyl lactate            | 1508 |
| 795  | Methyl sorbate             | 1488 |
| 796  | Diisobutyl acetal          | 1072 |
| 797* | Isopentyl ether            | 1067 |

---

---

|      |                                     |      |
|------|-------------------------------------|------|
| 798  | Angelica lactone                    | 1430 |
| 799  | o-Methyl anisole                    | 1432 |
| 800  | p-Methyl anisole                    | 1432 |
| 801^ | 2-Isopropyl-4-methylthiazole        | 1339 |
| 802* | Dimethyl succinate                  | 1558 |
| 803  | alpha-Phellandrene                  | 1177 |
| 804^ | Ethyl 2-methylpentanoate            | 1302 |
| 805* | Pyrrol-2-carboxaldehyde             | 1976 |
| 806  | 1,4-Cineole                         | 1185 |
| 807* | n-Heptyl mercaptan                  | 1256 |
| 808* | 4-Methyl-5-vinylthiazole            | 1500 |
| 809  | Isopropyl disulfide                 | 1264 |
| 810  | Methyl octyl ether                  | 1152 |
| 811  | p-Cimene                            | 1272 |
| 812^ | 2-Isopropyl-4-ethyl-5-methyloxazole | 1279 |
| 813^ | Carvomenthene                       | 1150 |
| 814* | Phenylacetaldehyde                  | 1646 |
| 815  | cis-Ocimene                         | 1228 |
| 816^ | beta-Phellandrene                   | 1216 |
| 817* | 1,8-Cineole                         | 1228 |
| 818  | Isobutyl angelate                   | 1289 |
| 819* | 2-Ethyl-1-hexyl vinyl ether         | 1165 |
| 820^ | trans-Ocimene                       | 1250 |
| 821  | Isobutyl n-valerate                 | 1252 |
| 822^ | trans-2-Octenal                     | 1427 |
| 823* | 1,4-Butanediol                      | 1861 |
| 824^ | Methyl hexyl acetaldehyde           | 1306 |
| 825* | 2-Acetylpyrrole                     | 1935 |
| 826* | 2,5-Dimethyl-4-ethylthiazole        | 1398 |
| 827  | p-Cresol                            | 2003 |
| 828* | m-Tolualdehyde                      | 1632 |
| 829  | o-Tolualdehyde                      | 1632 |
| 830  | n-Methyl-2-acetylpyrrole            | 1653 |
| 831^ | Tetrahydrofurfuryl acetate          | 1585 |
| 832^ | gamma-Terpinene                     | 1251 |
| 833* | n-Pentyl ether                      | 1165 |
| 834  | p-Tolualdehyde                      | 1652 |
| 835  | m-Diethylbenzene                    | 1339 |
| 836^ | Guaiacol                            | 1840 |
| 837* | Di-n-butyl sulfide                  | 1270 |

---

---

|      |                             |      |
|------|-----------------------------|------|
| 838^ | 2-Methyl-3-n-propylpyrazine | 1462 |
| 839  | 5-Nonanone                  | 1360 |
| 840* | Ethyl sorbate               | 1505 |
| 841* | p-Diethylbenzene            | 1353 |
| 842^ | alpha-p-dimethylstyrene     | 1278 |
| 843* | Fenchone                    | 1410 |
| 844* | trans-3-Heptenyl acetate    | 1388 |
| 845^ | cis-3-Heptenyl acetate      | 1400 |
| 846* | 2,6-Dimethyl phenol         | 1883 |
| 847^ | o-Diethylbenzene            | 1372 |
| 848* | 2-Nonanone                  | 1420 |
| 849* | Amyl vinyl carbinyl acetate | 1365 |
| 850^ | cis-3-Hexenyl ethyl acetal  | 1298 |
| 851* | Propyl disulfide            | 1358 |
| 852  | 2-Cyclohexyl ethanol        | 1668 |
| 853^ | Myrcenol                    | 1585 |
| 854  | 2-Buten-1,4-diol            | 1983 |
| 855* | Butyl acetoacetate          | 1798 |
| 856* | Maltol                      | 2030 |
| 857^ | alpha-Fenchyl alcohol       | 1574 |
| 858  | Norbornyl acetate           | 1476 |
| 859* | p-tert-Amylcyclohexanone    | 1800 |
| 860  | 2-Ethyl phenol              | 2028 |
| 861  | Methyl 2-hydroxyisobutyrate | 2054 |
| 862* | Methyl nicotinate           | 1779 |
| 863  | Octyl formate               | 1426 |
| 864^ | 2-methyl-1-octanol          | 1573 |
| 865^ | Dihydrolinalool             | 1512 |
| 866* | p-Methyl benzyl alcohol     | 1956 |
| 867  | Diethyl trisulfide          | 1535 |
| 868  | Propyl furoate              | 1700 |
| 869  | Cyclooctanol                | 1700 |
| 870* | Sabinol                     | 1683 |
| 871  | Dihydrocinnamic aldehyde    | 1783 |
| 872^ | 1,3-Dimethoxy benzene       | 1740 |
| 873^ | Menthone                    | 1478 |
| 874^ | p-Hydroxyacetophenone       | 1790 |
| 875  | Isopulegol                  | 1573 |
| 876^ | 3-Ethyl phenol              | 2150 |
| 877  | trans-2-Nonenal             | 1540 |

---

---

|      |                               |      |
|------|-------------------------------|------|
| 878* | Tetrahydrofurfuryl propionate | 1632 |
| 879^ | Lavandulol                    | 1662 |
| 880* | Isooctyl acetate              | 1419 |
| 881* | Isobutyl 2-methylpentanoate   | 1417 |
| 882  | Isoborneol                    | 1660 |
| 883^ | trans-2-nonenol               | 1691 |
| 884^ | Diethyl fumarate              | 1632 |
| 885* | Borneol                       | 1698 |
| 886* | Methyl 2-methylbenzoate       | 1709 |
| 887^ | cis-3-Heptenyl propionate     | 1472 |
| 888  | Amyl butyl carbinol           | 1550 |
| 889^ | Terpinene-4-ol                | 1628 |
| 890* | Isobutyl 2-hexenoate          | 1461 |
| 891  | 2-Pentanoylfuran              | 1747 |
| 892  | Methyl chavicol               | 1670 |
| 893  | Dihydrocarvone                | 1600 |
| 894^ | Methyl 3-methylbenzoate       | 1744 |
| 895* | Verbenone                     | 1730 |
| 896^ | Methyl 4-methylbenzoate       | 1755 |
| 897* | p-isopropylphenol             | 2178 |
| 898* | Dihydromyrcenyl acetate       | 1431 |
| 899  | Isobutyl disulfide            | 1435 |
| 900^ | Fenchyl acetate               | 1473 |
| 901^ | Benzyl ethyl carbinol         | 1882 |
| 902  | 2-Ethyl-1-hexyl acrylate      | 1494 |
| 903^ | 2,4-Dimethyl benzyl alcohol   | 2032 |
| 904  | Neral                         | 1680 |
| 905  | Isobornyl formate             | 1596 |
| 906* | Pulegone                      | 1662 |
| 907^ | 2,4-Dimethylacetophenone      | 1809 |
| 908* | Citronellyl vinyl ether       | 1445 |
| 909* | n-Butyl benzyl ether          | 1613 |
| 910* | Chavicol                      | 2300 |
| 911* | Bornyl formate                | 1610 |
| 912^ | Myrcenyl acetate              | 1574 |
| 913^ | Piperitone                    | 1739 |
| 914^ | Geranial                      | 1730 |
| 915^ | Methyl octyl acetaldehyde     | 1521 |
| 916* | n-Propyl benzoate             | 1745 |
| 917^ | cis-3-Heptenyl butyrate       | 1545 |

---

---

|      |                                  |      |
|------|----------------------------------|------|
| 918^ | 2-decen-1-ol                     | 1794 |
| 919  | trans-2-Decenol                  | 1792 |
| 920  | Diisoamyl ketone                 | 1528 |
| 921* | Methyl dihydrocinnamate          | 1842 |
| 922* | Isopulegyl acetate               | 1585 |
| 923  | Carvone oxide                    | 1805 |
| 924* | Sabinyl acetate                  | 1651 |
| 925^ | 4-Ethylguaiaicol                 | 2011 |
| 926^ | Neryl formate                    | 1663 |
| 927* | Safrole                          | 1876 |
| 928^ | 10-Undecen-1-al                  | 1642 |
| 929^ | Octyl propionate                 | 1536 |
| 930^ | Dihydrosafrole                   | 1822 |
| 931* | n-Nonanoic acid                  | 2110 |
| 932* | n-Nonyl acetate                  | 1560 |
| 933^ | Isopropyl salicylate             | 1773 |
| 934^ | Butyl disulfide                  | 1580 |
| 935  | 2-Nonyl-1-al dimethyl acetal     | 1666 |
| 936  | n-Propyl phenylacetate           | 1848 |
| 937^ | Anisyl formate                   | 1710 |
| 938^ | alpha-Methylcinnamic aldehyde    | 1992 |
| 939* | Cyclohexyl n-valerate            | 1595 |
| 940* | n-Hexyltiglate                   | 1621 |
| 941^ | Methyl 3-hydroxybutyrate         | 1454 |
| 942  | Methyl undecyl ether             | 1453 |
| 943^ | Dihydrocarvyl acetate            | 1670 |
| 944^ | Myrcenyl propionate              | 1625 |
| 945  | Cinnamyl formate                 | 2094 |
| 946^ | p-Isopropylacetophenone          | 1912 |
| 947^ | Terpinyl formate                 | 1666 |
| 948  | Methyllavender ketone            | 2067 |
| 949* | alpha-Methylcinnamyl alcohol     | 2252 |
| 950* | Hydroxycitronellol               | 2143 |
| 951^ | gamma-Ionone                     | 1882 |
| 952^ | 2,4-Dimethyl benzylacetate       | 1916 |
| 953* | Undecenol                        | 1889 |
| 954  | Methyl nonyl acetaldehyde        | 1621 |
| 955* | 2-Methyl-5-isopropylacetophenone | 1876 |
| 956* | Dihydrocoumarin                  | 2286 |
| 957  | Isosafrole                       | 2029 |

---

---

|      |  |      |
|------|--|------|
| 958^ | beta-Methylcinnamyl alcohol            | 2283 |
| 959  | 4-Methyl-5-thiazoleethanol acetate     | 2077 |
| 960* | Diphenyl                               | 1981 |
| 961^ | n-Pentadecanol                         | 2252 |
| 962  | Acetaldehyde citronellyl methyl acetal | 1596 |
| 963* | cis-3-Decenyl acetate                  | 1701 |
| 964  | Acetylcymene                           | 1928 |
| 965* | Isoamyl phenylethyl ether              | 1741 |
| 966* | p-Cresyl isovalerate                   | 1898 |
| 967* | Anisyl acetate                         | 2199 |
| 968* | alpha-Copaene                          | 1519 |
| 969* | Dihydrojasmane                         | 1892 |
| 970* | Cuminyl acetate                        | 1952 |
| 971* | beta-Bourbonene                        | 1546 |
| 972^ | m-Aminoacetophenone                    | 2181 |
| 973^ | Dimethyl suberate                      | 1985 |
| 974  | Isopropyl n-decanoate                  | 1615 |
| 975* | Dihydrocuminyll acetate                | 1900 |
| 976  | Linalyl n-butyrate                     | 1680 |
| 977^ | 2-Methyl-1-undecanol                   | 1875 |
| 978* | Phenylethyl butyrate                   | 1930 |
| 979* | Dimethyl phenyl ethyl carbinyl acetate | 1908 |
| 980^ | alpha-Cedrene                          | 1600 |
| 981  | Carvyl propionate                      | 1833 |
| 982  | beta-Copaene                           | 1626 |
| 983  | beta-Cedrene                           | 1633 |
| 984^ | Phenoxyethyl propionate                | 2126 |
| 985^ | 2-Decenal                              | 1842 |
| 986^ | Linalyl 2-methylbutyrate               | 1695 |
| 987* | Thujopsene                             | 1660 |
| 988^ | n-Amyl benzoate                        | 1940 |
| 989* | Di-n-heptyl ether                      | 1550 |
| 990* | alpha-Humulene                         | 1682 |
| 991^ | Citronellyl isobutyrate                | 1705 |
| 992  | Bornyl butyrate                        | 1760 |
| 993* | Dihydropseudo ionone                   | 1885 |
| 994  | p-Methoxybenzylacetone                 | 2236 |
| 995  | Alloaromadendrene                      | 1662 |
| 996^ | p-Cresyl isotiglate                    | 2011 |
| 997^ | alpha-Methylcinnamyl acetate           | 2158 |

---

---

|       |                             |        |
|-------|-----------------------------|--------|
| 998*  | Heliotropyl acetate         | 2325   |
| 999^  | Phenoxyethyl isobutyrate    | 2106   |
| 1000* | n–Undecyl acetate           | 1775   |
| 1001^ | Valencene                   | 1751   |
| 1002  | 3–Methylcoumarin            | 2424   |
| 1003  | 2–Cyclohexylcyclohexanone   | 1975   |
| 1004  | alpha–Muurolene             | 1730   |
| 1005^ | o–Methoxy cinnamic aldehyde | 2430   |
| 1006  | Methyl isovalerate          | 1734   |
| 1007* | Bornyl isovalerate          | 1774   |
| 1008^ | gamma–Cadinene              | 1766   |
| 1009^ | Calamenene                  | 1842   |
| 1010* | delta–Cadinene              | 1761   |
| 1011* | alpha–Nerolidol             | 1961   |
| 1012^ | Citronellyl isocrotonate    | 1833   |
| 1013^ | 3,4–Dimethoxyacetophenone   | 2393   |
| 1014  | Di–n–butyl succinate        | 2000   |
| 1015^ | gamma–Undecalactone         | 2210   |
| 1016  | Tangerinal                  | 1874.5 |
| 1017  | 6–Methylcoumarin            | 2630   |
| 1018* | cis–3–Hexenyl benzoate      | 2122   |
| 1019^ | 7–Methylcoumarin            | 2620   |
| 1020* | beta–Nerolidol              | 2000   |
| 1021^ | Citronellyl crotonate       | 1929   |
| 1022* | Di–n–amyl fumarate          | 2006   |
| 1023^ | n–Hexyl benzoate            | 2066   |
| 1024  | Triacetin                   | 2029   |
| 1025^ | n–Hexyl octanoate           | 1805   |
| 1026^ | Anisyl n–butyrate           | 2274   |
| 1027^ | Geranyl 2–methylbutyrate    | 1886   |
| 1028  | Neryl isovalerate           | 1864   |
| 1029* | alpha–Cedrene epoxide       | 1961   |
| 1030* | Phenylpropyl isovalerate    | 2100   |
| 1031  | n–Hexyl phenylacetate       | 2148   |
| 1032^ | cis–3–hexenyl phenylacetate | 2220   |
| 1033^ | Methyl n–tridecanoate       | 1895   |
| 1034^ | Isobutyl 10–undecenoate     | 1900   |
| 1035* | Dibenzyl ether              | 2323   |
| 1036^ | Geranyl tiglate             | 1985   |
| 1037^ | Isoamyl anthranilate        | 2447   |

---

---

|       |   |        |
|-------|---|--------|
| 1038  | Di-n-butyl adipate                                  | 2087   |
| 1039  | Furfuryl disulfide                                  | 2600   |
| 1040  | Methyl zingerone                                    | 2640   |
| 1041^ | Patchouli alcohol                                   | 2156   |
| 1042^ | Methyl N-propylanthranilate                         | 2575   |
| 1043* | Guaiacyl n-caproate                                 | 2296   |
| 1044^ | Isoamyl pyruvate                                    | 1910   |
| 1045  | Cyclohexadecanone                                   | 2392   |
| 1046  | Acetyl cedrene                                      | 2213   |
| 1047  | n-heptyl salicylate                                 | 2332   |
| 1048  | Neryl n-Heptanoate                                  | 2120   |
| 1049^ | Pentalide   | 2255   |
| 1050* | Geranyl n-heptanoate                                | 2157   |
| 1051  | Nerolidyl ethanol                                   | 2462   |
| 1052* | Octyl salicylate                                    | 2435   |
| 1053^ | Methyl hexadecanoate                                | 2204   |
| 1054^ | Phenylethyl n-decanoate                             | 2540   |
| 1055* | Methyl octadecanoate                                | 2418   |
| 1056  | Ethyl oleate  | 2489   |
| 1057^ | beta-Terpinenol                                     | 1616   |
| 1058* | delta-Terpinenol                                    | 1655   |
| 1059^ | sec-butylcyclohexanone                              | 1565   |
| 1060^ | 4-(2,4,6-Trimethyl-3-cyclohexen-1-yl)-3-buten-2-one | 1886.5 |
| 1061  | Allyl cinnamate                                     | 2156   |
| 1062* | Cedryl acetate                                      | 1882   |
| 1063  | 1,1-Dimethoxy-n-propane                             | 880    |
| 1064^ | n-Pentan-2-ol                                       | 1091   |
| 1065  | Nootketone  | 2250   |
| 1066* | Dibutyl butyrolactone                               | 2141   |
| 1067^ | 2-Propionylpyrrol                                   | 1990   |
| 1068  | Octyl-n-valerate                                    | 1719   |
| 1069^ | n-Propyl n-decanoate                                | 1697   |
| 1070  | Isoamyl n-nonanoate                                 | 1748   |
| 1071* | n-Butyl n-nonanoate                                 | 1700   |
| 1072^ | 1,1-Dimethoxy-n-butane                              | 969    |
| 1073  | 1,1-Dimethoxy-n-pentane                             | 1057   |
| 1074  | 2-n-Butylthiazole                                   | 1480   |
| 1075^ | n-gamma-heptalactone                                | 1860   |
| 1076  | n-Heptyl pivalate                                   | 1428   |
| 1077* | Isopropyl n-dodecanoate                             | 1814   |

---

---

|       |                              |      |
|-------|------------------------------|------|
| 1078  | n-Butyl n-tridecanoate       | 2118 |
| 1079  | 2,2-Dimethyl-n-pentanol      | 1405 |
| 1080^ | Isopropyl n-valerate         | 1125 |
| 1081  | Salicaldehyde                | 1705 |
| 1082  | n-Butyl propyl sulfide       | 1173 |
| 1083^ | Furfuryl n-propionate        | 1587 |
| 1084  | n-Propyl tiglate             | 1320 |
| 1085* | 2,4-Dimethyl-n-pentan-3-ol   | 1157 |
| 1086* | 2-Ethyl-2-n-hexanal          | 1336 |
| 1087  | 1,1-Dimethoxy-n-heptane      | 1261 |
| 1088^ | n-Decan-2-ol                 | 1585 |
| 1089* | D-n-hexyl ether              | 1349 |
| 1090  | n-Butyl angelate             | 1417 |
| 1091  | n-Amyl tiglate               | 1519 |
| 1092* | Benzyl n-decanoate           | 2460 |
| 1093  | n-Propyl anisate             | 2205 |
| 1094* | Anisyl n-propionate          | 2205 |
| 1095* | n-Propyl anthranilate        | 2320 |
| 1096^ | Neryl n-valerate             | 1930 |
| 1097* | Furfuryl n-heptanoate        | 1950 |
| 1098^ | Isopropyl n-nonanoate        | 1516 |
| 1099* | n-Heptyl valerate            | 1614 |
| 1100* | n-Amyl anthranilate          | 2510 |
| 1101^ | Cinnamyl n-heptanoate        | 2545 |
| 1102^ | trans-2-Hexenyl n-heptanoate | 1755 |
| 1103  | Citronellyl n-valerate       | 1880 |
| 1104^ | n-hexyl nonanoate            | 1900 |
| 1105^ | Linalyl n-heptanoate         | 1930 |
| 1106^ | Terpinyl n-valerate          | 1928 |
| 1107* | n-Heptyl 2-methylbutyrate    | 1518 |
| 1108^ | n-Amyl furfurylacrylate      | 2153 |
| 1109* | Geranyl n-valerate           | 1960 |
| 1110  | n-Amyl anisate               | 2405 |
| 1111^ | n-Butyl n-valerate           | 1305 |
| 1112^ | Isopropyl n-undecanoate      | 1715 |
| 1113* | Methyl n-dodecyl ketone      | 1893 |
| 1114  | Neryl n-hexanoate            | 2021 |
| 1115* | n-Heptyl phenylacetate       | 2265 |
| 1116  | n-Undecanal diethyl acetal   | 1712 |
| 1117^ | Methyl n-decyl ketone        | 1688 |

---

---

|                   |                                   |      |
|-------------------|-----------------------------------|------|
| 1118 <sup>^</sup> | n-Hexyl n-heptanoate              | 1699 |
| 1119 <sup>*</sup> | n-Amyl n-heptanoate               | 1600 |
| 1120 <sup>^</sup> | n-Undecanal dimethyl acetal       | 1668 |
| 1121 <sup>^</sup> | Isoamyl n-undecanoate             | 1948 |
| 1122 <sup>*</sup> | n-Amyl 10-undecenoate             | 2053 |
| 1123              | n-Butyl n-pentadecanoate          | 2330 |
| 1124 <sup>^</sup> | n-Butyl n-tetradecanoate          | 2229 |
| 1125 <sup>*</sup> | n-Propyl n-dodecanoate            | 1897 |
| 1126 <sup>^</sup> | n-Propyl n-nonanoate              | 1598 |
| 1127              | Furfuryl n-valerate               | 1752 |
| 1128 <sup>^</sup> | n-Hexyl n-valerate                | 1498 |
| 1129 <sup>*</sup> | alpha-n-Butyl cinnamic aldehyde   | 2160 |
| 1130 <sup>*</sup> | n-Dodecanal diethyl acetal        | 1811 |
| 1131 <sup>^</sup> | n-Butyl N-methylantranilate       | 2266 |
| 1132              | 2-Nonanoylfuran                   | 2163 |
| 1133 <sup>*</sup> | 2-Decanoylfuran                   | 2264 |
| 1134              | alpha-n-Heptylcinnamic aldehyde   | 2409 |
| 1135 <sup>^</sup> | 2-Ethyl-2-n-hexene-1-ol           | 1518 |
| 1136 <sup>*</sup> | Benzyl n-propyl carbinol          | 1655 |
| 1137              | Benzyl isopropyl carbinol         | 1913 |
| 1138              | n-Propyl cinnamic aldehyde        | 2111 |
| 1139              | cis-3-Hexenyl n-heptanoate        | 1743 |
| 1140 <sup>^</sup> | trans-2-Hexenyl n-valerate        | 1560 |
| 1141              | Cyclooctanyl acetate              | 1645 |
| 1142 <sup>*</sup> | cis-3-Decenol                     | 1765 |
| 1143 <sup>*</sup> | p-cresyl ethylcarbonate           | 1919 |
| 1144 <sup>^</sup> | 2-Ethyl-1-hexyl propionate        | 1450 |
| 1145 <sup>*</sup> | 2-Methyl-2-phenylhexan-4-one      | 1938 |
| 1146 <sup>^</sup> | 4-Methyl-5-hydroxy phenyl acetate | 1767 |
| 1147              | Jasmal                            | 1879 |
| 1148 <sup>^</sup> | cis-3-Octenol                     | 1563 |
| 1149              | trans-3-Octenol                   | 1541 |
| 1150 <sup>*</sup> | cis-3-Decenyl butyrate            | 1866 |
| 1151              | trans-2-Decenyl butyrate          | 1866 |
| 1152 <sup>^</sup> | n-Hexyl angelate                  | 1621 |
| 1153              | n-Propyl n-undecanoate            | 1797 |
| 1154 <sup>^</sup> | cis-3-Hexenyl n-nonanoate         | 1938 |
| 1155              | trans-2-Hexenyl n-nonanoate       | 1953 |
| 1156 <sup>^</sup> | cis-3-Hexenyl methyl acetal       | 1267 |
| 1157 <sup>*</sup> | trans-2-Heptenyl butyrate         | 1568 |

---

---

|       |  |      |
|-------|--|------|
| 1158* | trans-2-Heptenyl propionate                  | 1497 |
| 1159* | Citronellyl ethyl acetal                     | 1626 |
| 1160^ | o-Toluyl thiol                               | 1631 |
| 1161^ | Isoamyl anisate                              | 2333 |
| 1162^ | n-Amyl pivalate                              | 1228 |
| 1163^ | Isolongiforanone                             | 2092 |
| 1164* | Di-n-Propyl fumarate                         | 1806 |
| 1165* | N-Methyl-2-pyrrolaldehyde                    | 1616 |
| 1166^ | 1-Methyl 4-hepten-1-ol                       | 1433 |
| 1167^ | alpha-Methyl citronellol                     | 1540 |
| 1168^ | Phenylpropyl n-valerate                      | 2183 |
| 1169  | n-Propionyl methylantranilate                | 2350 |
| 1170  | 1,1-Diethoxyhexane                           | 1228 |
| 1171  | o-Methyl phenyl ethyl alcohol                | 2012 |
| 1172^ | omega-Decenyl butyrate                       | 1870 |
| 1173^ | omega-Methyl undecylenate                    | 1747 |
| 1174* | 4-Ethyl isohexanoate                         | 1181 |
| 1175* | Dihydrolinalyl acetaldehyde                  | 1659 |
| 1176^ | Methyl isogeranylacetate                     | 1765 |
| 1177  | Dihydrocumin alcohol                         | 1981 |
| 1178* | Methyl isohexyl carbinyl acetate             | 1300 |
| 1179  | Ethyl phenyl ethyl acetal                    | 1770 |
| 1180  | 2-Methyl-2-(4-methyl n-amyl) tetrahydrofuran | 1325 |
| 1181* | sec-Butyl ethyl formal                       | 992  |
| 1182  | 1,3-Butylene glycol formal                   | 1128 |
| 1183  | 1,4-Butylene glycol formal                   | 1176 |
| 1184* | 2,3-Butylene glycol formal                   | 1037 |

---

\* validación, ^ predicción

**Tabla 3A.** Nombres químicos, número de registro CAS e índices de retención experimentales de las 269 fragancias medidas en la columna polar DB-225MS

| No.             | Nombre  | CAS         | I      |
|-----------------|---|-------------|--------|
| 1 <sup>^</sup>  | 1,2-Cyclopentanedione, 3-methyl-                            | 000765-70-8 | 1464.7 |
| 2 <sup>^</sup>  | 1,6,10-Dodecatrien-3-ol, 3,7,11-trimethyl-                  | 007212-44-4 | 1877.2 |
| 3 <sup>*</sup>  | 1,6-Octadien-3-ol, 3,7-dimethyl-                            | 000078-70-6 | 1448.2 |
| 4 <sup>*</sup>  | 10-Undecenal  | 000112-45-8 | 1819.5 |
| 5 <sup>^</sup>  | 1-Butanol, 3-methyl-  | 000123-51-3 | 913.3  |
| 6 <sup>*</sup>  | 1-Butanol, 3-methyl-, acetate                               | 000123-92-2 | 1085.5 |
| 7 <sup>^</sup>  | 1-Butanol, 3-methyl-, formate                               | 000110-45-2 | 1045.3 |
| 8 <sup>^</sup>  | 1-Butanol, 3-methyl-, propanoate                            | 000105-68-0 | 1193.4 |
| 9 <sup>^</sup>  | 1-Cyclohexene-1-methanol, 4-(1-methylethenyl)-              | 000536-59-4 | 1750.9 |
| 10 <sup>*</sup> | 1-Decanol   | 000112-30-1 | 1551.4 |
| 11 <sup>^</sup> | 1-Hexanol   | 000111-27-3 | 1128.0 |
| 12 <sup>*</sup> | 1-Hexanol, 2-ethyl-   | 000104-76-7 | 1233.3 |
| 13              | 1-Hexanol, 3,5,5-trimethyl-                                 | 003452-97-9 | 1172.5 |
| 14 <sup>^</sup> | 1-Octanol   | 000111-87-5 | 1350.0 |
| 15 <sup>^</sup> | 1-Octen-3-ol  | 003391-86-4 | 1636.0 |
| 16              | 2(3H)-Furanone, 5-butyldihydro-                             | 000104-50-7 | 1801.6 |
| 17              | 2(3H)-Furanone, 5-ethyldihydro-                             | 000695-06-7 | 1574.9 |
| 18 <sup>*</sup> | 2(3H)-Furanone, 5-heptyldihydro-                            | 000104-67-6 | 2101.9 |
| 19              | 2(3H)-Furanone, 5-hexyldihydro-                             | 000706-14-9 | 2002.9 |
| 20              | 2(3H)-Furanone, 5-methyl-                                   | 000591-12-8 | 1282.5 |
| 21 <sup>*</sup> | 2(3H)-Furanone, dihydro-5-methyl-                           | 000108-29-2 | 1483.0 |
| 22 <sup>*</sup> | 2(3H)-Furanone, dihydro-5-propyl-                           | 000105-21-5 | 1698.5 |
| 23              | 2(4H)-Benzofuranone, 5,6,7,7a-tetrahydro-4,4,7a-trimethyl-  | 015356-74-8 | 1805.3 |
| 24              | 2(5H)-Furanone, 5-ethyl-3-hydroxy-4-methyl-                 | 000698-10-2 | 1793.3 |
| 25 <sup>*</sup> | 2,4-Decadienal, (E,E)-                                      | 025152-84-5 | 1633.7 |
| 26 <sup>^</sup> | 2,4-Heptadienal, (E,E)-                                     | 004313-03-5 | 1315.0 |
| 27              | 2,6,6-Trimethyl-2-cyclohexene-1,4-dione                     | 001125-21-9 | 1588.5 |
| 28 <sup>*</sup> | 2,6-Octadien-1-ol, 3,7-dimethyl-, (Z)-                      | 000106-25-2 | 1435.5 |
| 29 <sup>^</sup> | 2,6-Octadien-1-ol, 3,7-dimethyl-, formate, (E)-             | 000105-86-2 | 1602.9 |
| 30 <sup>*</sup> | 2-Acetyl-5-methylfuran                                      | 001193-79-9 | 1449.2 |
| 31 <sup>^</sup> | 2-Acetylthiazole  | 024295-03-2 | 1514.9 |
| 32              | 2-Butanone, 4-(4-hydroxyphenyl)-                            | 005471-51-2 | 1958.2 |
| 33              | 2-Buten-1-ol, 3-methyl-                                     | 000556-82-1 | 960.9  |
| 34 <sup>^</sup> | 2-Buten-1-one, 1-(2,6,6-trimethyl-1-cyclohexen-1-yl)-       | 035044-68-9 | 1752.2 |
| 35              | 2-Buten-1-one, 1-(2,6,6-trimethyl-2-cyclohexen-1-yl)-, (E)- | 024720-09-0 | 1752.2 |
| 36 <sup>*</sup> | 2-Cyclopenten-1-one, 3-methyl-2-(2-pentenyl)-, (Z)-         | 000488-10-8 | 1762.4 |

|                 |  |             |        |
|-----------------|--|-------------|--------|
| 37 <sup>^</sup> | 2-Ethyl-3-methoxypyrazine                              | 025680-58-4 | 1447.8 |
| 38              | 2-Furancarboxaldehyde, 5-methyl-                       | 000620-02-0 | 1425.5 |
| 39              | 2-Furanmethanol, acetate                               | 000623-17-6 | 1583.2 |
| 40 <sup>*</sup> | 2-Furfurylthiol  | 000098-02-2 | 1331.3 |
| 41 <sup>*</sup> | 2-Heptanone  | 000110-43-0 | 1144.0 |
| 42 <sup>^</sup> | 2-Hexen-1-ol, (E)-                                     | 000928-95-0 | 1173.3 |
| 43 <sup>^</sup> | 2-Hexen-1-ol, acetate, (E)-                            | 002497-18-9 | 1346.9 |
| 44 <sup>^</sup> | 2-Hexenal, (E)-  | 006728-26-3 | 1154.3 |
| 45 <sup>*</sup> | 2H-Pyran, tetrahydro-4-methyl-2-(2-methyl-1-propenyl)- | 016409-43-1 | 1664.5 |
| 46 <sup>^</sup> | 2H-Pyran-2-one, 6-hexyltetrahydro-                     | 000710-04-3 | 2075.9 |
| 47 <sup>*</sup> | 2H-Pyran-2-one, tetrahydro-6-methyl-                   | 000823-22-3 | 1573.9 |
| 48              | 2H-Pyran-2-one, tetrahydro-6-nonyl-                    | 002721-22-4 | 2370.8 |
| 49              | 2H-Pyran-2-one, tetrahydro-6-pentyl-                   | 000705-86-2 | 1982.0 |
| 50 <sup>^</sup> | 2H-Pyran-2-one, tetrahydro-6-propyl-                   | 000698-76-0 | 1773.1 |
| 51 <sup>^</sup> | 2-Methylheptanoic acid                                 | 001188-02-9 | 1416.0 |
| 52 <sup>^</sup> | 2-Naphthyl methyl ketone                               | 000093-08-3 | 2129.4 |
| 53 <sup>^</sup> | 2-Nonanone   | 000821-55-6 | 1372.3 |
| 54 <sup>^</sup> | 2-Nonen-1-ol, (E)-                                     | 031502-14-4 | 1491.6 |
| 55 <sup>*</sup> | 2-Nonenal, (Z)-  | 060784-31-8 | 1488.5 |
| 56              | 2-Octanone   | 000111-13-7 | 1263.3 |
| 57 <sup>*</sup> | 2-Octenal, (E)-  | 002548-87-0 | 1389.1 |
| 58 <sup>*</sup> | 2-Octynoic acid, methyl ester                          | 000111-12-6 | 1572.6 |
| 59 <sup>^</sup> | 2-Propanone, 1-(4-methoxyphenyl)-                      | 000122-84-9 | 1842.9 |
| 60 <sup>^</sup> | 2-Propen-1-ol, 3-phenyl-                               | 000104-54-1 | 1839.8 |
| 61 <sup>^</sup> | 2-Propen-1-ol, 3-phenyl-, propanoate                   | 000103-56-0 | 1974.3 |
| 62 <sup>^</sup> | 2-Propenal, 3-(2-methoxyphenyl)-                       | 001504-74-1 | 1957.8 |
| 63              | 2-Tridecanone  | 000593-08-8 | 1778.2 |
| 64 <sup>^</sup> | 2-Undecanone   | 000112-12-9 | 1580.3 |
| 65 <sup>*</sup> | 3-(4-Isopropylphenyl)-2-methylpropionaldehyde          | 000103-95-7 | 1857.7 |
| 66 <sup>*</sup> | 3-(Methylthio)propanoic acid methyl ester              | 013532-18-8 | 1344.1 |
| 67              | 3,4-Hexanedione  | 004437-51-8 | 1169.5 |
| 68              | 3-Hexen-1-ol   | 000544-12-7 | 1170.6 |
| 69              | 3-Hexen-1-ol, acetate, (Z)-                            | 003681-71-8 | 1339.7 |
| 70 <sup>^</sup> | 3-Hexen-1-ol, formate, (Z)-                            | 033467-73-1 | 1311.8 |
| 71 <sup>*</sup> | 3-Hexenoic acid, (E)-                                  | 001577-18-0 | 1353.0 |
| 72 <sup>*</sup> | 3-Phenylpropanol                                       | 000122-97-4 | 1803.4 |
| 73              | 4-Heptenal, (Z)-                                       | 006728-31-0 | 1278.2 |
| 74 <sup>^</sup> | 4H-Pyran-4-one, 2-ethyl-3-hydroxy-                     | 004940-11-8 | 1683.5 |
| 75              | 4-Methylthiazole                                       | 000693-95-8 | 1269.6 |
| 76              | 5,6,7,8-Tetrahydroquinoxaline                          | 034413-35-9 | 1743.1 |
| 77 <sup>^</sup> | 5,9-Undecadien-2-one, 6,10-dimethyl-                   | 000689-67-8 | 1655.4 |
| 78              | 5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-             | 003796-70-1 | 1655.4 |

|      |  |             |        |
|------|--|-------------|--------|
| 79   | 5H-5-Methyl-6,7-dihydrocyclopentapyrazine      | 023747-48-0 | 1656.1 |
| 80*  | 5-Heptenal, 2,6-dimethyl-                      | 000106-72-9 | 1299.6 |
| 81*  | 5-Methyl-2-phenyl-2-hexenal                    | 021834-92-4 | 1987.6 |
| 82*  | 5-Thiazoleethanol, 4-methyl-                   | 000137-00-8 | 1632.4 |
| 83*  | 6-Octen-1-ol, 3,7-dimethyl-, acetate           | 000150-84-5 | 1584.3 |
| 84   | 6-Octen-1-ol, 3,7-dimethyl-, formate           | 000105-85-1 | 1568.9 |
| 85^  | 6-Octen-1-ol, 3,7-dimethyl-, propanoate        | 000141-14-0 | 1669.2 |
| 86   | 6-Octenal, 3,7-dimethyl-, (R)-                 | 002385-77-5 | 1411.3 |
| 87   | 9,12-Octadecadienoic acid (Z,Z)-, methylester  | 000112-63-0 | 2607.8 |
| 88^  | Acetic acid, 2-phenylethyl ester               | 000103-45-7 | 1831.6 |
| 89^  | Acetic acid, 4-methylphenyl ester              | 000140-39-6 | 1654.6 |
| 90^  | Acetic acid, decyl ester                       | 000112-17-4 | 1741.9 |
| 91   | Acetic acid, heptyl ester                      | 000112-06-1 | 1416.2 |
| 92^  | Acetic acid, nonyl ester                       | 000143-13-5 | 1640.5 |
| 93   | Acetic acid, octyl ester                       | 000112-14-1 | 1530.7 |
| 94   | Acetic acid, phenyl-, isopentyl ester          | 000102-19-2 | 2032.0 |
| 95   | Acetophenone                                   | 000098-86-2 | 1582.6 |
| 96*  | Acetylpyrazine                                 | 022047-25-2 | 1434.9 |
| 97^  | Allyl nonanoate                                | 007493-72-3 | 1880.6 |
| 98^  | Anisyl propionate                              | 007549-33-9 | 1969.5 |
| 99*  | Benzaldehyde                                   | 000100-52-7 | 1582.4 |
| 100* | Benzaldehyde, 2-hydroxy-                       | 000090-02-8 | 1803.3 |
| 101^ | Benzaldehyde, 4-(1-methylethyl)-               | 000122-03-2 | 1674.2 |
| 102* | Benzaldehyde, 4-ethoxy-                        | 010031-82-0 | 1715.4 |
| 103  | Benzaldehyde, 4-ethyl-                         | 004748-78-1 | 1671.7 |
| 104* | Benzene, 1,1'-[oxybis(methylene)]bis-          | 000103-50-4 | 2338.1 |
| 105^ | Benzene, 1,2-dimethoxy-4-(1-propenyl)-         | 000093-16-3 | 1875.1 |
| 106^ | Benzene, 1,3-dimethoxy-                        | 000151-10-0 | 1619.5 |
| 107* | Benzene, 1-methoxy-4-methyl-                   | 000104-93-8 | 1480.2 |
| 108* | Benzene, 2-methoxy-4-methyl-1-(1-methylethyl)- | 001076-56-8 | 1592.4 |
| 109  | Benzene, ethoxy-                               | 000103-73-1 | 1478.1 |
| 110^ | Benzeneacetaldehyde, $\alpha$ -ethylidene-     | 004411-89-6 | 1801.3 |
| 111^ | Benzeneacetic acid, 2-methylpropyl ester       | 000102-13-6 | 1948.4 |
| 112  | Benzeneacetic acid, 2-phenylethyl ester        | 000102-20-5 | 2676.1 |
| 113^ | Benzeneacetic acid, methyl ester               | 000101-41-7 | 1832.4 |
| 114^ | Benzeneacetic acid, phenylmethyl ester         | 000102-16-9 | 2593.2 |
| 115^ | Benzeneethanol, $\alpha,\alpha$ -dimethyl-     | 000100-86-7 | 1710.8 |
| 116^ | Benzenemethanol, 4-methoxy-                    | 000105-13-5 | 1737.3 |
| 117* | Benzenemethanol, 4-methoxy-, acetate           | 000104-21-2 | 1887.2 |
| 118^ | Benzenemethanol, 4-methoxy-, formate           | 000122-91-8 | 1873.0 |
| 119^ | Benzenemethanol, $\alpha$ -methyl-, acetate    | 000093-92-5 | 1973.5 |
| 120  | Benzenepropanal                                | 000104-53-0 | 1789.1 |

|                  |   |                            |        |
|------------------|---|----------------------------|--------|
| 121 <sup>^</sup> | Benzoic acid  | 000065-85-0                | 1799.1 |
| 122              | Benzoic acid, 2-hydroxy-, 2-methylpropyl ester  | 000087-19-4                | 2000.0 |
| 123              | Benzoic acid, 2-hydroxy-, ethyl ester   | 000118-61-6                | 1901.8 |
| 124              | Benzoic acid, 2-hydroxy-, phenylmethyl ester  | 000118-58-1                | 2650.1 |
| 125              | Benzoic acid, ethyl ester   | 000093-89-0                | 1716.9 |
| 126 <sup>^</sup> | Benzophenone  | 000119-61-9                | 2311.4 |
| 127              | Benzyl alcohol  | 000100-51-6                | 1631.6 |
| 128 <sup>*</sup> | Benzyl Benzoate   | 000120-51-4                | 2475.3 |
| 129              | Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)-<br>Isobornyl acetate                 | 005655-61-8<br>000125-12-2 | 1993.9 |
| 130 <sup>*</sup> | Butanal, 3-methyl-  | 000590-86-3                | 912.8  |
| 131              | Butanedioic acid, dimethyl ester  | 000106-65-0                | 1427.4 |
| 132 <sup>*</sup> | Butanoic acid   | 000107-92-6                | 1084.4 |
| 133 <sup>*</sup> | Butanoic acid, 2-methyl-, ethyl ester   | 007452-79-1                | 1069.4 |
| 134 <sup>*</sup> | Butanoic acid, 2-methylpropyl ester   | 000539-90-2                | 1211.6 |
| 135 <sup>*</sup> | Butanoic acid, 3,7-dimethyl-2,6-octadienyl ester, (E)-  | 000106-29-6                | 1817.0 |
| 136              | Butanoic acid, 3-methyl-  | 000503-74-2                | 1116.5 |
| 137 <sup>*</sup> | Butanoic acid, 3-methyl-, 2-phenylethyl ester   | 000140-26-1                | 2036.0 |
| 138              | Butanoic acid, 3-methyl-, 3-methylbutyl ester   | 000659-70-1                | 1337.1 |
| 139 <sup>*</sup> | Butanoic acid, 3-methyl-, 3-phenyl-2-propenyl ester   | 000140-27-2                | 2175.8 |
| 140 <sup>*</sup> | Butanoic acid, 3-methyl-, butyl ester   | 000109-19-3                | 1322.9 |
| 141 <sup>^</sup> | Butanoic acid, 3-methyl-, ethyl ester   | 000108-64-5                | 1093.0 |
| 142              | Butanoic acid, 3-methyl-, phenylmethyl ester  | 000103-38-8                | 1950.8 |
| 143 <sup>^</sup> | Butanoic acid, 3-methylbutyl ester  | 000106-27-4                | 1310.7 |
| 144              | Butanoic acid, butyl ester  | 000109-21-7                | 1295.3 |
| 145              | Butanoic acid, ethyl ester  | 000105-54-4                | 1048.9 |
| 146 <sup>*</sup> | Butanoic acid, phenylmethyl ester   | 000103-37-7                | 1936.9 |
| 147              | Butanoic acid, propyl ester   | 000105-66-8                | 1182.7 |
| 148              | Caryophyllene   | 000087-44-5                | 1979.9 |
| 149              | Cedrol  | 000077-53-2                | 2047.9 |
| 150 <sup>*</sup> | Cinnamaldehyde, (E)-  | 014371-10-9                | 1812.5 |
| 151 <sup>*</sup> | Cinnamaldehyde, $\alpha$ -pentyl-   | 000122-40-7                | 2179.8 |
| 152              | Cinnamyl cinnamate  | 000122-69-0                | 2923.9 |
| 153 <sup>^</sup> | Citronellol   | 000106-22-9                | 1401.4 |
| 154              | Citronellyl butyrate  | 000141-16-2                | 1783.5 |
| 155 <sup>*</sup> | Creosol   | 000093-51-6                | 1667.8 |
| 156 <sup>^</sup> | Cyclohexane, 1-ethenyl-1-methyl-2, 4-bis (1-methylethenyl)-,[1S-(1 $\alpha$ ,2 $\beta$ ,4 $\beta$ )]- | 000515-13-9                | 1933.6 |
| 157 <sup>*</sup> | Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 $\alpha$ ,2 $\alpha$ ,5 $\alpha$ )-                     | 000491-02-1<br>015356-70-4 | 1729.4 |
| 158 <sup>*</sup> | Cyclohexanone, 2-(1-mercapto-1-methylethyl)-5-methyl-   | 038462-22-5                | 1592.6 |
| 159              | Cyclohexanone, 5-methyl-2-(1-methylethyl)-, (2R-  | 001196-31-2                | 1501.1 |

|      |  |             |        |
|------|--|-------------|--------|
|      | cis)-  |             |        |
| 160* | Cyclohexene, 1-methyl-4-(1-methylethylidene)-          | 000586-62-9 | 1392.5 |
| 161  | Cyclopentaneacetic acid, 3-oxo-2-pentyl-, methyl ester | 024851-98-7 | 2201.3 |
| 162  | D-Carvone  | 002244-16-8 | 1672.3 |
| 163  | Decanal  | 000112-31-2 | 1566.6 |
| 164  | Decanedioic acid, diethyl ester                        | 000110-40-7 | 2166.7 |
| 165* | Decanoic acid, ethyl ester                             | 000110-38-3 | 1736.0 |
| 166  | Diphenyl ether   | 000101-84-8 | 2153.7 |
| 167^ | D-Limonene   | 005989-27-5 | 1508.7 |
| 168^ | Dodecanal  | 000112-54-9 | 1771.0 |
| 169* | Dodecanoic acid, ethyl ester                           | 000106-33-2 | 1941.1 |
| 170* | Ethane, 1,1-diethoxy-                                  | 000105-57-7 | 808.3  |
| 171^ | Ethanone, 1-(2,4-dimethylphenyl)-                      | 000089-74-7 | 1601.0 |
| 172  | Ethanone, 1-(2-pyridinyl)-                             | 001122-62-9 | 1516.8 |
| 173^ | Ethanone, 1-(3-pyridinyl)-                             | 000350-03-8 | 1516.9 |
| 174* | Ethanone, 1-(4-methylphenyl)-                          | 000122-00-9 | 1592.0 |
| 175* | Ethyl Oleate   | 000111-62-6 | 2574.7 |
| 176* | Ethyl Vanillin   | 000121-32-4 | 1895.5 |
| 177  | Eucalyptol   | 000470-82-6 | 1513.9 |
| 178^ | Eugenol  | 000097-53-0 | 2004.7 |
| 179* | Formic acid, pentyl ester                              | 000638-49-3 | 1145.2 |
| 180^ | Formic acid, phenylmethyl ester                        | 000104-57-4 | 1731.9 |
| 181  | Furan, 2-[(methylthio)methyl]-                         | 001438-91-1 | 1558.6 |
| 182^ | Furan, 2-methyl-5-(methylthio)-                        | 013678-59-6 | 1438.6 |
| 183* | Heptanoic acid   | 000111-14-8 | 1418.3 |
| 184^ | Heptanoic acid, 3-methylbutyl ester                    | 000109-25-1 | 1635.2 |
| 185^ | Heptanoic acid, ethyl ester                            | 000106-30-9 | 1411.3 |
| 186* | Hexadecanoic acid, ethyl ester                         | 000628-97-7 | 2344.8 |
| 187* | Hexanal  | 000066-25-1 | 1121.3 |
| 188^ | Hexanoic acid, 2-methyl-                               | 004536-23-6 | 1310.1 |
| 189  | Hexanoic acid, 2-propenyl ester                        | 000123-68-2 | 1555.0 |
| 190^ | Hexanoic acid, ethyl ester                             | 000123-66-0 | 1293.9 |
| 191* | Hexanoic acid, pentyl ester                            | 000540-07-8 | 1625.3 |
| 192  | Humulene   | 006753-98-6 | 1830.2 |
| 193* | Hydrocoumarin  | 000119-84-6 | 1942.5 |
| 194^ | Isoamyl cinnamate                                      | 007779-65-9 | 2154.9 |
| 195^ | Isobutyl acetate                                       | 000110-19-0 | 969.1  |
| 196^ | Isopentyl hexanoate                                    | 002198-61-0 | 1529.5 |
| 197* | Isophorone   | 000078-59-1 | 1464.4 |
| 198  | Isopropyl myristate                                    | 000110-27-0 | 2398.4 |
| 199* | Menthyl isovalerate                                    | 016409-46-4 | 2088.5 |
| 200  | Methional  | 003268-49-3 | 1215.9 |

|      |   |             |        |
|------|---|-------------|--------|
| 201* | Methyl isovalerate  | 000556-24-1 | 1060.5 |
| 202* | Methyl salicylate   | 000119-36-8 | 1916.7 |
| 203* | Methyleugenol   | 000093-15-2 | 1986.4 |
| 204* | Naphthalene, 2-ethoxy-  | 000093-18-5 | 2048.2 |
| 205  | Naphthalene, 2-methoxy-                                       | 000093-04-9 | 2046.5 |
| 206^ | n-Decanoic acid   | 000334-48-5 | 1726.6 |
| 207  | Nonanal   | 000124-19-6 | 1458.9 |
| 208^ | Nonanoic acid   | 000112-05-0 | 1619.6 |
| 209  | Nonanoic acid, ethyl ester                                    | 000123-29-5 | 1627.4 |
| 210* | Nonanoic acid, methyl ester                                   | 001731-84-6 | 1612.1 |
| 211* | n-Propyl acetate  | 000109-60-4 | 922.0  |
| 212^ | Octanal, 7-hydroxy-3,7-dimethyl-                              | 000107-75-5 | 1530.4 |
| 213* | Octanoic acid   | 000124-07-2 | 1524.8 |
| 214^ | Octanoic acid, 3-methylbutyl ester                            | 002035-99-6 | 1742.4 |
| 215^ | Octanoic acid, ethyl ester                                    | 000106-32-1 | 1524.9 |
| 216* | Octanoic acid, methyl ester                                   | 000111-11-5 | 1508.1 |
| 217  | Oxacycloheptadec-8-en-2-one, (8Z)                             | 000123-69-3 | 2475.0 |
| 218* | Oxacyclohexadecan-2-one                                       | 000106-02-5 | 2339.3 |
| 219  | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester       | 000077-83-8 |        |
|      | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester, cis- | 019464-95-0 | 2286.1 |
| 220^ | p-Cymene  | 000099-87-6 | 1437.5 |
| 221^ | Pentanoic acid, 3-methyl-                                     | 000105-43-1 | 1212.9 |
| 222^ | Pentanoic acid, butyl ester                                   | 000591-68-4 | 1409.3 |
| 223^ | Pentanoic acid, ethyl ester                                   | 000539-82-2 | 1181.1 |
| 224  | Pentyl octanoate  | 000638-25-5 | 1841.7 |
| 225^ | Phenol, 2-(1-methylethyl)-                                    | 000088-69-7 | 1606.2 |
| 226  | Phenol, 2-methoxy-  | 000090-05-1 | 1672.5 |
| 227  | Phenol, 2-methyl-5-(1-methylethyl)-                           | 000499-75-2 | 1602.5 |
| 228^ | Phenylethyl Alcohol   | 000060-12-8 | 1708.9 |
| 229* | Piperonal   | 000120-57-0 | 1867.3 |
| 230* | Propanedioic acid, diethyl ester                              | 000105-53-3 | 1382.7 |
| 231* | Propanethioic acid, S-(2-furanylmethyl) ester                 | 059020-85-8 | 1776.4 |
| 232  | Propanoic acid, 2-methyl-                                     | 000079-31-2 | 976.3  |
| 233* | Propanoic acid, 2-methyl-, 2-phenylethyl ester                | 000103-48-0 | 1924.3 |
| 234^ | Propanoic acid, 2-methyl-, 3-phenyl-2-propenyl ester          | 000103-59-3 | 2056.2 |
| 235* | Propanoic acid, 2-methyl-, 4-formyl-2-methoxyphenyl ester     | 020665-85-4 | 2126.6 |
| 236  | Propanoic acid, 2-methyl-, ethyl ester                        | 000097-62-1 | 939.6  |
| 237^ | Propanoic acid, 2-methyl-, octyl ester                        | 000109-15-9 | 1628.8 |
| 238^ | Propanoic acid, 2-methyl-, phenylmethyl ester                 | 000103-28-6 | 1831.0 |
| 239  | Pyrazine, 2,3-diethyl-5-methyl-                               | 018138-04-0 | 1424.2 |

|      |   |             |        |
|------|---|-------------|--------|
| 240* | Pyrazine, 2,3-dimethyl-                 | 005910-89-4 | 1215.6 |
| 241^ | Pyrazine, 2-ethyl-3,5-dimethyl-         | 013925-07-0 | 1329.0 |
| 242^ | Pyrazine, 2-ethyl-3-methyl-             | 015707-23-0 | 1319.4 |
| 243* | Pyrazine, 2-methoxy-3-(2-methylpropyl)- | 024683-00-9 | 1596.1 |
| 244  | Pyrazine, 2-methyl-3-(methylthio)-      | 002882-20-4 | 1484.4 |
| 245* | Pyrazine, 2-methyl-6-(methylthio)-      | 002884-13-1 | 1485.9 |
| 246* | Pyrazine, 3-ethyl-2,5-dimethyl-         | 013360-65-1 | 1334.7 |
| 247  | Pyrazine, ethyl-                        | 013925-00-3 | 1291.5 |
| 248  | Pyrazine, tetramethyl-                  | 001124-11-4 | 1247.2 |
| 249  | Pyrazine, trimethyl-                    | 014667-55-1 | 1237.0 |
| 250^ | Tetradecanoic acid, ethyl ester         | 000124-06-1 | 2147.1 |
| 251  | Thiazole, 5-ethenyl-4-methyl-           | 001759-28-0 | 1498.4 |
| 252  | trans-Isoeugenol                        | 005932-68-3 | 1891.2 |
| 253  | Triacetin                               | 000102-76-1 | 1999.2 |
| 254  | Triethyl citrate                        | 000077-93-0 | 2219.5 |
| 255* | Undecanal                               | 000112-44-7 | 1672.8 |
| 256^ | $\alpha$ -Ionone                        | 000127-41-3 | 1753.5 |
| 257* | $\alpha$ -Irone                         | 000079-69-6 | 1754.8 |
| 258  | $\alpha$ -Phellandrene                  | 000099-83-2 | 1404.1 |
| 259^ | $\beta$ -Myrcene                        | 000123-35-3 | 1435.4 |
| 260^ | $\beta$ -Phenylethyl butyrate           | 000103-52-6 | 2025.2 |
| 261  | $\beta$ -Pinene                         | 000127-91-3 | 1598.0 |
| 262* | $\delta$ -Nonalactone                   | 003301-94-8 | 1882.5 |
| 263  | 2-Propenoic acid, ethyl ester           | 000140-88-5 | 1043.8 |
| 264* | Acetoin                                 | 000513-86-0 | 1210.4 |
| 265* | Pentanal                                | 000110-62-3 | 1002.3 |
| 266  | Propanoic acid                          | 000079-09-4 | 939.8  |
| 267* | 1-Propanol, 2-methyl-                   | 000078-83-1 | 774.7  |
| 268  | Ethyl Acetate                           | 000141-78-6 | 756.2  |
| 269* | Isopropyl acetate                       | 000108-21-4 | 1071.2 |

\* validación, ^ predicción

**Tabla 4A.** Nombres químicos, número de registro CAS e índices de retención experimentales de las 266 fragancias medidas en la columna ligeramente polar HP5–MS

| No. | Nombre  | CAS         | I      |
|-----|---|-------------|--------|
| 1*  | 1,2–Cyclopentanedione, 3–methyl–                            | 000765–70–8 | 1027   |
| 2*  | 1,6,10–Dodecatrien–3–ol, 3,7,11–trimethyl–                  | 007212–44–4 | 1534   |
| 3^  | 1,6–Octadien–3–ol, 3,7–dimethyl–                            | 000078–70–6 | 1099   |
| 4^  | 10–Undecenal  | 000112–45–8 | 1299   |
| 5^  | 1–Butanol, 3–methyl–  | 000123–51–3 | 734    |
| 6   | 1–Butanol, 3–methyl–, acetate                               | 000123–92–2 | 876    |
| 7^  | 1–Butanol, 3–methyl–, formate                               | 000110–45–2 | 792    |
| 8^  | 1–Butanol, 3–methyl–, propanoate                            | 000105–68–0 | 968    |
| 9   | 1–Cyclohexene–1–methanol, 4–(1–methylethenyl)–              | 000536–59–4 | 1301   |
| 10^ | 1–Decanol   | 000112–30–1 | 1271   |
| 11* | 1–Hexanol   | 000111–27–3 | 867    |
| 12* | 1–Hexanol, 2–ethyl–   | 000104–76–7 | 1028   |
| 13  | 1–Hexanol, 3,5,5–trimethyl–                                 | 003452–97–9 | 1047   |
| 14^ | 1–Octanol   | 000111–87–5 | 1069.3 |
| 15^ | 1–Octen–3–ol  | 003391–86–4 | 978    |
| 16* | 2(3H)–Furanone, 5–butyldihydro–                             | 000104–50–7 | 1259   |
| 17  | 2(3H)–Furanone, 5–ethylidihydro–                            | 000695–06–7 | 1054.5 |
| 18* | 2(3H)–Furanone, 5–heptyldihydro–                            | 000104–67–6 | 1576   |
| 19  | 2(3H)–Furanone, 5–hexyldihydro–                             | 000706–14–9 | 1471   |
| 20  | 2(3H)–Furanone, 5–methyl–                                   | 000591–12–8 | 869    |
| 21* | 2(3H)–Furanone, dihydro–5–methyl–                           | 000108–29–2 | 953    |
| 22^ | 2(3H)–Furanone, dihydro–5–propyl–                           | 000105–21–5 | 1154   |
| 23  | 2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–  | 015356–74–8 | 1537   |
| 24  | 2(5H)–Furanone, 5–ethyl–3–hydroxy–4–methyl–                 | 000698–10–2 | 1195   |
| 25* | 2,4–Decadienal, (E,E)–                                      | 025152–84–5 | 1316   |
| 26* | 2,4–Heptadienal, (E,E)–                                     | 004313–03–5 | 1010   |
| 27  | 2,6,6–Trimethyl–2–cyclohexene–1,4–dione                     | 001125–21–9 | 1144   |
| 28^ | 2,6–Octadien–1–ol, 3,7–dimethyl–, (Z)–                      | 000106–25–2 | 1228   |
| 29^ | 2,6–Octadien–1–ol, 3,7–dimethyl–, formate, (E)–             | 000105–86–2 | 1306   |
| 30* | 2–Acetyl–5–methylfuran                                      | 001193–79–9 | 1038   |
| 31  | 2–Acetylthiazole  | 024295–03–2 | 1019   |
| 32  | 2–Butanone, 4–(4–hydroxyphenyl)–                            | 005471–51–2 | 1553   |
| 33  | 2–Buten–1–ol, 3–methyl–                                     | 000556–82–1 | 776    |
| 34^ | 2–Buten–1–one, 1–(2,6,6–trimethyl–1–cyclohexen–1–yl)–       | 035044–68–9 | 1418   |
| 35  | 2–Buten–1–one, 1–(2,6,6–trimethyl–2–cyclohexen–1–yl)–, (E)– | 024720–09–0 | 1393   |
| 36* | 2–Cyclopenten–1–one, 3–methyl–2–(2–pentenyl)–, (Z)–         | 000488–10–8 | 1401   |

|     |  |             |      |
|-----|--|-------------|------|
| 37* | 2-Ethyl-3-methoxypyrazine                              | 025680-58-4 | 1053 |
| 38  | 2-Furancarboxaldehyde, 5-methyl-                       | 000620-02-0 | 963  |
| 39^ | 2-Furanmethanol, acetate                               | 000623-17-6 | 995  |
| 40* | 2-Furfurylthiol  | 000098-02-2 | 911  |
| 41* | 2-Heptanone  | 000110-43-0 | 891  |
| 42* | 2-Hexen-1-ol, (E)-                                     | 000928-95-0 | 865  |
| 43^ | 2-Hexen-1-ol, acetate, (E)-                            | 002497-18-9 | 1015 |
| 44^ | 2-Hexenal, (E)-  | 006728-26-3 | 852  |
| 45^ | 2H-Pyran, tetrahydro-4-methyl-2-(2-methyl-1-propenyl)- | 016409-43-1 | 1111 |
| 46^ | 2H-Pyran-2-one, 6-hexyltetrahydro-                     | 000710-04-3 | 1607 |
| 47  | 2H-Pyran-2-one, tetrahydro-6-methyl-                   | 000823-22-3 | 1095 |
| 48* | 2H-Pyran-2-one, tetrahydro-6-nonyl-                    | 002721-22-4 | 1926 |
| 49* | 2H-Pyran-2-one, tetrahydro-6-pentyl-                   | 000705-86-2 | 1499 |
| 50  | 2H-Pyran-2-one, tetrahydro-6-propyl-                   | 000698-76-0 | 1287 |
| 51^ | 2-Methylheptanoic acid                                 | 001188-02-9 | 1141 |
| 52* | 2-Naphthyl methyl ketone                               | 000093-08-3 | 1620 |
| 53* | 2-Nonanone   | 000821-55-6 | 1091 |
| 54^ | 2-Nonen-1-ol, (E)-                                     | 031502-14-4 | 1168 |
| 55  | 2-Nonenal, (Z)-  | 060784-31-8 | 1159 |
| 56  | 2-Octanone   | 000111-13-7 | 991  |
| 57* | 2-Octenal, (E)-  | 002548-87-0 | 1057 |
| 58^ | 2-Octynoic acid, methyl ester                          | 000111-12-6 | 1202 |
| 59* | 2-Propanone, 1-(4-methoxyphenyl)-                      | 000122-84-9 | 1386 |
| 60^ | 2-Propen-1-ol, 3-phenyl-                               | 000104-54-1 | 1306 |
| 61^ | 2-Propen-1-ol, 3-phenyl-, propanoate                   | 000103-56-0 | 1555 |
| 62* | 2-Propenal, 3-(2-methoxyphenyl)-                       | 001504-74-1 | 1533 |
| 63^ | 2-Tridecanone  | 000593-08-8 | 1495 |
| 64^ | 2-Undecanone   | 000112-12-9 | 1293 |
| 65  | 3-(4-Isopropylphenyl)-2-methylpropionaldehyde          | 000103-95-7 | 1464 |
| 66* | 3-(Methylthio)propanoic acid methyl ester              | 013532-18-8 | 1024 |
| 67  | 3,4-Hexanedione  | 004437-51-8 | 802  |
| 68  | 3-Hexen-1-ol   | 000544-12-7 | 855  |
| 69  | 3-Hexen-1-ol, acetate, (Z)-                            | 003681-71-8 | 1006 |
| 70^ | 3-Hexen-1-ol, formate, (Z)-                            | 033467-73-1 | 920  |
| 71* | 3-Hexenoic acid, (E)-                                  | 001577-18-0 | 1003 |
| 72^ | 3-Phenylpropanol                                       | 000122-97-4 | 1233 |
| 73  | 4-Heptenal, (Z)-                                       | 006728-31-0 | 901  |
| 74^ | 4H-Pyran-4-one, 2-ethyl-3-hydroxy-                     | 004940-11-8 | 1197 |
| 75  | 4-Methylthiazole                                       | 000693-95-8 | 817  |
| 76  | 5,6,7,8-Tetrahydroquinoxaline                          | 034413-35-9 | 1209 |
| 77  | 5,9-Undecadien-2-one, 6,10-dimethyl-                   | 000689-67-8 | 1434 |
| 78^ | 5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-             | 003796-70-1 | 1453 |

|      |  |             |        |
|------|--|-------------|--------|
| 79   | 5H-5-Methyl-6,7-dihydrocyclopentapyrazine      | 023747-48-0 | 1140   |
| 80*  | 5-Heptenal, 2,6-dimethyl-                      | 000106-72-9 | 1054   |
| 81^  | 5-Methyl-2-phenyl-2-hexenal                    | 021834-92-4 | 1493   |
| 82   | 5-Thiazoleethanol, 4-methyl-                   | 000137-00-8 | 1277   |
| 83*  | 6-Octen-1-ol, 3,7-dimethyl-, acetate           | 000150-84-5 | 1352   |
| 84   | 6-Octen-1-ol, 3,7-dimethyl-, formate           | 000105-85-1 | 1275   |
| 85*  | 6-Octen-1-ol, 3,7-dimethyl-, propanoate        | 000141-14-0 | 1448   |
| 86   | 6-Octenal, 3,7-dimethyl-, (R)-                 | 002385-77-5 | 1153   |
| 87   | 9,12-Octadecadienoic acid (Z,Z)-, methylester  | 000112-63-0 | 2094   |
| 88^  | Acetic acid, 2-phenylethyl ester               | 000103-45-7 | 1257.5 |
| 89^  | Acetic acid, 4-methylphenyl ester              | 000140-39-6 | 1170   |
| 90*  | Acetic acid, decyl ester                       | 000112-17-4 | 1409   |
| 91*  | Acetic acid, heptyl ester                      | 000112-06-1 | 1111   |
| 92^  | Acetic acid, nonyl ester                       | 000143-13-5 | 1309   |
| 93   | Acetic acid, octyl ester                       | 000112-14-1 | 1210   |
| 94   | Acetic acid, phenyl-, isopentyl ester          | 000102-19-2 | 1497   |
| 95   | Acetophenone                                   | 000098-86-2 | 1067   |
| 96*  | Acetylpyrazine                                 | 022047-25-2 | 1022   |
| 97^  | Allyl nonanoate                                | 007493-72-3 | 1377   |
| 98*  | Anisyl propionate                              | 007549-33-9 | 1514   |
| 99*  | Benzaldehyde                                   | 000100-52-7 | 961    |
| 100^ | Benzaldehyde, 2-hydroxy-                       | 000090-02-8 | 1044   |
| 101  | Benzaldehyde, 4-(1-methylethyl)-               | 000122-03-2 | 1242   |
| 102* | Benzaldehyde, 4-ethoxy-                        | 010031-82-0 | 1333   |
| 103^ | Benzaldehyde, 4-ethyl-                         | 004748-78-1 | 1179   |
| 104* | Benzene, 1,1'-[oxybis(methylene)]bis-          | 000103-50-4 | 1654   |
| 105  | Benzene, 1,2-dimethoxy-4-(1-propenyl)-         | 000093-16-3 | 1457   |
| 106* | Benzene, 1,3-dimethoxy-                        | 000151-10-0 | 1168   |
| 107^ | Benzene, 1-methoxy-4-methyl-                   | 000104-93-8 | 1021   |
| 108  | Benzene, 2-methoxy-4-methyl-1-(1-methylethyl)- | 001076-56-8 | 1235   |
| 109  | Benzene, ethoxy-                               | 000103-73-1 | 993    |
| 110^ | Benzeneacetaldehyde, $\alpha$ -ethylidene-     | 004411-89-6 | 1274   |
| 111^ | Benzeneacetic acid, 2-methylpropyl ester       | 000102-13-6 | 1392   |
| 112* | Benzeneacetic acid, 2-phenylethyl ester        | 000102-20-5 | 1922   |
| 113^ | Benzeneacetic acid, methyl ester               | 000101-41-7 | 1178   |
| 114  | Benzeneacetic acid, phenylmethyl ester         | 000102-16-9 | 1815   |
| 115* | Benzeneethanol, $\alpha,\alpha$ -dimethyl-     | 000100-86-7 | 1158   |
| 116  | Benzenemethanol, 4-methoxy-                    | 000105-13-5 | 1284   |
| 117^ | Benzenemethanol, 4-methoxy-, acetate           | 000104-21-2 | 1421   |
| 118* | Benzenemethanol, 4-methoxy-, formate           | 000122-91-8 | 1334   |
| 119^ | Benzenemethanol, $\alpha$ -methyl-, acetate    | 000093-92-5 | 1194   |
| 120  | Benzenepropanal                                | 000104-53-0 | 1163   |

|                  |   |                            |        |
|------------------|---|----------------------------|--------|
| 121 <sup>^</sup> | Benzoic acid  | 000065-85-0                | 1178   |
| 122 <sup>*</sup> | Benzoic acid, 2-hydroxy-, 2-methylpropyl ester  | 000087-19-4                | 1475   |
| 123              | Benzoic acid, 2-hydroxy-, ethyl ester   | 000118-61-6                | 1273   |
| 124 <sup>*</sup> | Benzoic acid, 2-hydroxy-, phenylmethyl ester  | 000118-58-1                | 1876   |
| 125              | Benzoic acid, ethyl ester   | 000093-89-0                | 1172   |
| 126 <sup>^</sup> | Benzophenone  | 000119-61-9                | 1635   |
| 127 <sup>^</sup> | Benzyl alcohol  | 000100-51-6                | 1034   |
| 128 <sup>^</sup> | Benzyl Benzoate   | 000120-51-4                | 1770   |
| 129              | Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)-<br>Isobornyl acetate                 | 005655-61-8<br>000125-12-2 | 1289   |
| 130 <sup>*</sup> | Butanal, 3-methyl-  | 000590-86-3                | 900    |
| 131              | Butanedioic acid, dimethyl ester  | 000106-65-0                | 1032   |
| 132 <sup>*</sup> | Butanoic acid   | 000107-92-6                | 794    |
| 133 <sup>*</sup> | Butanoic acid, 2-methyl-, ethyl ester   | 007452-79-1                | 849    |
| 134              | Butanoic acid, 2-methylpropyl ester   | 000539-90-2                | 954    |
| 135 <sup>^</sup> | Butanoic acid, 3,7-dimethyl-2,6-octadienyl ester, (E)-  | 000106-29-6                | 1561.5 |
| 136              | Butanoic acid, 3-methyl-  | 000503-74-2                | 850    |
| 137 <sup>^</sup> | Butanoic acid, 3-methyl-, 2-phenylethyl ester   | 000140-26-1                | 1494   |
| 138 <sup>*</sup> | Butanoic acid, 3-methyl-, 3-methylbutyl ester   | 000659-70-1                | 1104.5 |
| 139 <sup>*</sup> | Butanoic acid, 3-methyl-, 3-phenyl-2-propenyl ester   | 000140-27-2                | 1686   |
| 140              | Butanoic acid, 3-methyl-, butyl ester   | 000109-19-3                | 1045   |
| 141 <sup>^</sup> | Butanoic acid, 3-methyl-, ethyl ester   | 000108-64-5                | 852    |
| 142              | Butanoic acid, 3-methyl-, phenylmethyl ester  | 000103-38-8                | 1396   |
| 143 <sup>^</sup> | Butanoic acid, 3-methylbutyl ester  | 000106-27-4                | 1055   |
| 144 <sup>*</sup> | Butanoic acid, butyl ester  | 000109-21-7                | 995    |
| 145              | Butanoic acid, ethyl ester  | 000105-54-4                | 804    |
| 146 <sup>*</sup> | Butanoic acid, phenylmethyl ester   | 000103-37-7                | 1347   |
| 147 <sup>^</sup> | Butanoic acid, propyl ester   | 000105-66-8                | 899    |
| 148              | Caryophyllene   | 000087-44-5                | 1426   |
| 149              | Cedrol  | 000077-53-2                | 1611   |
| 150 <sup>*</sup> | Cinnamaldehyde, (E)-  | 014371-10-9                | 1272   |
| 151 <sup>*</sup> | Cinnamaldehyde, $\alpha$ -pentyl-   | 000122-40-7                | 1651   |
| 152              | Cinnamyl cinnamate  | 000122-69-0                | 2416   |
| 153 <sup>^</sup> | Citronellol   | 000106-22-9                | 1227   |
| 154              | Citronellyl butyrate  | 000141-16-2                | 1528   |
| 155 <sup>^</sup> | Creosol   | 000093-51-6                | 1193   |
| 156 <sup>*</sup> | Cyclohexane, 1-ethenyl-1-methyl-2, 4-bis (1-methylethenyl)-,[1S-(1 $\alpha$ ,2 $\beta$ ,4 $\beta$ )]- | 000515-13-9                | 1426   |
| 157 <sup>*</sup> | Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 $\alpha$ ,2 $\alpha$ ,5 $\alpha$ )-                     | 000491-02-1<br>015356-70-4 | 1174   |
| 158 <sup>*</sup> | Cyclohexanone, 2-(1-mercapto-1-methylethyl)-5-methyl-   | 038462-22-5                | 1367   |
| 159              | Cyclohexanone, 5-methyl-2-(1-methylethyl)-, (2R-  | 001196-31-2                | 1155   |

|      |  |             |      |
|------|--|-------------|------|
|      | cis)-  |             |      |
| 160* | Cyclohexene, 1-methyl-4-(1-methylethylidene)-          | 000586-62-9 | 1089 |
| 161* | Cyclopentaneacetic acid, 3-oxo-2-pentyl-, methyl ester | 024851-98-7 | 1657 |
| 162* | D-Carvone  | 002244-16-8 | 1246 |
| 163* | Decanal  | 000112-31-2 | 1205 |
| 164  | Decanedioic acid, diethyl ester                        | 000110-40-7 | 1787 |
| 165  | Decanoic acid, ethyl ester                             | 000110-38-3 | 1394 |
| 166  | Diphenyl ether   | 000101-84-8 | 1404 |
| 167^ | D-Limonene   | 005989-27-5 | 1029 |
| 168  | Dodecanal  | 000112-54-9 | 1408 |
| 169^ | Dodecanoic acid, ethyl ester                           | 000106-33-2 | 1593 |
| 170* | Ethane, 1,1-diethoxy-                                  | 000105-57-7 | 729  |
| 171^ | Ethanone, 1-(2,4-dimethylphenyl)-                      | 000089-74-7 | 1253 |
| 172  | Ethanone, 1-(2-pyridinyl)-                             | 001122-62-9 | 1033 |
| 173^ | Ethanone, 1-(3-pyridinyl)-                             | 000350-03-8 | 1111 |
| 174* | Ethanone, 1-(4-methylphenyl)-                          | 000122-00-9 | 1186 |
| 175* | Ethyl Oleate   | 000111-62-6 | 2167 |
| 176^ | Ethyl Vanillin   | 000121-32-4 | 1459 |
| 177  | Eucalyptol   | 000470-82-6 | 1032 |
| 178  | Eugenol  | 000097-53-0 | 1359 |
| 179* | Formic acid, pentyl ester                              | 000638-49-3 | 826  |
| 180^ | Formic acid, phenylmethyl ester                        | 000104-57-4 | 1078 |
| 181  | Furan, 2-[(methylthio)methyl]-                         | 001438-91-1 | 1001 |
| 182* | Furan, 2-methyl-5-(methylthio)-                        | 013678-59-6 | 951  |
| 183^ | Heptanoic acid   | 000111-14-8 | 1084 |
| 184* | Heptanoic acid, 3-methylbutyl ester                    | 000109-25-1 | 1347 |
| 185^ | Heptanoic acid, ethyl ester                            | 000106-30-9 | 1127 |
| 186^ | Hexadecanoic acid, ethyl ester                         | 000628-97-7 | 1993 |
| 187^ | Hexanal  | 000066-25-1 | 803  |
| 188^ | Hexanoic acid, 2-methyl-                               | 004536-23-6 | 1043 |
| 189* | Hexanoic acid, 2-propenyl ester                        | 000123-68-2 | 1079 |
| 190  | Hexanoic acid, ethyl ester                             | 000123-66-0 | 999  |
| 191^ | Hexanoic acid, pentyl ester                            | 000540-07-8 | 1287 |
| 192  | Humulene   | 006753-98-6 | 1460 |
| 193* | Hydrocoumarin  | 000119-84-6 | 1387 |
| 194^ | Isoamyl cinnamate                                      | 007779-65-9 | 1745 |
| 195^ | Isobutyl acetate                                       | 000110-19-0 | 773  |
| 196^ | Isopentyl hexanoate                                    | 002198-61-0 | 1249 |
| 197* | Isophorone   | 000078-59-1 | 1122 |
| 198  | Isopropyl myristate                                    | 000110-27-0 | 1824 |
| 199  | Menthyl isovalerate                                    | 016409-46-4 | 1518 |
| 200* | Methional  | 003268-49-3 | 907  |

|      |   |             |        |
|------|---|-------------|--------|
| 201* | Methyl isovalerate  | 000556-24-1 | 777    |
| 202* | Methyl salicylate   | 000119-36-8 | 1196   |
| 203* | Methyleugenol   | 000093-15-2 | 1404.5 |
| 204^ | Naphthalene, 2-ethoxy-  | 000093-18-5 | 1528   |
| 205  | Naphthalene, 2-methoxy-                                       | 000093-04-9 | 1454.5 |
| 206^ | n-Decanoic acid   | 000334-48-5 | 1370   |
| 207  | Nonanal   | 000124-19-6 | 1104   |
| 208^ | Nonanoic acid   | 000112-05-0 | 1274   |
| 209* | Nonanoic acid, ethyl ester                                    | 000123-29-5 | 1295   |
| 210^ | Nonanoic acid, methyl ester                                   | 001731-84-6 | 1223   |
| 211  | n-Propyl acetate  | 000109-60-4 | 715    |
| 212^ | Octanal, 7-hydroxy-3,7-dimethyl-                              | 000107-75-5 | 1288   |
| 213^ | Octanoic acid   | 000124-07-2 | 1179   |
| 214^ | Octanoic acid, 3-methylbutyl ester                            | 002035-99-6 | 1445   |
| 215^ | Octanoic acid, ethyl ester                                    | 000106-32-1 | 1196   |
| 216  | Octanoic acid, methyl ester                                   | 000111-11-5 | 1125   |
| 217  | Oxacycloheptadec-8-en-2-one, (8Z)                             | 000123-69-3 | 1937.5 |
| 218* | Oxacyclohexadecan-2-one                                       | 000106-02-5 | 1839   |
| 219  | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester       | 000077-83-8 |        |
|      | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester, cis- | 019464-95-0 | 1517   |
| 220^ | p-Cymene  | 000099-87-6 | 1025   |
| 221^ | Pentanoic acid, 3-methyl-                                     | 000105-43-1 | 944    |
| 222^ | Pentanoic acid, butyl ester                                   | 000591-68-4 | 1093   |
| 223  | Pentanoic acid, ethyl ester                                   | 000539-82-2 | 852    |
| 224* | Pentyl octanoate  | 000638-25-5 | 1484   |
| 225^ | Phenol, 2-(1-methylethyl)-                                    | 000088-69-7 | 1199   |
| 226* | Phenol, 2-methoxy-  | 000090-05-1 | 1090   |
| 227* | Phenol, 2-methyl-5-(1-methylethyl)-                           | 000499-75-2 | 1291   |
| 228* | Phenylethyl Alcohol   | 000060-12-8 | 1114   |
| 229* | Piperonal   | 000120-57-0 | 1336   |
| 230* | Propanedioic acid, diethyl ester                              | 000105-53-3 | 1070   |
| 231  | Propanethioic acid, S-(2-furanylmethyl) ester                 | 059020-85-8 | 1256   |
| 232  | Propanoic acid, 2-methyl-                                     | 000079-31-2 | 765    |
| 233* | Propanoic acid, 2-methyl-, 2-phenylethyl ester                | 000103-48-0 | 1396   |
| 234^ | Propanoic acid, 2-methyl-, 3-phenyl-2-propenyl ester          | 000103-59-3 | 1584   |
| 235* | Propanoic acid, 2-methyl-, 4-formyl-2-methoxyphenyl ester     | 020665-85-4 | 1671   |
| 236  | Propanoic acid, 2-methyl-, ethyl ester                        | 000097-62-1 | 757    |
| 237^ | Propanoic acid, 2-methyl-, octyl ester                        | 000109-15-9 | 1344   |
| 238^ | Propanoic acid, 2-methyl-, phenylmethyl ester                 | 000103-28-6 | 1298   |
| 239^ | Pyrazine, 2,3-diethyl-5-methyl-                               | 018138-04-0 | 1155   |

|      |   |             |        |
|------|---|-------------|--------|
| 240  | Pyrazine, 2,3-dimethyl-                 | 005910-89-4 | 918    |
| 241  | Pyrazine, 2-ethyl-3,5-dimethyl-         | 013925-07-0 | 1084   |
| 242  | Pyrazine, 2-ethyl-3-methyl-             | 015707-23-0 | 1003   |
| 243* | Pyrazine, 2-methoxy-3-(2-methylpropyl)- | 024683-00-9 | 1181   |
| 244^ | Pyrazine, 2-methyl-3-(methylthio)-      | 002882-20-4 | 1169   |
| 245* | Pyrazine, 2-methyl-6-(methylthio)-      | 002884-13-1 | 1187   |
| 246^ | Pyrazine, 3-ethyl-2,5-dimethyl-         | 013360-65-1 | 1078   |
| 247* | Pyrazine, ethyl-                        | 013925-00-3 | 915    |
| 248  | Pyrazine, tetramethyl-                  | 001124-11-4 | 1087   |
| 249  | Pyrazine, trimethyl-                    | 014667-55-1 | 1004   |
| 250* | Tetradecanoic acid, ethyl ester         | 000124-06-1 | 1792.3 |
| 251* | Thiazole, 5-ethenyl-4-methyl-           | 001759-28-0 | 1026   |
| 252* | trans-Isoeugenol                        | 005932-68-3 | 1451   |
| 253  | Triacetin                               | 000102-76-1 | 1352   |
| 254  | Triethyl citrate                        | 000077-93-0 | 1664   |
| 255^ | Undecanal                               | 000112-44-7 | 1306   |
| 256* | $\alpha$ -Ionone                        | 000127-41-3 | 1431   |
| 257^ | $\alpha$ -Irone                         | 000079-69-6 | 1519   |
| 258  | $\alpha$ -Phellandrene                  | 000099-83-2 | 1006   |
| 259^ | $\beta$ -Myrcene                        | 000123-35-3 | 991    |
| 260^ | $\beta$ -Phenylethyl butyrate           | 000103-52-6 | 1443   |
| 261  | $\beta$ -Pinene                         | 000127-91-3 | 978    |
| 262* | $\delta$ -Nonalactone                   | 003301-94-8 | 1394   |
| 263* | 2-Propenoic acid, ethyl ester           | 000140-88-5 | 702    |
| 264* | Acetoin                                 | 000513-86-0 | 714    |
| 265  | Pentanal                                | 000110-62-3 | 704    |
| 266  | Propanoic acid                          | 000079-09-4 | 706    |

\* validación, ^ predicción

**Tabla 5A.** Nombres químicos, número de registro CAS e índices de retención experimentales de las 262 fragancias medidas en la columna apolar HP-1

| No.             | Nombre  | CAS         | I      |
|-----------------|---|-------------|--------|
| 1 <sup>^</sup>  | 1,2-Cyclopentanedione, 3-methyl-                            | 000765-70-8 | 982.5  |
| 2 <sup>^</sup>  | 1,6,10-Dodecatrien-3-ol, 3,7,11-trimethyl-                  | 007212-44-4 | 1506.8 |
| 3 <sup>*</sup>  | 1,6-Octadien-3-ol, 3,7-dimethyl-                            | 000078-70-6 | 1099.7 |
| 4 <sup>*</sup>  | 10-Undecenal  | 000112-45-8 | 1298.7 |
| 5 <sup>*</sup>  | 1-Butanol, 3-methyl-  | 000123-51-3 | 688.7  |
| 6 <sup>^</sup>  | 1-Butanol, 3-methyl-, acetate                               | 000123-92-2 | 847.6  |
| 7 <sup>^</sup>  | 1-Butanol, 3-methyl-, formate                               | 000110-45-2 | 770.8  |
| 8 <sup>^</sup>  | 1-Butanol, 3-methyl-, propanoate                            | 000105-68-0 | 986.6  |
| 9               | 1-Cyclohexene-1-methanol, 4-(1-methylethenyl)-              | 000536-59-4 | 1248.0 |
| 10 <sup>*</sup> | 1-Decanol   | 000112-30-1 | 1216.9 |
| 11 <sup>*</sup> | 1-Hexanol   | 000111-27-3 | 843.8  |
| 12 <sup>*</sup> | 1-Hexanol, 2-ethyl-   | 000104-76-7 | 1018.7 |
| 13              | 1-Hexanol, 3,5,5-trimethyl-                                 | 003452-97-9 | 945.5  |
| 14 <sup>^</sup> | 1-Octanol   | 000111-87-5 | 1032.9 |
| 15 <sup>^</sup> | 1-Octen-3-ol  | 003391-86-4 | 1025.8 |
| 16              | 2(3H)-Furanone, 5-butyldihydro-                             | 000104-50-7 | 1124.1 |
| 17              | 2(3H)-Furanone, 5-ethyldihydro-                             | 000695-06-7 | 929.1  |
| 18 <sup>*</sup> | 2(3H)-Furanone, 5-heptyldihydro-                            | 000104-67-6 | 1393.5 |
| 19              | 2(3H)-Furanone, 5-hexyldihydro-                             | 000706-14-9 | 1311.6 |
| 20              | 2(3H)-Furanone, 5-methyl-                                   | 000591-12-8 | 799.4  |
| 21 <sup>*</sup> | 2(3H)-Furanone, dihydro-5-methyl-                           | 000108-29-2 | 800.0  |
| 22 <sup>^</sup> | 2(3H)-Furanone, dihydro-5-propyl-                           | 000105-21-5 | 1010.7 |
| 23              | 2(4H)-Benzofuranone, 5,6,7,7a-tetrahydro-4,4,7a-trimethyl-  | 015356-74-8 | 1299.0 |
| 24              | 2(5H)-Furanone, 5-ethyl-3-hydroxy-4-methyl-                 | 000698-10-2 | 1142.3 |
| 25 <sup>*</sup> | 2,4-Decadienal, (E,E)-                                      | 025152-84-5 | 1198.5 |
| 26 <sup>^</sup> | 2,4-Heptadienal, (E,E)-                                     | 004313-03-5 | 930.0  |
| 27 <sup>*</sup> | 2,6,6-Trimethyl-2-cyclohexene-1,4-dione                     | 001125-21-9 | 1115.6 |
| 28 <sup>^</sup> | 2,6-Octadien-1-ol, 3,7-dimethyl-, (Z)-                      | 000106-25-2 | 1123.2 |
| 29 <sup>^</sup> | 2,6-Octadien-1-ol, 3,7-dimethyl-, formate, (E)-             | 000105-86-2 | 1245.5 |
| 30              | 2-Acetyl-5-methylfuran                                      | 001193-79-9 | 989.3  |
| 31 <sup>*</sup> | 2-Acetylthiazole  | 024295-03-2 | 1031.0 |
| 32              | 2-Butanone, 4-(4-hydroxyphenyl)-                            | 005471-51-2 | 1338.3 |
| 33              | 2-Buten-1-ol, 3-methyl-                                     | 000556-82-1 | 700.1  |
| 34 <sup>^</sup> | 2-Buten-1-one, 1-(2,6,6-trimethyl-1-cyclohexen-1-yl)-       | 035044-68-9 | 1351.7 |
| 35              | 2-Buten-1-one, 1-(2,6,6-trimethyl-2-cyclohexen-1-yl)-, (E)- | 024720-09-0 | 1351.7 |

|     |  |             |        |
|-----|--|-------------|--------|
| 36  | 2-Cyclopenten-1-one, 3-methyl-2-(2-pentenyl)-, (Z)-    | 000488-10-8 | 1315.4 |
| 37* | 2-Ethyl-3-methoxypyrazine                              | 025680-58-4 | 1073.5 |
| 38* | 2-Furancarboxaldehyde, 5-methyl-                       | 000620-02-0 | 928.6  |
| 39* | 2-Furanmethanol, acetate                               | 000623-17-6 | 1043.7 |
| 40  | 2-Furfurylthiol  | 000098-02-2 | 842.8  |
| 41* | 2-Heptanone  | 000110-43-0 | 866.4  |
| 42^ | 2-Hexen-1-ol, (E)-                                     | 000928-95-0 | 852.0  |
| 43^ | 2-Hexen-1-ol, acetate, (E)-                            | 002497-18-9 | 1009.8 |
| 44* | 2-Hexenal, (E)-  | 006728-26-3 | 815.0  |
| 45^ | 2H-Pyran, tetrahydro-4-methyl-2-(2-methyl-1-propenyl)- | 016409-43-1 | 1125.8 |
| 46^ | 2H-Pyran-2-one, 6-hexyltetrahydro-                     | 000710-04-3 | 1400.1 |
| 47* | 2H-Pyran-2-one, tetrahydro-6-methyl-                   | 000823-22-3 | 892.4  |
| 48  | 2H-Pyran-2-one, tetrahydro-6-nonyl-                    | 002721-22-4 | 1663.1 |
| 49  | 2H-Pyran-2-one, tetrahydro-6-pentyl-                   | 000705-86-2 | 1291.6 |
| 50^ | 2H-Pyran-2-one, tetrahydro-6-propyl-                   | 000698-76-0 | 1098.9 |
| 51^ | 2-Methylheptanoic acid                                 | 001188-02-9 | 1073.7 |
| 52^ | 2-Naphthyl methyl ketone                               | 000093-08-3 | 1452.1 |
| 53^ | 2-Nonanone   | 000821-55-6 | 1058.7 |
| 54* | 2-Nonen-1-ol, (E)-                                     | 031502-14-4 | 1140.1 |
| 55  | 2-Nonenal, (Z)-  | 060784-31-8 | 1115.1 |
| 56  | 2-Octanone   | 000111-13-7 | 979.0  |
| 57  | 2-Octenal, (E)-  | 002548-87-0 | 1009.6 |
| 58^ | 2-Octynoic acid, methyl ester                          | 000111-12-6 | 1135.9 |
| 59^ | 2-Propanone, 1-(4-methoxyphenyl)-                      | 000122-84-9 | 1259.4 |
| 60* | 2-Propen-1-ol, 3-phenyl-                               | 000104-54-1 | 1239.9 |
| 61* | 2-Propen-1-ol, 3-phenyl-, propanoate                   | 000103-56-0 | 1392.0 |
| 62^ | 2-Propenal, 3-(2-methoxyphenyl)-                       | 001504-74-1 | 1324.9 |
| 63  | 2-Tridecanone  | 000593-08-8 | 1428.5 |
| 64^ | 2-Undecanone   | 000112-12-9 | 1245.2 |
| 65  | 3-(4-Isopropylphenyl)-2-methylpropionaldehyde          | 000103-95-7 | 1441.2 |
| 66* | 3-(Methylthio)propanoic acid methyl ester              | 013532-18-8 | 967.3  |
| 67* | 3,4-Hexanedione  | 004437-51-8 | 911.6  |
| 68  | 3-Hexen-1-ol   | 000544-12-7 | 852.0  |
| 69  | 3-Hexen-1-ol, acetate, (Z)-                            | 003681-71-8 | 1009.8 |
| 70^ | 3-Hexen-1-ol, formate, (Z)-                            | 033467-73-1 | 938.6  |
| 71* | 3-Hexenoic acid, (E)-                                  | 001577-18-0 | 913.4  |
| 72^ | 3-Phenylpropanol                                       | 000122-97-4 | 1232.9 |
| 73  | 4-Heptenal, (Z)-                                       | 006728-31-0 | 927.2  |
| 74* | 4H-Pyran-4-one, 2-ethyl-3-hydroxy-                     | 004940-11-8 | 1171.6 |
| 75  | 4-Methylthiazole                                       | 000693-95-8 | 863.0  |
| 76  | 5,6,7,8-Tetrahydroquinoxaline                          | 034413-35-9 | 1229.2 |

|                  |  |             |        |
|------------------|--|-------------|--------|
| 77 <sup>^</sup>  | 5,9–Undecadien–2–one, 6,10–dimethyl–           | 000689–67–8 | 1331.7 |
| 78 <sup>*</sup>  | 5,9–Undecadien–2–one, 6,10–dimethyl–, (E)–     | 003796–70–1 | 1331.7 |
| 79               | 5H–5–Methyl–6,7–dihydrocyclopentapyrazine      | 023747–48–0 | 1192.8 |
| 80 <sup>^</sup>  | 5–Heptenal, 2,6–dimethyl–                      | 000106–72–9 | 1028.8 |
| 81 <sup>^</sup>  | 5–Methyl–2–phenyl–2–hexenal                    | 021834–92–4 | 1467.5 |
| 82 <sup>*</sup>  | 5–Thiazoleethanol, 4–methyl–                   | 000137–00–8 | 1151.5 |
| 83               | 6–Octen–1–ol, 3,7–dimethyl–, acetate           | 000150–84–5 | 1305.6 |
| 84 <sup>*</sup>  | 6–Octen–1–ol, 3,7–dimethyl–, formate           | 000105–85–1 | 1245.0 |
| 85 <sup>^</sup>  | 6–Octen–1–ol, 3,7–dimethyl–, propanoate        | 000141–14–0 | 1433.6 |
| 86 <sup>*</sup>  | 6–Octenal, 3,7–dimethyl–, (R)–                 | 002385–77–5 | 1096.5 |
| 87               | 9,12–Octadecadienoic acid (Z,Z)–, methylester  | 000112–63–0 | 2075.0 |
| 88 <sup>^</sup>  | Acetic acid, 2–phenylethyl ester               | 000103–45–7 | 1281.4 |
| 89 <sup>^</sup>  | Acetic acid, 4–methylphenyl ester              | 000140–39–6 | 1172.8 |
| 90 <sup>^</sup>  | Acetic acid, decyl ester                       | 000112–17–4 | 1403.5 |
| 91 <sup>*</sup>  | Acetic acid, heptyl ester                      | 000112–06–1 | 1123.1 |
| 92 <sup>^</sup>  | Acetic acid, nonyl ester                       | 000143–13–5 | 1317.6 |
| 93               | Acetic acid, octyl ester                       | 000112–14–1 | 1210.4 |
| 94               | Acetic acid, phenyl–, isopentyl ester          | 000102–19–2 | 1513.0 |
| 95 <sup>*</sup>  | Acetophenone                                   | 000098–86–2 | 1070.5 |
| 96 <sup>^</sup>  | Acetylpyrazine                                 | 022047–25–2 | 1004.9 |
| 97 <sup>^</sup>  | Allyl nonanoate                                | 007493–72–3 | 1443.1 |
| 98 <sup>*</sup>  | Anisyl propionate                              | 007549–33–9 | 1470.0 |
| 99 <sup>^</sup>  | Benzaldehyde                                   | 000100–52–7 | 1021.1 |
| 100 <sup>*</sup> | Benzaldehyde, 2–hydroxy–                       | 000090–02–8 | 1146.2 |
| 101              | Benzaldehyde, 4–(1–methylethyl)–               | 000122–03–2 | 1219.2 |
| 102 <sup>*</sup> | Benzaldehyde, 4–ethoxy–                        | 010031–82–0 | 1220.3 |
| 103 <sup>^</sup> | Benzaldehyde, 4–ethyl–                         | 004748–78–1 | 1179.6 |
| 104 <sup>*</sup> | Benzene, 1,1'–[oxybis(methylene)]bis–          | 000103–50–4 | 1650.4 |
| 105              | Benzene, 1,2–dimethoxy–4–(1–propenyl)–         | 000093–16–3 | 1344.7 |
| 106 <sup>*</sup> | Benzene, 1,3–dimethoxy–                        | 000151–10–0 | 1085.6 |
| 107 <sup>^</sup> | Benzene, 1–methoxy–4–methyl–                   | 000104–93–8 | 1010.1 |
| 108 <sup>*</sup> | Benzene, 2–methoxy–4–methyl–1–(1–methylethyl)– | 001076–56–8 | 1223.7 |
| 109              | Benzene, ethoxy–                               | 000103–73–1 | 1044.0 |
| 110 <sup>^</sup> | Benzeneacetaldehyde, $\alpha$ –ethylidene–     | 004411–89–6 | 1253.0 |
| 111 <sup>*</sup> | Benzeneacetic acid, 2–methylpropyl ester       | 000102–13–6 | 1442.1 |
| 112              | Benzeneacetic acid, 2–phenylethyl ester        | 000102–20–5 | 1891.9 |
| 113 <sup>^</sup> | Benzeneacetic acid, methyl ester               | 000101–41–7 | 1207.4 |
| 114 <sup>*</sup> | Benzeneacetic acid, phenylmethyl ester         | 000102–16–9 | 1821.7 |
| 115 <sup>*</sup> | Benzeneethanol, $\alpha,\alpha$ –dimethyl–     | 000100–86–7 | 1185.1 |
| 116 <sup>*</sup> | Benzenemethanol, 4–methoxy–                    | 000105–13–5 | 1180.7 |
| 117 <sup>^</sup> | Benzenemethanol, 4–methoxy–, acetate           | 000104–21–2 | 1341.7 |
| 118 <sup>^</sup> | Benzenemethanol, 4–methoxy–, formate           | 000122–91–8 | 1273.3 |

|      |  |                            |        |
|------|--|----------------------------|--------|
| 119* | Benzenemethanol, $\alpha$ -methyl-, acetate  | 000093-92-5                | 1268.9 |
| 120  | Benzenepropanal  | 000104-53-0                | 1194.8 |
| 121^ | Benzoic acid   | 000065-85-0                | 1145.2 |
| 122^ | Benzoic acid, 2-hydroxy-, 2-methylpropyl ester   | 000087-19-4                | 1465.2 |
| 123^ | Benzoic acid, 2-hydroxy-, ethyl ester  | 000118-61-6                | 1318.5 |
| 124  | Benzoic acid, 2-hydroxy-, phenylmethyl ester   | 000118-58-1                | 1850.2 |
| 125  | Benzoic acid, ethyl ester  | 000093-89-0                | 1220.3 |
| 126^ | Benzophenone   | 000119-61-9                | 1598.8 |
| 127  | Benzyl alcohol   | 000100-51-6                | 1076.5 |
| 128^ | Benzyl Benzoate  | 000120-51-4                | 1757.7 |
| 129  | Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)-<br>Isobornyl acetate                  | 005655-61-8<br>000125-12-2 | 1383.0 |
| 130^ | Butanal, 3-methyl-   | 000590-86-3                | 651.8  |
| 131  | Butanedioic acid, dimethyl ester   | 000106-65-0                | 929.2  |
| 132* | Butanoic acid  | 000107-92-6                | 703.4  |
| 133^ | Butanoic acid, 2-methyl-, ethyl ester  | 007452-79-1                | 912.7  |
| 134* | Butanoic acid, 2-methylpropyl ester  | 000539-90-2                | 982.8  |
| 135* | Butanoic acid, 3,7-dimethyl-2,6-octadienyl ester, (E)-   | 000106-29-6                | 1503.0 |
| 136  | Butanoic acid, 3-methyl-   | 000503-74-2                | 748.0  |
| 137^ | Butanoic acid, 3-methyl-, 2-phenylethyl ester  | 000140-26-1                | 1512.2 |
| 138* | Butanoic acid, 3-methyl-, 3-methylbutyl ester  | 000659-70-1                | 1107.2 |
| 139^ | Butanoic acid, 3-methyl-, 3-phenyl-2-propenyl ester  | 000140-27-2                | 1620.6 |
| 140^ | Butanoic acid, 3-methyl-, butyl ester  | 000109-19-3                | 1061.9 |
| 141* | Butanoic acid, 3-methyl-, ethyl ester  | 000108-64-5                | 850.9  |
| 142  | Butanoic acid, 3-methyl-, phenylmethyl ester   | 000103-38-8                | 1439.9 |
| 143  | Butanoic acid, 3-methylbutyl ester   | 000106-27-4                | 1063.6 |
| 144  | Butanoic acid, butyl ester   | 000109-21-7                | 1015.5 |
| 145  | Butanoic acid, ethyl ester   | 000105-54-4                | 794.9  |
| 146  | Butanoic acid, phenylmethyl ester  | 000103-37-7                | 1402.4 |
| 147  | Butanoic acid, propyl ester  | 000105-66-8                | 927.7  |
| 148  | Caryophyllene  | 000087-44-5                | 1485.1 |
| 149  | Cedrol   | 000077-53-2                | 1593.0 |
| 150^ | Cinnamaldehyde, (E)-   | 014371-10-9                | 1194.8 |
| 151* | Cinnamaldehyde, $\alpha$ -pentyl-  | 000122-40-7                | 1604.0 |
| 152  | Cinnamyl cinnamate   | 000122-69-0                | 2108.8 |
| 153^ | Citronellol  | 000106-22-9                | 1119.4 |
| 154  | Citronellyl butyrate   | 000141-16-2                | 1502.5 |
| 155^ | Creosol  | 000093-51-6                | 1143.7 |
| 156^ | Cyclohexane, 1-ethenyl-1-methyl-2, 4-bis (1-methylethenyl)-, [1S-(1 $\alpha$ ,2 $\beta$ ,4 $\beta$ )]- | 000515-13-9                | 1434.7 |
| 157* | Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 $\alpha$ ,2 $\alpha$ ,5 $\alpha$ )-                      | 000491-02-1<br>015356-70-4 | 1170.0 |

|      |  |             |        |
|------|--|-------------|--------|
| 158* | Cyclohexanone, 2-(1-mercapto-1-methylethyl)-5-methyl-  | 038462-22-5 | 1166.4 |
| 159  | Cyclohexanone, 5-methyl-2-(1-methylethyl)-, (2R-cis)-  | 001196-31-2 | 1146.7 |
| 160  | Cyclohexene, 1-methyl-4-(1-methylethylidene)-          | 000586-62-9 | 1089.8 |
| 161  | Cyclopentaneacetic acid, 3-oxo-2-pentyl-, methyl ester | 024851-98-7 | 1642.3 |
| 162* | D-Carvone  | 002244-16-8 | 1183.2 |
| 163  | Decanal  | 000112-31-2 | 1195.7 |
| 164  | Decanedioic acid, diethyl ester                        | 000110-40-7 | 1745.5 |
| 165* | Decanoic acid, ethyl ester                             | 000110-38-3 | 1411.3 |
| 166  | Diphenyl ether   | 000101-84-8 | 1484.0 |
| 167* | D-Limonene   | 005989-27-5 | 1089.8 |
| 168* | Dodecanal  | 000112-54-9 | 1379.8 |
| 169* | Dodecanoic acid, ethyl ester                           | 000106-33-2 | 1599.9 |
| 170* | Ethane, 1,1-diethoxy-                                  | 000105-57-7 | 717.9  |
| 171^ | Ethanone, 1-(2,4-dimethylphenyl)-                      | 000089-74-7 | 1146.7 |
| 172  | Ethanone, 1-(2-pyridinyl)-                             | 001122-62-9 | 1051.1 |
| 173* | Ethanone, 1-(3-pyridinyl)-                             | 000350-03-8 | 1041.6 |
| 174^ | Ethanone, 1-(4-methylphenyl)-                          | 000122-00-9 | 1108.8 |
| 175* | Ethyl Oleate   | 000111-62-6 | 2150.9 |
| 176  | Ethyl Vanillin   | 000121-32-4 | 1346.9 |
| 177  | Eucalyptol   | 000470-82-6 | 1158.9 |
| 178* | Eugenol  | 000097-53-0 | 1328.8 |
| 179* | Formic acid, pentyl ester                              | 000638-49-3 | 838.4  |
| 180^ | Formic acid, phenylmethyl ester                        | 000104-57-4 | 1143.4 |
| 181* | Furan, 2-[(methylthio)methyl]-                         | 001438-91-1 | 1028.0 |
| 182^ | Furan, 2-methyl-5-(methylthio)-                        | 013678-59-6 | 1017.1 |
| 183* | Heptanoic acid   | 000111-14-8 | 1013.5 |
| 184^ | Heptanoic acid, 3-methylbutyl ester                    | 000109-25-1 | 1374.0 |
| 185* | Heptanoic acid, ethyl ester                            | 000106-30-9 | 1132.7 |
| 186* | Hexadecanoic acid, ethyl ester                         | 000628-97-7 | 1968.9 |
| 187^ | Hexanal  | 000066-25-1 | 811.7  |
| 188* | Hexanoic acid, 2-methyl-                               | 004536-23-6 | 998.2  |
| 189* | Hexanoic acid, 2-propenyl ester                        | 000123-68-2 | 1139.0 |
| 190^ | Hexanoic acid, ethyl ester                             | 000123-66-0 | 1015.5 |
| 191* | Hexanoic acid, pentyl ester                            | 000540-07-8 | 1331.7 |
| 192  | Humulene   | 006753-98-6 | 1453.6 |
| 193* | Hydrocoumarin  | 000119-84-6 | 1278.9 |
| 194  | Isoamyl cinnamate                                      | 007779-65-9 | 1627.9 |
| 195^ | Isobutyl acetate                                       | 000110-19-0 | 758.0  |
| 196^ | Isopentyl hexanoate                                    | 002198-61-0 | 1262.9 |
| 197  | Isophorone   | 000078-59-1 | 994.7  |

|      |   |             |        |
|------|---|-------------|--------|
| 198  | Isopropyl myristate   | 000110-27-0 | 1831.3 |
| 199* | Menthyl isovalerate   | 016409-46-4 | 1581.6 |
| 200  | Methional   | 003268-49-3 | 842.8  |
| 201* | Methyl isovalerate  | 000556-24-1 | 755.7  |
| 202* | Methyl salicylate   | 000119-36-8 | 1245.1 |
| 203* | Methyleugenol   | 000093-15-2 | 1344.7 |
| 204* | Naphthalene, 2-ethoxy-  | 000093-18-5 | 1450.1 |
| 205  | Naphthalene, 2-methoxy-                                       | 000093-04-9 | 1373.2 |
| 206^ | n-Decanoic acid   | 000334-48-5 | 1273.7 |
| 207  | Nonanal   | 000124-19-6 | 1113.4 |
| 208^ | Nonanoic acid   | 000112-05-0 | 1195.1 |
| 209  | Nonanoic acid, ethyl ester                                    | 000123-29-5 | 1326.9 |
| 210* | Nonanoic acid, methyl ester                                   | 001731-84-6 | 1244.1 |
| 211  | n-Propyl acetate  | 000109-60-4 | 686.7  |
| 212^ | Octanal, 7-hydroxy-3,7-dimethyl-                              | 000107-75-5 | 1162.1 |
| 213^ | Octanoic acid   | 000124-07-2 | 1091.0 |
| 214^ | Octanoic acid, 3-methylbutyl ester                            | 002035-99-6 | 1454.3 |
| 215^ | Octanoic acid, ethyl ester                                    | 000106-32-1 | 1217.9 |
| 216  | Octanoic acid, methyl ester                                   | 000111-11-5 | 1133.7 |
| 217  | Oxacycloheptadec-8-en-2-one, (8Z)                             | 000123-69-3 | 1867.9 |
| 218* | Oxacyclohexadecan-2-one                                       | 000106-02-5 | 1768.5 |
| 219  | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester       | 000077-83-8 |        |
|      | Oxiranecarboxylic acid, 3-methyl-3-phenyl-, ethyl ester, cis- | 019464-95-0 | 1593.1 |
| 220* | p-Cymene  | 000099-87-6 | 1089.8 |
| 221^ | Pentanoic acid, 3-methyl-                                     | 000105-43-1 | 879.1  |
| 222^ | Pentanoic acid, butyl ester                                   | 000591-68-4 | 1138.0 |
| 223^ | Pentanoic acid, ethyl ester                                   | 000539-82-2 | 927.7  |
| 224  | Pentyl octanoate  | 000638-25-5 | 1523.0 |
| 225* | Phenol, 2-(1-methylethyl)-                                    | 000088-69-7 | 1147.6 |
| 226^ | Phenol, 2-methoxy-  | 000090-05-1 | 1072.6 |
| 227^ | Phenol, 2-methyl-5-(1-methylethyl)-                           | 000499-75-2 | 1215.8 |
| 228* | Phenylethyl Alcohol   | 000060-12-8 | 1130.5 |
| 229* | Piperonal   | 000120-57-0 | 1240.1 |
| 230* | Propanedioic acid, diethyl ester                              | 000105-53-3 | 1001.6 |
| 231  | Propanethioic acid, S-(2-furanylmethyl) ester                 | 059020-85-8 | 1310.3 |
| 232  | Propanoic acid, 2-methyl-                                     | 000079-31-2 | 681.9  |
| 233^ | Propanoic acid, 2-methyl-, 2-phenylethyl ester                | 000103-48-0 | 1451.6 |
| 234* | Propanoic acid, 2-methyl-, 3-phenyl-2-propenyl ester          | 000103-59-3 | 1558.9 |
| 235  | Propanoic acid, 2-methyl-, 4-formyl-2-methoxyphenyl ester     | 020665-85-4 | 1588.1 |
| 236  | Propanoic acid, 2-methyl-, ethyl ester                        | 000097-62-1 | 771.5  |

|                  |   |             |        |
|------------------|---|-------------|--------|
| 237 <sup>^</sup> | Propanoic acid, 2-methyl-, octyl ester        | 000109-15-9 | 1386.8 |
| 238 <sup>^</sup> | Propanoic acid, 2-methyl-, phenylmethyl ester | 000103-28-6 | 1376.8 |
| 239              | Pyrazine, 2,3-diethyl-5-methyl-               | 018138-04-0 | 1203.3 |
| 240              | Pyrazine, 2,3-dimethyl-                       | 005910-89-4 | 912.6  |
| 241 <sup>*</sup> | Pyrazine, 2-ethyl-3,5-dimethyl-               | 013925-07-0 | 1074.4 |
| 242              | Pyrazine, 2-ethyl-3-methyl-                   | 015707-23-0 | 1014.9 |
| 243 <sup>*</sup> | Pyrazine, 2-methoxy-3-(2-methylpropyl)-       | 024683-00-9 | 1210.6 |
| 244 <sup>*</sup> | Pyrazine, 2-methyl-3-(methylthio)-            | 002882-20-4 | 1111.9 |
| 245 <sup>^</sup> | Pyrazine, 2-methyl-6-(methylthio)-            | 002884-13-1 | 1107.2 |
| 246 <sup>^</sup> | Pyrazine, 3-ethyl-2,5-dimethyl-               | 013360-65-1 | 1107.5 |
| 247 <sup>*</sup> | Pyrazine, ethyl-                              | 013925-00-3 | 948.7  |
| 248              | Pyrazine, tetramethyl-                        | 001124-11-4 | 1054.3 |
| 249 <sup>*</sup> | Pyrazine, trimethyl-                          | 014667-55-1 | 983.3  |
| 250 <sup>^</sup> | Tetradecanoic acid, ethyl ester               | 000124-06-1 | 1785.1 |
| 251              | Thiazole, 5-ethenyl-4-methyl-                 | 001759-28-0 | 1030.9 |
| 252 <sup>^</sup> | trans-Isoeugenol                              | 005932-68-3 | 1328.8 |
| 253              | Triacetin                                     | 000102-76-1 | 1327.7 |
| 254              | Triethyl citrate                              | 000077-93-0 | 1722.6 |
| 255 <sup>^</sup> | Undecanal                                     | 000112-44-7 | 1297.6 |
| 256 <sup>^</sup> | $\alpha$ -Ionone                              | 000127-41-3 | 1345.2 |
| 257 <sup>*</sup> | $\alpha$ -Irone                               | 000079-69-6 | 1435.1 |
| 258              | $\alpha$ -Phellandrene                        | 000099-83-2 | 1089.8 |
| 259              | $\beta$ -Myrcene                              | 000123-35-3 | 1025.3 |
| 260 <sup>^</sup> | $\beta$ -Phenylethyl butyrate                 | 000103-52-6 | 1475.5 |
| 261              | $\beta$ -Pinene                               | 000127-91-3 | 1112.4 |
| 262 <sup>*</sup> | $\delta$ -Nonalactone                         | 003301-94-8 | 1215.2 |

\* validación, <sup>^</sup> predicción

**Tabla 6A.** Nombres químicos, número de registro CAS e índices de retención experimentales de 137 compuestos volátiles de arroz medidos en SPME/GC/MS

| No. | Nombre                 | CAS        | I     |
|-----|------------------------|------------|-------|
| 1   | Ethanol                | 64-17-5    | 250   |
| 2   | Acetone                | 67-64-1    | 259   |
| 3   | Dimethyl sulfide       | 75-18-3    | 308   |
| 4*  | Hexane                 | 110-54-3   | 600   |
| 5   | Butanal                | 123-72-8   | 602   |
| 6   | Acetic acid            | 64-19-7    | 622   |
| 7^  | 3-Methyl-butanal       | 590-86-3   | 652   |
| 8*  | 2-Methyl-butanal       | 96-17-3    | 660   |
| 9   | Heptane                | 142-82-5   | 700   |
| 10* | Pentanal               | 110-62-3   | 701   |
| 11  | Methyl butanoate       | 623-42-7   | 710   |
| 12  | 3-Methyl-butanol       | 123-51-3   | 730   |
| 13^ | 2-Methyl-butanol       | 137-32-6   | 730   |
| 14  | Dimethyl disulfide     | 624-92-0   | 741   |
| 15  | Pyridine               | 110-86-1   | 763   |
| 16* | 3-Hexanone             | 589-38-8   | 744   |
| 17^ | 1-Pentanol             | 71-41-0    | 761   |
| 18^ | Toluene                | 108-88-3   | 764   |
| 19* | 2,4-Pentandione        | 123-54-6   | 778   |
| 20  | 2,3-Butandiol          | 513-85-9   | 788   |
| 21* | 1,3-Butandiol          | 107-88-0   | 791   |
| 22* | Octane                 | 111-65-9   | 800   |
| 23  | Hexanal                | 66-25-1    | 800   |
| 24* | Methyl pentanoate      | 624-24-8   | 821   |
| 25* | Ethenyl cyclohexane    | 100-40-3   | 832   |
| 26  | Butanoic acid          | 107-92-6   | 638   |
| 27^ | (E)-2-Hexenal          | 6728-26-3  | 850   |
| 28* | (Z)-3-hexenal          | 6789-80-6  | 853   |
| 29  | Ethyl benzene          | 100-41-4   | 858   |
| 30* | 1-Hexanol              | 111-27-3   | 865   |
| 31^ | 1,4-Dimethyl benzene   | 106-42-3   | 868   |
| 32^ | Pentanoic acid         | 109-52-4   | 879   |
| 33  | 2,2,4-Trimethylheptane | 14720-74-2 | 878.5 |
| 34^ | 2-Heptanone            | 110-43-0   | 888   |
| 35  | 2-Butylfuran           | 4466-24-4  | 890   |
| 36^ | Styrene                | 100-42-5   | 892   |
| 37* | 1,2-dimethylbenzene    | 94-47-6    | 893   |
| 38^ | Nonane                 | 111-84-2   | 900   |

|                 |                                    |             |      |
|-----------------|------------------------------------|-------------|------|
| 39              | Heptanal                           | 111-71-7    | 902  |
| 40 <sup>^</sup> | 2-Butoxyethanol                    | 111-76-2    | 905  |
| 41 <sup>*</sup> | 2,3,6-Trimethylheptane             | 4032-93-3   | 913  |
| 42              | 2-Acetyl-1-pyrroline               | 85213-22-05 | 920  |
| 43              | Methyl hexanoate                   | 106-70-7    | 922  |
| 44              | alpha-Pinene                       | 80-56-8     | 932  |
| 45 <sup>^</sup> | 1-Butoxy-2-propanol                | 5131-66-8   | 938  |
| 46 <sup>*</sup> | (E)-2-Heptenal                     | 57266-86-1  | 956  |
| 47 <sup>^</sup> | 1-Ethyl, 4-methylbenzene           | 611-14-3    | 959  |
| 48              | Benzaldehyde                       | 100-52-7    | 962  |
| 49 <sup>^</sup> | 1-Heptanol                         | 111-70-6    | 969  |
| 50              | Dimethyl trisulfide                | 3658-80-8   | 970  |
| 51 <sup>^</sup> | Hexanoic acid                      | 142-62-1    | 983  |
| 52 <sup>^</sup> | 6-Methyl-5-hepten-2-one            | 110-93-0    | 983  |
| 53 <sup>*</sup> | 2-Pentylfuran                      | 3777-69-3   | 990  |
| 54              | 2,3,6-Trimethylpyridine            | 1462-84-6   | 989  |
| 55 <sup>*</sup> | 1,2,4-Trimethylbenzene             | 95-63-6     | 994  |
| 56 <sup>*</sup> | Ethyl hexanoate                    | 123-66-0    | 998  |
| 57              | Decane                             | 124-18-5    | 1000 |
| 58 <sup>^</sup> | Octanal                            | 124-13-0    | 1004 |
| 59 <sup>^</sup> | 2-Ethyl-1-hexanol                  | 104-76-7    | 1029 |
| 60 <sup>*</sup> | Limonene                           | 138-86-3    | 1030 |
| 61              | Indane                             | 496-11-7    | 1035 |
| 62 <sup>*</sup> | Benzyl alcohol                     | 100-51-6    | 1036 |
| 63 <sup>*</sup> | (E)-3-Octen-2-one                  | 18402-82-9  | 1038 |
| 64 <sup>^</sup> | Benzeneacetaldehyde                | 122-78-1    | 1045 |
| 65 <sup>*</sup> | 5-Ethyldihydro-2(3H)-furanone      | 695-06-7    | 1049 |
| 66 <sup>^</sup> | 2-Octenal                          | 2548-87-0   | 1058 |
| 67 <sup>*</sup> | 1-Octanol                          | 111-87-5    | 1071 |
| 68 <sup>*</sup> | 2-Methyl-1-propenyl benzene        | 768-49-0    | 1082 |
| 69              | 1-Methyl,2-(1-methylethyl)-benzene | 527-84-4    | 1084 |
| 70 <sup>*</sup> | Undecane                           | 1120-21-4   | 1100 |
| 71 <sup>^</sup> | Nonanal                            | 124-19-6    | 1106 |
| 72 <sup>^</sup> | 1,2,3,4-Tetramethylbenzene         | 488-23-3    | 1123 |
| 73              | Methyl octanoate                   | 111-11-5    | 1128 |
| 74 <sup>*</sup> | 5-Ethyl-6-methyl-(E)3-hepten-2-one | 57283-79-1  | 1147 |
| 75 <sup>*</sup> | 3-Methyl-2-heptyl acetate          | 72218-58-7  | 1151 |
| 76 <sup>^</sup> | 5-Propyldihydro-2(3H)-furanone     | 105-21-5    | 1156 |
| 77 <sup>*</sup> | Beta-Terpineol                     | 138-87-4    | 1157 |
| 78              | (E)-2-nonenal                      | 18829-56-6  | 1166 |
| 79 <sup>^</sup> | Nonanol                            | 143-08-8    | 1176 |
| 80 <sup>*</sup> | Naphthalene                        | 91-20-3     | 1190 |

|      |   |            |      |
|------|---|------------|------|
| 81^  | Ethyl octanoate                               | 106-32-1   | 1196 |
| 82^  | Alpha-Terpineol                               | 98-55-5    | 1198 |
| 83^  | Dodecane                                      | 112-40-3   | 1200 |
| 84   | Gamma-terpineol                               | 586-81-2   | 1203 |
| 85   | Decanal                                       | 112-31-2   | 1207 |
| 86^  | (E,E)-2,4-Nonadienal                          | 5910-87-2  | 1218 |
| 87   | Benzothiazole                                 | 85-16-9    | 1234 |
| 88*  | 2-Hexyl-1-octanol                             | 19780-79-1 | 1254 |
| 89   | 5-Butyldihydro-2(3H)-furanone                 | 104-50-7   | 1259 |
| 90^  | (E)-2-Decenal                                 | 3913-81-3  | 1267 |
| 91^  | 1-Tridecene                                   | 2437-56-1  | 1293 |
| 92*  | 3-(t-Butyl)-phenol                            | 585-34-2   | 1295 |
| 93^  | Ethyl nonanoate                               | 123-29-5   | 1296 |
| 94^  | 2-Undecanone                                  | 112-12-9   | 1300 |
| 95*  | Indole  | 120-72-9   | 1300 |
| 96   | Tridecane                                     | 629-50-5   | 1300 |
| 97*  | Undecanal                                     | 112-44-7   | 1310 |
| 98*  | Methyl decanoate                              | 110-42-9   | 1325 |
| 99*  | 5-Pentyldihydro-2(3H)-furanone                | 104-61-0   | 1368 |
| 100^ | 2-Butyl-2-octenal                             | 13019-16-4 | 1375 |
| 101* | Tetradec-1-ene                                | 1120-36-1  | 1392 |
| 102  | Ethyl decanoate                               | 110-38-3   | 1394 |
| 103^ | Tetradecane                                   | 629-59-4   | 1400 |
| 104  | Isolongifolene                                | 1135-66-6  | 1408 |
| 105  | trans-caryophyllene                           | 87-44-5    | 1433 |
| 106  | Bis-(1-methylethyl)-hexadecanoate             | 6938-94-9  | 1448 |
| 107^ | (E)-6,10-dimethyl-5,9-undecadien-2-one        | 3879-26-3  | 1450 |
| 108* | 2,6-Bis-(t-butyl)-2,5-cyclohexadien-1,4-dione | 719-22-2   | 1471 |
| 109* | 1-Hexadecene                                  | 629-73-2   | 1492 |
| 110^ | Pentadecane                                   | 629-62-9   | 1500 |
| 111  | BHT   | 128-37-0   | 1510 |
| 112^ | Methyl dodecanoate                            | 111-82-0   | 1523 |
| 113  | 1-S, cis-calamenene                           | 483-77-2   | 1534 |
| 114* | Ionol 2                                       | 4130-42-1  | 1561 |
| 115* | Ethyl dodecanoate                             | 106-33-2   | 1592 |
| 116  | Hexadecane                                    | 544-76-3   | 1600 |
| 117^ | 2,6-bis-(t-butyl)-2,5-cyclohexadien-1-one     | 6378-27-8  | 1637 |
| 118^ | 2,5,10-Trimethylpentadecane                   | 3892-00-0  | 1646 |
| 119^ | Heptadecane                                   | 629-78-7   | 1700 |
| 120  | 2,5,10,14-Tetramethylpentadecane              | 1921-70-6  | 1703 |
| 121^ | Methyl tetradecanoate                         | 124-10-7   | 1724 |
| 122  | Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl  | 3910-35-8  | 1731 |

|      |                                   |            |      |
|------|-----------------------------------|------------|------|
| 123* | 2,6-Diisopropylnaphthalene        | 24157-81-1 | 1744 |
| 124* | 2-Methyl-2,4-diphenylpentane      | 31516-55-9 | 1762 |
| 125  | Ethyl tetradecanoate              | 124-06-1   | 1793 |
| 126* | Octadecane                        | 593-45-3   | 1800 |
| 127* | 2,6,10,14-Tetramethylhexadecane   | 638-36-8   | 1807 |
| 128* | 6,10,14-Trimethyl-2-pentadecanone | 502-69-2   | 1844 |
| 129  | Nonadecane                        | 629-92-5   | 1900 |
| 130* | Methyl hexadecanoate              | 112-39-0   | 1925 |
| 131^ | Ethyl hexadecanoate               | 628-97-7   | 1993 |
| 132^ | Hexadecanoic acid                 | 57-10-3    | 1995 |
| 133^ | Eicosane                          | 112-95-8   | 2000 |
| 134* | Methyl linoleate                  | 2566-97-4  | 2051 |
| 135  | Methyl oleate                     | 112-62-9   | 2052 |
| 136^ | Ethyl linoleate                   | 544-35-4   | 2081 |
| 137  | Ethyl oleate                      | 111-62-6   | 2086 |

\* validación, ^ predicción

**Tabla 7A.** Nombres químicos y dulzor relativo experimental de 233 compuestos dulcificantes

| No.             | Nombre                         | log(RS) |
|-----------------|--------------------------------|---------|
| 1               | Sucrose                        | 0.000   |
| 2               | Sucralose                      | 2.778   |
| 3               | Alitame                        | 3.301   |
| 4 <sup>^</sup>  | Aspartame                      | 2.255   |
| 5               | Tagatose                       | -0.046  |
| 6 <sup>^</sup>  | Maltitol                       | -0.046  |
| 7               | Isomaltulose                   | -0.319  |
| 8 <sup>^</sup>  | Trehalose                      | -0.347  |
| 9               | Neohesperidine dihydrochalcone | 3.255   |
| 10              | Neotame                        | 3.903   |
| 11              | Lactitol                       | -0.398  |
| 12              | Isomalt                        | -0.260  |
| 13              | Sorbitol                       | -0.222  |
| 14 <sup>^</sup> | Mannitol                       | -0.155  |
| 15              | Erythritol                     | -0.155  |
| 16              | Xylitol                        | 0.000   |
| 17              | Stevioside                     | 2.477   |
| 18              | Steviolbioside                 | 2.051   |
| 19              | Rebaudioside A                 | 2.544   |
| 20              | Rebaudioside B                 | 2.512   |
| 21 <sup>^</sup> | Rebaudioside C                 | 1.929   |
| 22              | Rebaudioside D                 | 2.512   |
| 23              | Rebaudioside E                 | 2.352   |
| 24 <sup>^</sup> | Dulcoside A                    | 1.929   |
| 25              | Glycyrrhizin                   | 1.875   |
| 26              | Hernandulcin                   | 3.000   |
| 27              | Fructose                       | 0.057   |
| 28              | Glucose                        | -0.161  |
| 29              | Galactose                      | -0.201  |
| 30 <sup>^</sup> | Mannose                        | -0.229  |
| 31 <sup>^</sup> | Xylose                         | -0.174  |
| 32              | Maltose                        | -0.337  |
| 33 <sup>^</sup> | Lactose                        | -0.398  |
| 34 <sup>^</sup> | Lactulose                      | -0.222  |
| 35 <sup>^</sup> | Lactosucrose                   | -0.347  |
| 36 <sup>^</sup> | Galactitol                     | -0.387  |
| 37              | Raffinose                      | -0.658  |

|     |  |        |
|-----|--|--------|
| 38^ | Rhamnose   | -0.481 |
| 39  | Glycerol   | -0.097 |
| 40  | 2-amino-4-nitro-propoxybenzene   | 3.602  |
| 41  | Dulcin   | 2.301  |
| 42^ | Dihydroquercetin 3-acetate 4'-(methylether)  | 2.602  |
| 43^ | 1,3-benzodioxane   | 3.477  |
| 44  | 1,4-benzodioxane   | 2.653  |
| 45  | Isoflavan  | 2.477  |
| 46  | 2-(3-hydroxy-4-methoxyphenyl)-1,3-benzodioxole                                     | 2.176  |
| 47^ | Bibenzyl derivative  | 2.477  |
| 48^ | Flavan   | 2.544  |
| 49^ | Compound 5   | 2.544  |
| 50^ | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzoxathiine                                 | 2.699  |
| 51  | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzodithiole                                 | 2.301  |
| 52^ | 2-(3-Hydroxy-4-methoxyphenyl)-1,3-benzoxathiole                                    | 2.477  |
| 53^ | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiine                        | 2.398  |
| 54  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzodithiine                        | 2.699  |
| 55^ | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol                    | 2.699  |
| 56  | 2-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran                      | 2.301  |
| 57  | 2-(3-Mercapto-4-methoxyphenyl)-4H-3,1-benzoxathiine                                | 1.699  |
| 58  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiine S,S-Dioxide            | 1.699  |
| 59^ | Haematoxylin   | 2.079  |
| 60  | 9-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,10-diol                        | 1.699  |
| 61^ | Compound 8   | 1.699  |
| 62  | Compound 18  | 2.079  |
| 63  | Compound 20  | 2.778  |
| 64^ | Compound 21  | 3.041  |
| 65  | 2-(4-methoxybenzoyl) benzoic acid  | 2.176  |
| 66  | 2-(3,4-Dihydroxybenzoyl) benzoic acid  | 2.000  |
| 67  | 2-(3-Hydroxy-4-methoxybenzoyl) benzoic acid  | 2.398  |
| 68  | 2-(4-Methoxyphenylthio) benzoic acid S-Oxide                                       | 1.699  |
| 69^ | Phthalic Acid Mono(4-methoxyphenyl) Ester  | 1.000  |
| 70  | Compound 9   | 2.301  |
| 71  | 3-Methoxy-1,3,5-estratrien-4-ol  | 1.699  |
| 72  | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-methoxy-6'hydroxy-[1H]-2',3'-dihydroindene)] | 1.699  |
| 73^ | Compound 14  | 4.255  |
| 74  | Compound 15  | 4.301  |

|                  |  |       |
|------------------|--|-------|
| 75               | Phyllodulcin   | 2.903 |
| 76 <sup>^</sup>  | Compound 4   | 4.301 |
| 77               | Compound 6   | 3.176 |
| 78               | Compound 9   | 2.176 |
| 79               | Compound 10  | 2.000 |
| 80 <sup>^</sup>  | 3',4'-dihydroxyphenyl-3,1-benzoxathiane                                | 2.301 |
| 81               | 2-[3-(4-nitrophenoxy)phenyl]-3,1-benzoxathiane                         | 2.477 |
| 82               | 2-[3-(4-nitrophenylcarbonylmethoxy)-4-methoxyphenyl]-3,1-benzoxathiane | 2.477 |
| 83               | 2-[3-(4-nitrophenylcarbamoxyloxy-4-methoxy)phenyl]-3,1-benzodioxane    | 2.176 |
| 84               | 3-(3-Hydroxy-4-methoxyphenyl)-1-(3-carboxyphenyl)propan-1-one          | 1.903 |
| 85               | 2-(3-hydroxy-4-methoxyphenoxy)methyl)benzoic acid                      | 1.000 |
| 86 <sup>^</sup>  | Compound 32  | 2.823 |
| 87               | Compound 42  | 1.699 |
| 88 <sup>^</sup>  | Dihydroquercetin 4'-(methyl ether)                                     | 1.602 |
| 89               | Compound V   | 1.875 |
| 90               | Compound VII   | 2.000 |
| 91               | Compound XIV   | 1.000 |
| 92 <sup>^</sup>  | Compound XXI   | 2.477 |
| 93               | Compound XXII  | 2.176 |
| 94               | Compound XXIII   | 2.176 |
| 95               | Compound XXIV  | 2.398 |
| 96 <sup>^</sup>  | Compound XXVII   | 1.000 |
| 97 <sup>^</sup>  | Compound XXX   | 2.079 |
| 98               | Compound XXXI  | 1.477 |
| 99               | Compound XXXII   | 1.477 |
| 100              | Compound XL  | 2.477 |
| 101 <sup>^</sup> | Compound XLI   | 2.176 |
| 102 <sup>^</sup> | Compound XLII  | 2.097 |
| 103              | Compound XLIII   | 1.000 |
| 104 <sup>^</sup> | Compound XLVII   | 1.000 |
| 105 <sup>^</sup> | Compound XLIX  | 1.000 |
| 106 <sup>^</sup> | Compound LIV   | 1.000 |
| 107              | Compound LIX   | 1.477 |
| 108 <sup>^</sup> | Compound LXV   | 1.000 |
| 109 <sup>^</sup> | Compound LXVI  | 1.000 |
| 110              | Compound LXXVIII   | 2.000 |
| 111              | Compound LXXIX   | 2.398 |
| 112              | Compound LXXX  | 1.000 |
| 113              | Compound LXXXV   | 1.875 |

---

|      |   |        |
|------|---|--------|
| 114^ | Compound LXXXIX                         | 2.544  |
| 115  | Compound XC                             | 2.176  |
| 116^ | Compound XCI                            | 1.778  |
| 117  | Compound XCV                            | 2.000  |
| 118^ | Compound XCVI                           | 1.000  |
| 119  | Glycol                                  | 0.114  |
| 120^ | Arabitol                                | 0.000  |
| 121  | Stachyose                               | -1.000 |
| 122^ | 6-O-Methyl sucrose                      | 0.000  |
| 123  | 6,6'-di-O-Methyl sucrose                | 0.000  |
| 124  | 6-Chloro-6-deoxy-D-fructofuranose       | 0.000  |
| 125^ | 1,6-dichloro-1,6-deoxy-D-fructofuranose | 0.000  |
| 126  | Methyl $\alpha$ -D-glucopyranoside      | -0.602 |
| 127  | 4-Chloro-galactosucrose                 | 0.699  |
| 128  | 6'-Chloro-sucrose                       | 1.301  |
| 129  | 1'-Chloro-sucrose                       | 1.301  |
| 130^ | 1',6'-Dichloro-sucrose                  | 1.845  |
| 131  | 1',4-Dichloro-galactosucrose            | 2.079  |
| 132  | 1,6-Di-S-1,6-dithio-D-fructofuranose    | 1.301  |
| 133  | Alanine                                 | -0.005 |
| 134  | Arginine                                | 0.624  |
| 135  | Asparagine                              | 0.301  |
| 136  | glutamine                               | 0.322  |
| 137  | Glycine                                 | -0.182 |
| 138  | Histidine                               | 0.746  |
| 139  | Isoleucine                              | -0.104 |
| 140  | Leucine                                 | 0.599  |
| 141  | Methionine                              | 0.294  |
| 142  | Norvaline                               | 0.230  |
| 143  | Phenylalanine                           | 0.689  |
| 144  | Proline                                 | -0.268 |
| 145  | Hydroxy proline                         | 0.262  |
| 146  | Serine                                  | -0.210 |
| 147  | Threonine                               | -0.569 |
| 148  | Tryptophane                             | 1.556  |
| 149^ | Tyrosine                                | 0.740  |
| 150  | Aspartylmalonic acid diester 1          | 2.778  |
| 151  | Aspartylmalonic acid diester 2          | 2.944  |
| 152  | Aspartylmalonic acid diester 3          | 4.521  |
| 153  | N-(L-aspartyl)-1,1-diaminoalkane 1      | 2.000  |

---

---

|      |                                       |       |
|------|---------------------------------------|-------|
| 154  | N-(L-aspartyl)-1,1-diaminoalkane 2    | 2.845 |
| 155  | N-(L-aspartyl)-1,1-diaminoalkane 3    | 2.000 |
| 156  | N-(L-aspartyl)-1,1-diaminoalkane 4    | 3.000 |
| 157  | N-(L-aspartyl)-1,1-diaminoalkane 5    | 4.643 |
| 158  | N-(L-aspartyl)-1,1-diaminoalkane 6    | 1.875 |
| 159^ | N-(L-aspartyl)-1,1-diaminoalkane 7    | 2.301 |
| 160  | N-(L-aspartyl)-1,1-diaminoalkane 8    | 2.000 |
| 161  | N-(L-aspartyl)-1,1-diaminoalkane 9    | 1.176 |
| 162^ | Super aspartame                       | 3.903 |
| 163  | Sucrononic acid                       | 5.301 |
| 164  | Cyanoarylurea aspartame               | 3.892 |
| 165^ | Aspartic acid fenchyl ester           | 4.699 |
| 166  | Glycine analog of cyanosuosan         | 3.845 |
| 167  | Glycine analog of suosan              | 1.000 |
| 168  | Nitroaniline (2-amino-4-nitrobenzene) | 1.602 |
| 169  | Methoxy-nitroaniline                  | 2.418 |
| 170  | Ethoxy-nitroaniline                   | 3.047 |
| 171^ | Butoxy-nitroaniline                   | 3.000 |
| 172  | Fluoro-nitroaniline                   | 1.602 |
| 173  | Chloro-nitroaniline                   | 2.602 |
| 174  | Bromo-nitroaniline                    | 2.903 |
| 175^ | Iodo-nitroaniline                     | 3.097 |
| 176  | Allyl oxy-nitroaniline                | 3.301 |
| 177^ | iso Propoxy-nitroaniline              | 2.778 |
| 178  | Hydroxy-nitroaniline                  | 2.204 |
| 179  | Saccharine derivative 13              | 2.653 |
| 180  | Saccharine derivative 23              | 2.352 |
| 181  | Thiophenesaccharine                   | 3.000 |
| 182^ | Suosan derivative 1                   | 2.845 |
| 183  | Suosan derivative 11                  | 3.380 |
| 184  | Suosan derivative 13                  | 1.301 |
| 185  | Suosan derivative 16                  | 2.653 |
| 186  | Oxathiazinon dioxide derivative 1     | 2.114 |
| 187  | Oxathiazinon dioxide derivative 2     | 2.176 |
| 188  | Oxathiazinon dioxide derivative 3     | 2.114 |
| 189  | Oxathiazinon dioxide derivative 4     | 2.398 |
| 190^ | Oxathiazinon dioxide derivative 5     | 2.114 |
| 191  | Oxathiazinon dioxide derivative 6     | 1.301 |
| 192  | 2-aminobenzoic acid                   | 2.079 |
| 193^ | 1',4',6'-3Cl-sucrose                  | 2.000 |

---

---

|      |  |        |
|------|--|--------|
| 194^ | 1',4'-2Cl-sucrose                      | 1.477  |
| 195  | 4'-F-4,1',6'-3Cl-sucrose               | 3.000  |
| 196^ | 4,1',4',6'-4Cl-sucrose                 | 3.322  |
| 197  | 4,1',4'-3Cl-sucrose                    | 2.342  |
| 198  | 4,1',6'-3Br-sucrose                    | 2.903  |
| 199^ | 4,1',6'-3Cl-sucrose (trichlorosucrose) | 2.813  |
| 200  | 4,1',6'-3F-sucrose                     | 1.602  |
| 201^ | 4,4',6'-3Cl-sucrose                    | 2.204  |
| 202  | 4,6,1',6'-4Cl-sucrose                  | 2.301  |
| 203  | 4-F-1',4',6'-3Cl-sucrose               | 2.778  |
| 204  | 4-Hydroxy glycosides DHC               | 2.000  |
| 205  | 6,1',6'-3Cl-sucrose                    | 1.398  |
| 206^ | Psicose                                | -0.155 |
| 207  | Ribose                                 | -0.155 |
| 208  | DHQ-3-SIFA                             | 1.903  |
| 209^ | Hesperidin-7-glucose DHC               | 2.000  |
| 210  | Sorbose                                | 0.000  |
| 211  | Arabinose                              | -0.161 |
| 212  | Fucose                                 | -0.161 |
| 213  | Liquiritin                             | 2.398  |
| 214^ | Naringin DHC                           | 2.000  |
| 215  | Periandrin I                           | 1.929  |
| 216  | Periandrin II                          | 2.000  |
| 217^ | Periandrin III                         | 2.000  |
| 218  | Periandrin IV                          | 2.000  |
| 219  | Perillartine                           | 3.301  |
| 220  | Polypodosides A                        | 2.778  |
| 221  | Rubusoside                             | 2.057  |
| 222  | Siamenoside I                          | 2.748  |
| 223  | Yang Li glycosides DHC                 | 1.602  |
| 224^ | Acesulfame potassium_CH3CH2_H          | 1.301  |
| 225  | Acesulfame potassium_H_H               | 1.000  |
| 226  | Baiyunoside                            | 2.699  |
| 227^ | Mogroside II                           | 1.924  |
| 228  | Mogroside IV                           | 2.097  |
| 229  | Mogroside V                            | 2.477  |
| 230^ | Osladin                                | 3.477  |
| 231  | Rebaudioside F                         | 2.051  |
| 232^ | Saccharin                              | 2.653  |
| 233  | Xylobiose                              | -0.398 |

---

^ predicción

**Tabla 8A.** Nombres químicos y respuesta experimental de los 508 compuestos dulces y amargos

| No.             | Nombre                         | Clase  |
|-----------------|--------------------------------|--------|
| 1               | Sucrose                        | Dulce  |
| 2               | Sucralose                      | Dulce  |
| 3               | Alitame                        | Dulce  |
| 4               | Aspartame                      | Dulce  |
| 5               | Tagatose                       | Dulce  |
| 6               | Maltitol                       | Dulce  |
| 7               | Isomaltulose (Palatinose)      | Dulce  |
| 8               | Trehalose                      | Dulce  |
| 9               | Neohesperidine dihydrochalcone | Dulce  |
| 10              | Neotame                        | Dulce  |
| 11              | Isomalt (Palatinit)            | Dulce  |
| 12              | Sorbitol                       | Dulce  |
| 13 <sup>^</sup> | Erythritol                     | Dulce  |
| 14 <sup>^</sup> | Xylitol                        | Dulce  |
| 15              | Stevioside                     | Dulce  |
| 16              | Steviolbioside                 | Dulce  |
| 17 <sup>^</sup> | Rebaudioside A                 | Dulce  |
| 18              | Rebaudioside B                 | Dulce  |
| 19 <sup>^</sup> | Rebaudioside C                 | Dulce  |
| 20              | Rebaudioside D                 | Dulce  |
| 21              | Rebaudioside E                 | Dulce  |
| 22              | Dulcoside A                    | Dulce  |
| 23              | Glycyrrhizin                   | Dulce  |
| 24              | Hernandulcin                   | Dulce  |
| 25              | Fructose                       | Dulce  |
| 26 <sup>^</sup> | Glucose                        | Dulce  |
| 27              | Xylose                         | Dulce  |
| 28              | Maltose                        | Dulce  |
| 29              | Lactulose                      | Dulce  |
| 30              | Lactosucrose                   | Dulce  |
| 31 <sup>^</sup> | Raffinose                      | Dulce  |
| 32              | Rhamnose                       | Dulce  |
| 33              | Naringin                       | Amargo |
| 34 <sup>^</sup> | Gentiobiose                    | Amargo |
| 35              | Limonin                        | Amargo |
| 36              | Oleuropein                     | Amargo |
| 37 <sup>^</sup> | Caffeine                       | Amargo |
| 38              | Poncirin                       | Amargo |
| 39              | Neoeriocitrin                  | Amargo |
| 40              | Quinine                        | Amargo |
| 41 <sup>^</sup> | Divinyl glycol                 | Amargo |
| 42              | Theobromine                    | Amargo |
| 43              | Picrocrocin                    | Amargo |
| 44              | Glycerol                       | Dulce  |
| 45              | Saccharin derivative 1         | Dulce  |
| 46 <sup>^</sup> | Saccharin derivative 2         | Dulce  |

|     |  |        |
|-----|--|--------|
| 47^ | 2-amino-4-nitro-propoxybenzene   | Dulce  |
| 48  | 2,4-dinitro-propoxybenzene   | Amargo |
| 49^ | Dulcin (p-ethoxyphenylurea)  | Dulce  |
| 50  | 4-Ethoxyphenylthiourea   | Amargo |
| 51  | Anti-Anisaldehyde Oxime  | Dulce  |
| 52  | Dihydroquercetin 3-acetate 4'-(methylether)  | Dulce  |
| 53  | Flavan   | Dulce  |
| 54  | Compound 5   | Dulce  |
| 55^ | 1,3-benzodioxane   | Dulce  |
| 56^ | 1,4-benzodioxane   | Dulce  |
| 57  | Bibenzyl derivative  | Dulce  |
| 58  | Isoflavan or 3-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1benzopyran              | Dulce  |
| 59^ | 2-(3-hydroxy-4-methoxyphenyl)-1,3-benzodioxole                                     | Dulce  |
| 60  | 2-(3-Hydroxy-4-methoxyphenyl)-4-H-1,3-benzodithiine                                | Dulce  |
| 61^ | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzoxathiine                                 | Dulce  |
| 62  | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzodithiole                                 | Dulce  |
| 63  | 2-(3-Hydroxy-4-methoxyphenyl)-1,3-benzoxathiole                                    | Dulce  |
| 64^ | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiine                        | Dulce  |
| 65  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzodithiine                        | Dulce  |
| 66  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol                    | Dulce  |
| 67  | 2-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran                      | Dulce  |
| 68  | 2-(3-Mercapto-4-methoxyphenyl)-4H-3,1-benzoxathiine                                | Dulce  |
| 69  | Sulfone  | Dulce  |
| 70^ | Haematoxylin   | Dulce  |
| 71  | 9-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,10-diol                        | Dulce  |
| 72^ | Tetramethylhaematoxylin  | Amargo |
| 73^ | Compound 8   | Dulce  |
| 74  | Compound 18  | Dulce  |
| 75  | Compound 20  | Dulce  |
| 76  | Compound 21  | Dulce  |
| 77  | Compound 26  | Dulce  |
| 78^ | 2-(4-methoxybenzoyl) benzoic acid  | Dulce  |
| 79  | Compound 3   | Dulce  |
| 80  | Compound 4   | Dulce  |
| 81  | Compound 7   | Dulce  |
| 82  | Compound 9   | Dulce  |
| 83^ | 2-(4-Nitrobenzoyl)benzoic Acid   | Amargo |
| 84  | Compound 12  | Dulce  |
| 85  | 2-(3,4-Dihydroxybenzoyl)benzoic Acid   | Dulce  |
| 86  | 2-(3-Hydroxy-4-methoxybenzoyl)benzoic Acid   | Dulce  |
| 87  | 2-(2,4-Dimethoxybenzoyl)benzoic Acid   | Amargo |
| 88^ | 2-(4-Methoxyphenylthio) benzoic Acid S-Oxide                                       | Dulce  |
| 89^ | Phthalic Acid Mono(4-methoxyphenyl) Ester  | Dulce  |
| 90  | 3-Methoxy-1,3,5-estratrien-4-ol  | Dulce  |
| 91^ | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-methoxy-6'hydroxy-[1H]-2',3'-dihydroindene)] | Dulce  |

|      |  |        |
|------|--|--------|
| 92   | Compound 14  | Dulce  |
| 93   | Compound 15  | Dulce  |
| 94   | Phyllo dulcin  | Dulce  |
| 95^  | Compound 6   | Dulce  |
| 96   | Compound 9   | Dulce  |
| 97^  | 3',4'-dihydroxyphenyl-3,1-benzoxathiane                                | Dulce  |
| 98   | 2-[3-(4-nitrophenoxy)phenyl]-3,1-benzoxathiane                         | Dulce  |
| 99   | 2-[3-(4-nitrophenylcarbonylmethoxy)-4-methoxyphenyl]-3,1-benzoxathiane | Dulce  |
| 100^ | 2-[3-(4-nitrophenylcarbamoyloxy)-4-methoxyphenyl]-3,1-benzodioxane     | Dulce  |
| 101^ | Compound 32  | Dulce  |
| 102  | 3-(3-Hydroxy-4-methoxyphenyl)-1-(3-carboxyphenyl)propan-1-one          | Dulce  |
| 103  | Compound 34  | Dulce  |
| 104  | 2-(3-hydroxy-4-methoxyphenoxy)methyl)benzoic acid                      | Dulce  |
| 105^ | Compound 42  | Dulce  |
| 106  | Dihydroquercetin 4'-(methyl ether)                                     | Dulce  |
| 107^ | Compound III   | Amargo |
| 108  | Compound V   | Dulce  |
| 109  | Compound VII   | Dulce  |
| 110  | Compound XIV   | Dulce  |
| 111  | Compound XVIII   | Amargo |
| 112  | Compound XIX   | Amargo |
| 113  | Compound XX  | Amargo |
| 114  | Compound XXI   | Dulce  |
| 115^ | Compound XXII  | Dulce  |
| 116  | Compound XXIII   | Dulce  |
| 117^ | Compound XXIV  | Dulce  |
| 118^ | Compound XXVII   | Dulce  |
| 119  | Compound XXIX  | Amargo |
| 120^ | Compound XXX   | Dulce  |
| 121  | Compound XXXI  | Dulce  |
| 122  | Compound XXXII   | Dulce  |
| 123^ | Compound XL  | Dulce  |
| 124^ | Compound XLI   | Dulce  |
| 125  | Compound XLII  | Dulce  |
| 126  | Compound XLIII   | Dulce  |
| 127^ | Compound XLIV  | Amargo |
| 128  | Compound XLVII   | Dulce  |
| 129  | Compound XLIX  | Dulce  |
| 130  | Compound LIV   | Dulce  |
| 131  | Compound LIX   | Dulce  |
| 132  | Compound LXV   | Dulce  |
| 133^ | Compound LXVI  | Dulce  |
| 134  | Compound LXXII   | Amargo |
| 135^ | Compound LXXIII  | Amargo |
| 136^ | Compound LXXIV   | Amargo |
| 137  | Compound LXXV  | Amargo |
| 138  | Compound LXXVI   | Amargo |

---

|      |  |        |
|------|--|--------|
| 139^ | Compound LXXVIII                                 | Dulce  |
| 140^ | Compound LXXIX                                   | Dulce  |
| 141  | Compound LXXX                                    | Dulce  |
| 142  | Compound LXXXV                                   | Dulce  |
| 143  | Compound LXXXIX                                  | Dulce  |
| 144  | Compound XC                                      | Dulce  |
| 145  | Compound XCI                                     | Dulce  |
| 146  | Compound XCV                                     | Dulce  |
| 147  | Compound XCVI                                    | Dulce  |
| 148^ | Compound XCVII                                   | Amargo |
| 149  | Nitrobenzene                                     | Dulce  |
| 150  | m-Nitrobenzene                                   | Amargo |
| 151  | Phenylthiourea                                   | Amargo |
| 152  | Dimethyl urea                                    | Dulce  |
| 153^ | Urea   | Amargo |
| 154  | Resorcinol                                       | Dulce  |
| 155  | Phloroglucinol                                   | Dulce  |
| 156  | Glycol   | Dulce  |
| 157  | 1,4-Anhydro-ribitol                              | Dulce  |
| 158^ | 1,4-Anhydro-mannitol                             | Dulce  |
| 159  | 1,5-Anhydro-mannitol                             | Dulce  |
| 160^ | Viboquercitol                                    | Dulce  |
| 161  | Allo-inositol                                    | Dulce  |
| 162^ | Dihydroxyacetone                                 | Dulce  |
| 163  | Monohydroxyacetone                               | Dulce  |
| 164^ | Methyl dihydroxyacetone                          | Dulce  |
| 165  | Stachyose  | Dulce  |
| 166  | 4-O-Methyl sucrose                               | Dulce  |
| 167  | 6-O-Methyl sucrose                               | Dulce  |
| 168^ | 6,6'-di-O-Methyl sucrose                         | Dulce  |
| 169  | 4,6'-di-O-Methyl sucrose                         | Dulce  |
| 170  | 4,6-di-O-Methyl sucrose                          | Dulce  |
| 171  | 1',6'-di-O-Methyl sucrose                        | Dulce  |
| 172  | Methyl 2-deoxy- $\alpha$ -D-gluco-pyranoside     | Dulce  |
| 173  | Methyl 3-deoxy- $\alpha$ -D-gluco-pyranoside     | Dulce  |
| 174^ | Methyl 4-deoxy- $\alpha$ -D-gluco-pyranoside     | Dulce  |
| 175^ | Methyl 6-deoxy- $\alpha$ -D-gluco-pyranoside     | Dulce  |
| 176  | Methyl- $\alpha$ -D-2,6-dideoxy-gluco-pyranoside | Amargo |
| 177  | Methyl- $\alpha$ -D-3,6-dideoxy-gluco-pyranoside | Amargo |
| 178^ | Methyl- $\alpha$ -D-4,6-dideoxy-gluco-pyranoside | Amargo |
| 179  | 6-Chloro-6-deoxy-D-fructofuranose                | Dulce  |
| 180  | 1,6-dichloro-1,6-deoxy-D-fructofuranose          | Dulce  |
| 181  | Methyl glucopyranoside                           | Dulce  |
| 182  | 4-Chloro-galactosucrose                          | Dulce  |
| 183  | 6'-Chloro-sucrose                                | Dulce  |
| 184^ | 1'-Chloro-sucrose                                | Dulce  |
| 185  | 1',6'-Dichloro-sucrose                           | Dulce  |
| 186  | 1',4-Dichloro-galactosucrose                     | Dulce  |
| 187  | 1,6-Di-S-1,6-dithio-D-fructofuranose             | Dulce  |
| 188  | Alanine  | Dulce  |

---

---

|      |  |        |
|------|--|--------|
| 189  | Asparagine                             | Dulce  |
| 190  | glutamine                              | Dulce  |
| 191^ | Glycine                                | Dulce  |
| 192^ | Histidine                              | Dulce  |
| 193  | Methionine                             | Dulce  |
| 194  | Hydroxy proline                        | Dulce  |
| 195  | Serine                                 | Dulce  |
| 196^ | Threonine                              | Dulce  |
| 197  | L-Aspartyl-aminomalonic acid diester 1 | Dulce  |
| 198  | L-Aspartyl-aminomalonic acid diester 2 | Dulce  |
| 199  | L-Aspartyl-aminomalonic acid diester 3 | Dulce  |
| 200  | N-(L-aspartyl)-1,1-diaminoalkane 1     | Dulce  |
| 201^ | N-(L-aspartyl)-1,1-diaminoalkane 2     | Dulce  |
| 202  | N-(L-aspartyl)-1,1-diaminoalkane 3     | Dulce  |
| 203^ | N-(L-aspartyl)-1,1-diaminoalkane 4     | Dulce  |
| 204  | N-(L-aspartyl)-1,1-diaminoalkane 5     | Dulce  |
| 205^ | N-(L-aspartyl)-1,1-diaminoalkane 6     | Dulce  |
| 206  | N-(L-aspartyl)-1,1-diaminoalkane 7     | Dulce  |
| 207  | N-(L-aspartyl)-1,1-diaminoalkane 8     | Dulce  |
| 208  | N-(L-aspartyl)-1,1-diaminoalkane 9     | Dulce  |
| 209^ | Super aspartame                        | Dulce  |
| 210  | Sucrononic acid                        | Dulce  |
| 211  | Cyanoarylurea aspartame                | Dulce  |
| 212^ | Aspartic acid fenchyl ester            | Dulce  |
| 213  | Glycine analog of cyanosuosan          | Dulce  |
| 214  | Glycine analog of suosan               | Dulce  |
| 215  | Nitroaniline (2-amino-4-nitrobenzene)  | Dulce  |
| 216^ | Methoxy-nitroaniline                   | Dulce  |
| 217  | Ethoxy-nitroaniline                    | Dulce  |
| 218^ | Butoxy-nitroaniline                    | Dulce  |
| 219^ | Fluoro-nitroaniline                    | Dulce  |
| 220  | Chloro-nitroaniline                    | Dulce  |
| 221^ | Bromo-nitroaniline                     | Dulce  |
| 222^ | Iodo-nitroaniline                      | Dulce  |
| 223  | Methyl-nitroaniline                    | Dulce  |
| 224  | Allyl oxy-nitroaniline                 | Dulce  |
| 225^ | iso Propoxy-nitroaniline               | Dulce  |
| 226  | Hydroxy-nitroaniline                   | Dulce  |
| 227^ | Saccharine derivative 1                | Amargo |
| 228  | Saccharine derivative 2                | Amargo |
| 229  | Saccharine derivative 3                | Amargo |
| 230  | Saccharine derivative 4                | Amargo |
| 231  | Saccharine derivative 5                | Dulce  |
| 232^ | Saccharine derivative 13               | Dulce  |
| 233  | Saccharine derivative 14               | Amargo |
| 234  | Saccharine derivative 15               | Dulce  |
| 235^ | Saccharine derivative 16               | Dulce  |
| 236^ | Saccharine derivative 20               | Amargo |
| 237  | Saccharine derivative 22               | Amargo |
| 238  | Saccharine derivative 23               | Dulce  |

---

---

|      |                                       |        |
|------|---------------------------------------|--------|
| 239  | Saccharine derivative 25              | Dulce  |
| 240  | Saccharine derivative 27              | Amargo |
| 241  | Thiophenesaccharine                   | Dulce  |
| 242  | anti-Furfuraldehyde oxime             | Dulce  |
| 243  | 5-Benzyl-2-Furfuraldehyde oxime       | Dulce  |
| 244  | 1-Cycloheptene-1-carboxaldehyde oxime | Amargo |
| 245^ | anti-Cyclohexane carboxaldehyde       | Amargo |
| 246^ | Dihydroquercetin 3-acetate            | Dulce  |
| 247  | Hesperitin                            | Dulce  |
| 248  | Neoastilbin                           | Dulce  |
| 249^ | meta-Tolylurea                        | Amargo |
| 250^ | para-Tolylurea                        | Dulce  |
| 251  | 4-Propoxyphenyl urea                  | Dulce  |
| 252  | Suosan derivative 1 (suosan)          | Dulce  |
| 253^ | Suosan derivative 11                  | Dulce  |
| 254  | Suosan derivative 12                  | Amargo |
| 255^ | Suosan derivative 13                  | Dulce  |
| 256^ | Suosan derivative 16                  | Dulce  |
| 257  | Pyrrolidine                           | Amargo |
| 258  | Pyrrole                               | Amargo |
| 259  | Imidazole                             | Amargo |
| 260  | Pyrazole                              | Amargo |
| 261^ | Piperidine                            | Amargo |
| 262  | Piperazine                            | Amargo |
| 263  | Purine                                | Amargo |
| 264^ | Pyridazine                            | Amargo |
| 265  | Pyrazine                              | Amargo |
| 266  | Pyridine                              | Amargo |
| 267^ | Pyrimidine                            | Amargo |
| 268  | Strichnine                            | Amargo |
| 269  | Brucine                               | Amargo |
| 270^ | Solanine                              | Amargo |
| 271  | Diterpene derivative 1                | Amargo |
| 272  | Diterpene derivative 2                | Amargo |
| 273  | Diterpene derivative 3                | Amargo |
| 274  | Diterpene derivative 4                | Amargo |
| 275  | Denatonium chloride                   | Amargo |
| 276  | Denatonium chloride derivative 1      | Amargo |
| 277  | Denatonium chloride derivative 2      | Amargo |
| 278  | Denatonium chloride derivative 3      | Amargo |
| 279  | Denatonium chloride derivative 4      | Amargo |
| 280  | Denatonium chloride derivative 5      | Amargo |
| 281  | Denatonium chloride derivative 6      | Amargo |
| 282  | Denatonium chloride derivative 7      | Amargo |
| 283  | Denatonium chloride derivative 8      | Amargo |
| 284  | Denatonium chloride derivative 9      | Amargo |
| 285^ | Denatonium chloride derivative 10     | Amargo |
| 286  | Denatonium chloride derivative 11     | Amargo |
| 287^ | Denatonium chloride derivative 12     | Amargo |
| 288^ | Denatonium chloride derivative 13     | Amargo |

---

|      |  |        |
|------|--|--------|
| 289  | Methyl-β-D-Fructopyranose  | Dulce  |
| 290^ | 5-deoxy-D-threo-Hexulose   | Dulce  |
| 291  | Oxathiazinon dioxide derivative 1                                      | Dulce  |
| 292^ | Oxathiazinon dioxide derivative 2                                      | Dulce  |
| 293  | Oxathiazinon dioxide derivative 3                                      | Dulce  |
| 294  | Oxathiazinon dioxide derivative 4                                      | Dulce  |
| 295^ | Oxathiazinon dioxide derivative 5                                      | Dulce  |
| 296^ | Oxathiazinon dioxide derivative 6                                      | Dulce  |
| 297  | 3-amino-4-chlorobenzoic acid   | Amargo |
| 298^ | 2-aminobenzoic acid  | Dulce  |
| 299^ | 2-nitroaniline   | Amargo |
| 300  | Germacrolide   | Amargo |
| 301  | 3-anilino-2-styryl-3H-naphthol[1, 2-d]imidazole-5-sulfonate (SSN)      | Dulce  |
| 302  | Perillartine   | Dulce  |
| 303^ | 1',4',6'-3Cl-sucrose   | Dulce  |
| 304^ | 1',4'-2Cl-sucrose  | Dulce  |
| 305  | 4'-Br-4,1',6'-3Cl-sucrose  | Dulce  |
| 306^ | 4'-F-4,1',6'-3Cl-sucrose   | Dulce  |
| 307  | 4,1',4',6'-4Br-sucrose   | Dulce  |
| 308  | 4,1',4',6'-4Cl-sucrose   | Dulce  |
| 309^ | 4,1',4'-3Cl-sucrose  | Dulce  |
| 310  | 4,1',6'-3Br-sucrose  | Dulce  |
| 311  | 4,1',6'-3F-sucrose   | Dulce  |
| 312  | 4,1',6'-3I-sucrose   | Dulce  |
| 313  | 4,4',6'-3Cl-sucrose  | Dulce  |
| 314  | 4,6,1',6'-4Cl-sucrose  | Dulce  |
| 315^ | 4-F-1',4',6'-3Br-sucrose   | Dulce  |
| 316^ | 4-F-1',4',6'-3Cl-sucrose   | Dulce  |
| 317  | 4-O-n-propyl glycosides DHC  | Dulce  |
| 318  | 4β,10α-dimethyl-1,2,3,4,5,10-hexahydrofluorene-4α,6α-dicarboxylic acid | Dulce  |
| 319  | 6,1',6'-3Cl-sucrose  | Dulce  |
| 320  | 6-chloro-tryptophan  | Dulce  |
| 321  | Liquiritin   | Dulce  |
| 322  | N-Acetyl-D-glucosamine   | Dulce  |
| 323  | Periandrin I   | Dulce  |
| 324  | Periandrin II  | Dulce  |
| 325  | Periandrin III   | Dulce  |
| 326  | Periandrin IV  | Dulce  |
| 327  | Abrusoside B   | Dulce  |
| 328^ | Abrusoside C   | Dulce  |
| 329^ | Abrusoside D   | Dulce  |
| 330  | Acesulfame potassium_CH3CH2CH2_CH3                                     | Dulce  |
| 331  | Acesulfame potassium_CH3CH2_H  | Dulce  |
| 332^ | Acesulfame potassium_H_H   | Dulce  |
| 333  | Albiziasaponin B   | Dulce  |
| 334  | Arabinose  | Dulce  |
| 335  | Asp 1  | Dulce  |
| 336^ | Asp 2  | Dulce  |

---

|      |        |       |
|------|--------|-------|
| 337  | Asp 3  | Dulce |
| 338  | Asp 4  | Dulce |
| 339^ | Asp 5  | Dulce |
| 340  | Asp 6  | Dulce |
| 341  | Asp 7  | Dulce |
| 342^ | Asp 8  | Dulce |
| 343  | Asp 9  | Dulce |
| 344  | Asp 10 | Dulce |
| 345^ | Asp 11 | Dulce |
| 346  | Asp 14 | Dulce |
| 347^ | Asp 15 | Dulce |
| 348^ | Asp 16 | Dulce |
| 349^ | Asp 17 | Dulce |
| 350  | Asp 18 | Dulce |
| 351  | Asp 20 | Dulce |
| 352  | Asp 21 | Dulce |
| 353  | Asp 22 | Dulce |
| 354  | Asp 23 | Dulce |
| 355  | Asp 24 | Dulce |
| 356  | Asp 52 | Dulce |
| 357^ | Asp 53 | Dulce |
| 358  | Asp 55 | Dulce |
| 359^ | Asp 56 | Dulce |
| 360  | Asp 58 | Dulce |
| 361^ | Asp 60 | Dulce |
| 362  | Asp 61 | Dulce |
| 363^ | Asp 62 | Dulce |
| 364  | Asp 63 | Dulce |
| 365  | Asp 64 | Dulce |
| 366  | Asp 65 | Dulce |
| 367  | Asp 66 | Dulce |
| 368  | Asp 67 | Dulce |
| 369  | Asp 68 | Dulce |
| 370  | Asp 69 | Dulce |
| 371^ | Asp 70 | Dulce |
| 372^ | Asp 71 | Dulce |
| 373^ | Asp 72 | Dulce |
| 374  | Asp 73 | Dulce |
| 375^ | Asp 74 | Dulce |
| 376  | Asp 75 | Dulce |
| 377  | Asp 76 | Dulce |
| 378  | Asp 77 | Dulce |
| 379  | Asp 78 | Dulce |
| 380^ | Asp 79 | Dulce |
| 381  | Asp 80 | Dulce |
| 382  | Asp 81 | Dulce |
| 383  | Asp 82 | Dulce |
| 384  | Asp 83 | Dulce |
| 385^ | Asp 84 | Dulce |
| 386^ | Asp 85 | Dulce |

---

---

|      |         |       |
|------|---------|-------|
| 387^ | Asp 86  | Dulce |
| 388  | Asp 87  | Dulce |
| 389^ | Asp 88  | Dulce |
| 390^ | Asp 89  | Dulce |
| 391  | Asp 90  | Dulce |
| 392  | Asp 91  | Dulce |
| 393^ | Asp 92  | Dulce |
| 394  | Asp 93  | Dulce |
| 395  | Asp 94  | Dulce |
| 396  | Asp 95  | Dulce |
| 397  | Asp 96  | Dulce |
| 398  | Asp 97  | Dulce |
| 399^ | Asp 98  | Dulce |
| 400  | Asp 99  | Dulce |
| 401  | Asp 100 | Dulce |
| 402  | Asp 101 | Dulce |
| 403  | Asp 102 | Dulce |
| 404  | Asp 103 | Dulce |
| 405  | Asp 104 | Dulce |
| 406  | Asp 105 | Dulce |
| 407  | Asp 106 | Dulce |
| 408  | Asp 107 | Dulce |
| 409  | Asp 108 | Dulce |
| 410  | Asp 109 | Dulce |
| 411  | Asp 110 | Dulce |
| 412  | Asp 112 | Dulce |
| 413^ | Asp 122 | Dulce |
| 414^ | Asp 123 | Dulce |
| 415^ | Asp 124 | Dulce |
| 416  | Asp 125 | Dulce |
| 417  | Asp 126 | Dulce |
| 418  | Asp 128 | Dulce |
| 419  | Asp 129 | Dulce |
| 420  | Asp 130 | Dulce |
| 421^ | Asp 131 | Dulce |
| 422  | Asp 132 | Dulce |
| 423^ | Asp 134 | Dulce |
| 424^ | Asp 135 | Dulce |
| 425^ | Asp 136 | Dulce |
| 426  | Asp 137 | Dulce |
| 427  | Asp 138 | Dulce |
| 428^ | Asp 139 | Dulce |
| 429  | Asp 140 | Dulce |
| 430  | Asp 141 | Dulce |
| 431^ | Asp 142 | Dulce |
| 432  | Asp 143 | Dulce |
| 433  | Asp 153 | Dulce |
| 434^ | Asp 154 | Dulce |
| 435  | Asp 155 | Dulce |
| 436^ | Asp 156 | Dulce |

---

---

|      |                          |       |
|------|--------------------------|-------|
| 437  | Asp 157                  | Dulce |
| 438  | Asp 159                  | Dulce |
| 439  | Asp 160                  | Dulce |
| 440  | Asp 161                  | Dulce |
| 441  | Asp 163                  | Dulce |
| 442^ | Asp 164                  | Dulce |
| 443  | Asp 165                  | Dulce |
| 444  | Asp 166                  | Dulce |
| 445  | Asp 167                  | Dulce |
| 446  | Asp 168                  | Dulce |
| 447  | Asp 169                  | Dulce |
| 448^ | Asp 170                  | Dulce |
| 449  | Asp 171                  | Dulce |
| 450  | Asp 172                  | Dulce |
| 451  | Asp 173                  | Dulce |
| 452  | Asp 181                  | Dulce |
| 453  | Asp 182                  | Dulce |
| 454  | Asp 183                  | Dulce |
| 455  | Asp 184                  | Dulce |
| 456^ | Asp 185                  | Dulce |
| 457  | Asp 186                  | Dulce |
| 458  | Asp 187                  | Dulce |
| 459^ | Asp 188                  | Dulce |
| 460^ | Asp 189                  | Dulce |
| 461  | Asp 190                  | Dulce |
| 462^ | Asp 191                  | Dulce |
| 463  | Asp 192                  | Dulce |
| 464  | Asp 193                  | Dulce |
| 465  | Asp 194                  | Dulce |
| 466  | Asp 197                  | Dulce |
| 467^ | Asp 198                  | Dulce |
| 468^ | Asp 204                  | Dulce |
| 469  | Asp 207                  | Dulce |
| 470  | Asp 208                  | Dulce |
| 471^ | Asp 211                  | Dulce |
| 472  | Asp 212                  | Dulce |
| 473  | Asp 213                  | Dulce |
| 474  | Asp 214                  | Dulce |
| 475  | Asp 215                  | Dulce |
| 476^ | Baiyunoside              | Dulce |
| 477  | Benzenepropanoic acid    | Dulce |
| 478^ | Bernardame               | Dulce |
| 479  | cyclamate                | Dulce |
| 480  | Cyclocarioside A         | Dulce |
| 481  | Gaudichaudioside A       | Dulce |
| 482^ | Hesperinin glucoside DHC | Dulce |
| 483  | Homoeohesperidin DHC     | Dulce |
| 484  | Isovanillic 18           | Dulce |
| 485  | Isovanillic 24           | Dulce |
| 486  | isovanillic 25           | Dulce |

---

---

|                  |                      |       |
|------------------|----------------------|-------|
| 487              | Leucrose             | Dulce |
| 488              | Lugdunane            | Dulce |
| 489              | Mogroside II         | Dulce |
| 490              | Mogroside IV         | Dulce |
| 491              | Mogroside V          | Dulce |
| 492              | Naringin DHC         | Dulce |
| 493 <sup>^</sup> | Neohesperidin DHC    | Dulce |
| 494              | Osladin              | Dulce |
| 495              | Polypodoside A       | Dulce |
| 496              | Prunin DHC           | Dulce |
| 497 <sup>^</sup> | Pterocaryoside A     | Dulce |
| 498              | Pterocaryoside B     | Dulce |
| 499 <sup>^</sup> | Rebaudioside F       | Dulce |
| 500 <sup>^</sup> | Rubusoside           | Dulce |
| 501              | Saccharin            | Dulce |
| 502              | Selligueain A        | Dulce |
| 503 <sup>^</sup> | Siamenoside I        | Dulce |
| 504              | Sorbose              | Dulce |
| 505              | Telosmoside A        | Dulce |
| 506              | trans-anethole       | Dulce |
| 507              | trans-cinnamaldehyde | Dulce |
| 508              | Xylobiose            | Dulce |

---

<sup>^</sup> predicción

**Tabla 9A.** Nombres químicos y respuesta experimental de los 566 compuestos dulces e insípidos

| No. | Nombre                                      | Clase    |
|-----|---|----------|
| 1   | Sucrose                                     | Dulce    |
| 2   | Sucralose                                   | Dulce    |
| 3^  | Alitame                                     | Dulce    |
| 4   | Aspartame                                   | Dulce    |
| 5   | Tagatose                                    | Dulce    |
| 6^  | Maltitol                                    | Dulce    |
| 7   | Isomaltulose (Palatinose)                   | Dulce    |
| 8   | Trehalose                                   | Dulce    |
| 9   | Neohesperidine dihydrochalcone              | Dulce    |
| 10  | Neotame                                     | Dulce    |
| 11  | Isomalt (Palatinit)                         | Dulce    |
| 12  | Sorbitol                                    | Dulce    |
| 13  | Erythritol                                  | Dulce    |
| 14^ | Xylitol                                     | Dulce    |
| 15^ | Stevioside                                  | Dulce    |
| 16^ | Steviolbioside                              | Dulce    |
| 17  | Rebaudioside A                              | Dulce    |
| 18  | Rebaudioside B                              | Dulce    |
| 19  | Rebaudioside C                              | Dulce    |
| 20^ | Rebaudioside D                              | Dulce    |
| 21  | Rebaudioside E                              | Dulce    |
| 22^ | Dulcoside A                                 | Dulce    |
| 23  | Glycyrrhizin                                | Dulce    |
| 24^ | Hernandulcin                                | Dulce    |
| 25  | Fructose                                    | Dulce    |
| 26  | Glucose                                     | Dulce    |
| 27  | Xylose                                      | Dulce    |
| 28  | Maltose                                     | Dulce    |
| 29  | Lactulose                                   | Dulce    |
| 30  | Lactosucrose                                | Dulce    |
| 31  | Raffinose                                   | Dulce    |
| 32  | Rhamnose                                    | Dulce    |
| 33^ | Glycerol                                    | Dulce    |
| 34^ | Saccharin derivative 1                      | Dulce    |
| 35  | Saccharin derivative 2                      | Dulce    |
| 36  | Saccharin derivative 3                      | Insípido |
| 37  | 3-nitro-o-toluidine                         | Insípido |
| 38^ | 3-nitro-p-toluidine                         | Insípido |
| 39  | 2-amino-4-nitro-propoxybenzene              | Dulce    |
| 40  | 2-nitro-4-amino-propoxybenzene              | Insípido |
| 41^ | Dulcin                                      | Dulce    |
| 42  | o-ethoxyphenylurea                          | Insípido |
| 43  | Dihydroquercetin 3-acetate 4'-(methylether) | Dulce    |
| 44  | Flavan                                      | Dulce    |
| 45  | Compound 5                                  | Dulce    |
| 46  | 1,3-benzodioxane                            | Dulce    |

|     |  |          |
|-----|--|----------|
| 47  | 1,4-benzodioxane   | Dulce    |
| 48^ | Bibenzyl derivative  | Dulce    |
| 49^ | 1,4-benzodioxin  | Insípido |
| 50  | Isoflavan or 3-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1benzopyran      | Dulce    |
| 51  | 2-(3-Hydroxy-4-methoxyphenyl)-4,5-dihydro-1,3-benzodioxepine               | Insípido |
| 52^ | 3-(3-Hydroxy-4-methoxyphenyl)-1,5-dihydro-2,4-benzodioxepine               | Insípido |
| 53  | 2-(3-hydroxy-4-methoxyphenyl)-1,3-benzodioxole                             | Dulce    |
| 54  | Compound 23  | Insípido |
| 55  | Compound 24  | Insípido |
| 56^ | Compound 25  | Insípido |
| 57  | 2-(3-Hydroxy-4-methoxyphenyl)-4-H-1,3-benzodithiine                        | Dulce    |
| 58^ | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzoxathiine                         | Dulce    |
| 59  | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzodithiole                         | Dulce    |
| 60  | 2-(3-Hydroxy-4-methoxyphenyl)-1,3-benzoxathiole                            | Dulce    |
| 61  | 7,8-Dimethyl-3-(3-hydroxy-4-methoxyphenyl)-1,5-dihydro-2,4-benzodithiepine | Insípido |
| 62  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiine                | Dulce    |
| 63  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzodithiine                | Dulce    |
| 64^ | 2-(3-Hydroxy-4-methylthiophenyl)-2,3-dihydro-1,4-benzoxathiine             | Insípido |
| 65^ | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol            | Dulce    |
| 66  | 2-(3-Hydroxy-4-methylthiophenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol         | Insípido |
| 67  | 2-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran              | Dulce    |
| 68^ | 3-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran              | Insípido |
| 69^ | 2-(3-Mercapto-4-methoxyphenyl)-4H-3,1-benzoxathiine                        | Dulce    |
| 70^ | Sulfone  | Dulce    |
| 71  | Haematoxylin   | Dulce    |
| 72^ | 9-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,10-diol                | Dulce    |
| 73  | 10-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,9-diol                | Insípido |
| 74  | 3,9,10-Trimethoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-4,6a-diol         | Insípido |
| 75^ | 4,9,10-Trimethoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-3,6a-diol         | Insípido |
| 76  | Compound 8   | Dulce    |
| 77  | Compound 17  | Insípido |
| 78^ | Compound 18  | Dulce    |
| 79  | Compound 19  | Insípido |
| 80  | Compound 20  | Dulce    |
| 81^ | Compound 21  | Dulce    |
| 82  | Compound 25  | Insípido |
| 83^ | Compound 26  | Dulce    |
| 84  | Compound 27  | Insípido |
| 85^ | Compound 28  | Insípido |

|      |   |          |
|------|---|----------|
| 86   | Compound 29   | Insípido |
| 87^  | Compound 30   | Insípido |
| 88^  | 2-(4-methoxybenzoyl) benzoic acid   | Dulce    |
| 89^  | Compound 2  | Insípido |
| 90^  | Compound 3  | Dulce    |
| 91^  | Compound 4  | Dulce    |
| 92   | Compound 5  | Insípido |
| 93   | Compound 6  | Insípido |
| 94^  | Compound 7  | Dulce    |
| 95   | Compound 8  | Insípido |
| 96   | Compound 9  | Dulce    |
| 97   | Compound 11   | Insípido |
| 98   | Compound 12   | Dulce    |
| 99   | Compound 13   | Insípido |
| 100^ | 2-(3,4-Dihydroxybenzoyl)benzoic Acid  | Dulce    |
| 101  | 2-(3-Hydroxy-4-methoxybenzoyl)benzoic Acid  | Dulce    |
| 102^ | 2-(4-Hydroxy-3-methoxybenzoyl)benzoic Acid  | Insípido |
| 103^ | 2-(3,4-Dimethoxybenzoyl)benzoic Acid  | Insípido |
| 104  | 2-(4-Hydroxy-2-methoxybenzoyl)benzoic Acid  | Insípido |
| 105  | Compound 20   | Insípido |
| 106  | Compound 21   | Insípido |
| 107  | Compound 22   | Insípido |
| 108^ | 2-(4-Methoxyphenylthio) benzoic Acid S-Oxide  | Dulce    |
| 109  | Compound 24   | Insípido |
| 110  | Compound 25   | Insípido |
| 111^ | 2-Hydroxymethylphenyl-4-methoxyphenylmethanol                                       | Insípido |
| 112^ | Phthalic Acid Mono(4-methoxyphenyl) Ester   | Dulce    |
| 113^ | 3-(4-Methoxyphenyl)benzo[2]furan-1-one  | Insípido |
| 114  | Compound 29   | Insípido |
| 115  | 3-Methoxy-1,3,5-estratrien-4-ol   | Dulce    |
| 116  | Substituted pterocarpan   | Insípido |
| 117  | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-methoxy-6'-hydroxy-[1H]-2',3'-dihydroindene)] | Dulce    |
| 118  | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-hydroxy-6'-methoxy-[1H]-2',3'-dihydroindene)] | Insípido |
| 119^ | Spiro[(4-H-1,3-benzodithian)-2,1'-(4'-hydroxy-5'-methoxy-[1H]-2',3'-dihydroindene)] | Insípido |
| 120  | 2-Phenoxymethyl-2H-7,8-dihydrobenzopyran  | Insípido |
| 121^ | 2-Methyl-2-(3-hydroxy-4-methoxyphenyl)-1,4-dihydrobenzodioxane                      | Insípido |
| 122^ | Compound 4  | Dulce    |
| 123  | Compound 6  | Dulce    |
| 124  | Compound 9  | Dulce    |
| 125  | 3',4'-dihydroxyphenyl-3,1-benzoxathiane   | Dulce    |
| 126^ | 2-[3-(4-nitrophenoxycarboxy)phenyl]-3,1-benzoxathiane                               | Dulce    |
| 127^ | 2-[3-(4-nitrophenylcarbonylmethoxy)-4-methoxyphenyl]-3,1-benzoxathiane              | Dulce    |
| 128^ | 2-[3-(4-nitrophenylcarbamoxyloxy-4-methoxy)phenyl]-3,1-benzodioxane                 | Dulce    |
| 129  | Compound 32   | Dulce    |
| 130  | 3-(3-Hydroxy-4-methoxyphenyl)-1-(3-carboxyphenyl)propan-                            | Dulce    |

---

|      |  |          |
|------|--|----------|
|      | 1-one  |          |
| 131  | Compound 34                                      | Dulce    |
| 132  | 2-(3-hydroxy-4-methoxyphenoxymethyl)benzoic acid | Dulce    |
| 133^ | Compound 42                                      | Dulce    |
| 134  | Dihydroquercetin 4'-(methyl ether)               | Dulce    |
| 135  | Dihydroquercetin                                 | Insípido |
| 136  | Compound II or Hydrangenol                       | Insípido |
| 137^ | Compound IV                                      | Insípido |
| 138  | Compound V                                       | Dulce    |
| 139^ | Compound VI                                      | Insípido |
| 140  | Compound VII                                     | Dulce    |
| 141^ | Compound VIII                                    | Insípido |
| 142  | Compound IX                                      | Insípido |
| 143^ | Compound X                                       | Insípido |
| 144  | Compound XI                                      | Insípido |
| 145  | Compound XII                                     | Insípido |
| 146  | Compound XIV                                     | Dulce    |
| 147  | Compound XV                                      | Insípido |
| 148^ | Compound XVI                                     | Insípido |
| 149  | Compound XVII                                    | Insípido |
| 150  | Compound XXI                                     | Dulce    |
| 151  | Compound XXII                                    | Dulce    |
| 152  | Compound XXIII                                   | Dulce    |
| 153  | Compound XXIV                                    | Dulce    |
| 154^ | Compound XXV                                     | Insípido |
| 155  | Compound XXVI                                    | Insípido |
| 156  | Compound XXVII                                   | Dulce    |
| 157  | Compound XXVIII                                  | Insípido |
| 158  | Compound XXX                                     | Dulce    |
| 159^ | Compound XXXI                                    | Dulce    |
| 160  | Guaiacol   | Insípido |
| 161  | Compound XXXII                                   | Dulce    |
| 162  | Compound XXXIII                                  | Insípido |
| 163^ | Compound XXXIV                                   | Insípido |
| 164  | Compound XXXV                                    | Insípido |
| 165  | Compound XXXVII                                  | Insípido |
| 166^ | Compound XXXVIII                                 | Insípido |
| 167  | Compound XXXIX                                   | Insípido |
| 168  | Compound XL                                      | Dulce    |
| 169^ | Compound XLI                                     | Dulce    |
| 170  | Compound XLII                                    | Dulce    |
| 171  | Compound XLIII                                   | Dulce    |
| 172  | Compound XLV                                     | Insípido |
| 173  | Compound XLVI                                    | Insípido |
| 174  | Compound XLVII                                   | Dulce    |
| 175  | Compound XLVIII                                  | Insípido |
| 176^ | Compound XLIX                                    | Dulce    |
| 177^ | Compound L                                       | Insípido |
| 178  | Compound LI                                      | Insípido |
| 179  | Compound LII                                     | Insípido |

---

---

|      |                      |          |
|------|----------------------|----------|
| 180  | Compound LIII        | Insípido |
| 181  | Compound LIV         | Dulce    |
| 182^ | Compound LV          | Insípido |
| 183  | Compound LVI         | Insípido |
| 184  | Compound LVII        | Insípido |
| 185  | Compound LVIII       | Insípido |
| 186^ | Compound LIX         | Dulce    |
| 187  | Compound LX          | Insípido |
| 188  | Compound LXI         | Insípido |
| 189  | Compound LXII        | Insípido |
| 190^ | Compound LXIII       | Insípido |
| 191  | Compound LXIV        | Insípido |
| 192  | Compound LXV         | Dulce    |
| 193  | Compound LXVI        | Dulce    |
| 194^ | Compound LXVII       | Insípido |
| 195^ | Compound LXVIII      | Insípido |
| 196  | Compound LXIX        | Insípido |
| 197  | Compound LXX         | Insípido |
| 198  | Compound LXXI        | Insípido |
| 199  | Compound LXXVII      | Insípido |
| 200^ | Compound LXXVIII     | Dulce    |
| 201  | Compound LXXIX       | Dulce    |
| 202^ | Compound LXXX        | Dulce    |
| 203^ | Compound LXXXI       | Insípido |
| 204  | Compound LXXXII      | Insípido |
| 205  | Compound LXXXIII     | Insípido |
| 206  | Compound LXXXIV      | Insípido |
| 207^ | Compound LXXXV       | Dulce    |
| 208  | Compound LXXXVIII    | Insípido |
| 209^ | Compound LXXXIX      | Dulce    |
| 210^ | Compound XC          | Dulce    |
| 211^ | Compound XCI         | Dulce    |
| 212  | Compound XCII        | Insípido |
| 213^ | Compound XCIII       | Insípido |
| 214  | Compound XCV         | Dulce    |
| 215  | Compound XCVI        | Dulce    |
| 216  | Compound XCVIII      | Insípido |
| 217^ | Compound CX          | Insípido |
| 218  | Compound C           | Insípido |
| 219  | Compound CI          | Insípido |
| 220  | Compound CII         | Insípido |
| 221  | Nitrobenzene         | Dulce    |
| 222  | Dimethyl urea        | Dulce    |
| 223^ | Resorcinol           | Dulce    |
| 224^ | Phloroglucinol       | Dulce    |
| 225  | Glycol               | Dulce    |
| 226  | 1,4-Anhydro-ribitol  | Dulce    |
| 227^ | 1,4-Anhydro-mannitol | Dulce    |
| 228  | 1,5-Anhydro-mannitol | Dulce    |
| 229  | Viboquercitol        | Dulce    |

---

|      |  |         |
|------|--|---------|
| 230  | Allo-inositol                                | Dulce   |
| 231  | Dihydroxyacetone                             | Dulce   |
| 232  | Monohydroxyacetone                           | Dulce   |
| 233  | Methyl dihydroxyacetone                      | Dulce   |
| 234  | Stachyose                                    | Dulce   |
| 235  | 4-O-Methyl sucrose                           | Dulce   |
| 236  | 6-O-Methyl sucrose                           | Dulce   |
| 237  | 6,6'-di-O-Methyl sucrose                     | Dulce   |
| 238^ | 4,6'-di-O-Methyl sucrose                     | Dulce   |
| 239  | 4,6-di-O-Methyl sucrose                      | Dulce   |
| 240  | 1',6'-di-O-Methyl sucrose                    | Dulce   |
| 241^ | Methyl 2-deoxy- $\alpha$ -D-gluco-pyranoside | Dulce   |
| 242  | Methyl 3-deoxy- $\alpha$ -D-gluco-pyranoside | Dulce   |
| 243  | Methyl 4-deoxy- $\alpha$ -D-gluco-pyranoside | Dulce   |
| 244  | Methyl 6-deoxy- $\alpha$ -D-gluco-pyranoside | Dulce   |
| 245^ | 6-Chloro-6-deoxy-D-galactose                 | Inspido |
| 246  | 6-Chloro-6-deoxy-D-fructofuranose            | Dulce   |
| 247  | 1,6-dichloro-1,6-deoxy-D-fructofuranose      | Dulce   |
| 248  | Methyl $\alpha$ -D-glucopyranoside           | Dulce   |
| 249  | 6'-Chloro-sucrose                            | Dulce   |
| 250^ | 1'-Chloro-sucrose                            | Dulce   |
| 251  | 1',6'-Dichloro-sucrose                       | Dulce   |
| 252^ | 1,6-Di-S-1,6-dithio-D-fructofuranose         | Dulce   |
| 253  | Alanine                                      | Dulce   |
| 254  | Arginine                                     | Dulce   |
| 255  | Asparagine                                   | Dulce   |
| 256^ | Glutamine                                    | Dulce   |
| 257^ | Glycine                                      | Dulce   |
| 258  | Histidine                                    | Dulce   |
| 259  | Isoleucine                                   | Dulce   |
| 260  | Leucine                                      | Dulce   |
| 261  | Methionine                                   | Dulce   |
| 262  | Norvaline                                    | Dulce   |
| 263  | Phenylalanine                                | Dulce   |
| 264  | Proline                                      | Dulce   |
| 265  | Serine                                       | Dulce   |
| 266  | Threonine                                    | Dulce   |
| 267  | Tryptophane                                  | Dulce   |
| 268  | Tyrosine                                     | Dulce   |
| 269^ | Valine                                       | Dulce   |
| 270^ | Alanine tertiary butyl ester                 | Dulce   |
| 271  | Aspartyl-aminomalonic acid diester 1         | Dulce   |
| 272  | Aspartyl-aminomalonic acid diester 2         | Dulce   |
| 273  | Aspartyl-aminomalonic acid diester 3         | Dulce   |
| 274^ | N-(L-aspartyl)-1,1-diaminoalkane 1           | Dulce   |
| 275  | N-(L-aspartyl)-1,1-diaminoalkane 2           | Dulce   |
| 276  | N-(L-aspartyl)-1,1-diaminoalkane 3           | Dulce   |
| 277  | N-(L-aspartyl)-1,1-diaminoalkane 4           | Dulce   |
| 278^ | N-(L-aspartyl)-1,1-diaminoalkane 5           | Dulce   |
| 279  | N-(L-aspartyl)-1,1-diaminoalkane 6           | Dulce   |

---

|      |   |          |
|------|---|----------|
| 280  | N-(L-aspartyl)-1,1-diaminoalkane 7  | Dulce    |
| 281  | N-(L-aspartyl)-1,1-diaminoalkane 8  | Dulce    |
| 282  | N-(L-aspartyl)-1,1-diaminoalkane 9  | Dulce    |
| 283  | Super aspartame   | Dulce    |
| 284  | Sucrononic acid   | Dulce    |
| 285^ | Cyanoarylurea aspartame   | Dulce    |
| 286  | Aspartic acid fenchyl ester   | Dulce    |
| 287^ | Glycine analog of cyanosuosan   | Dulce    |
| 288  | Glycine analog of suosan  | Dulce    |
| 289  | (R)-3-amino-3-phenylpropionic acid analog of N-(4-cyanophenyl)-N' carboxyethyl urea | Insípido |
| 290  | Nitroaniline (2-amino-4-nitrobenzene)   | Dulce    |
| 291  | Methoxy-nitroaniline  | Dulce    |
| 292^ | Ethoxy-nitroaniline   | Dulce    |
| 293  | Butoxy-nitroaniline   | Dulce    |
| 294  | Fluoro-nitroaniline   | Dulce    |
| 295^ | Chloro-nitroaniline   | Dulce    |
| 296  | Bromo-nitroaniline  | Dulce    |
| 297^ | Iodo-nitroaniline   | Dulce    |
| 298^ | Methyl-nitroaniline   | Dulce    |
| 299  | Allyl oxy-nitroaniline  | Dulce    |
| 300  | iso Propoxy-nitroaniline  | Dulce    |
| 301^ | Hydroxy-nitroaniline  | Dulce    |
| 302  | Saccharine derivative 5   | Dulce    |
| 303^ | Saccharine derivative 7   | Insípido |
| 304  | Saccharine derivative 8   | Insípido |
| 305  | Saccharine derivative 9   | Insípido |
| 306^ | Saccharine derivative 10  | Insípido |
| 307^ | Saccharine derivative 11  | Insípido |
| 308  | Saccharine derivative 12  | Insípido |
| 309  | Saccharine derivative 13  | Dulce    |
| 310  | Saccharine derivative 15  | Dulce    |
| 311  | Saccharine derivative 16  | Dulce    |
| 312  | Saccharine derivative 17  | Insípido |
| 313  | Saccharine derivative 18  | Insípido |
| 314  | Saccharine derivative 21  | Insípido |
| 315  | Saccharine derivative 23  | Dulce    |
| 316  | Saccharine derivative 25  | Dulce    |
| 317  | Thiophenesaccharine   | Dulce    |
| 318  | DL-4-(2-Chloro-2-propyl)-1-carboxaldehyde   | Insípido |
| 319^ | Furfuraldehyde oxime  | Dulce    |
| 320  | syn-hydroxymethyl Furfuraldehyde oxime  | Insípido |
| 321^ | 5-Benzyl-2-Furfuraldehyde oxime   | Dulce    |
| 322  | Dihydroquercetin 3-acetate  | Dulce    |
| 323  | Hesperitin  | Dulce    |
| 324  | Neoastilbin   | Dulce    |
| 325  | ortho-Tolylurea   | Insípido |
| 326  | para-Tolylurea  | Dulce    |
| 327  | 4-Propoxyphenyl urea  | Dulce    |
| 328^ | Suosan derivative 1 (suosan)  | Dulce    |

---

|      |   |          |
|------|---|----------|
| 329  | Suosan derivative 2   | Insípido |
| 330  | Suosan derivative 3   | Insípido |
| 331^ | Suosan derivative 4   | Insípido |
| 332  | Suosan derivative 6   | Insípido |
| 333  | Suosan derivative 7   | Insípido |
| 334  | Suosan derivative 8   | Insípido |
| 335^ | Suosan derivative 9   | Insípido |
| 336  | Suosan derivative 10  | Insípido |
| 337^ | Suosan derivative 11  | Dulce    |
| 338^ | Suosan derivative 13  | Dulce    |
| 339^ | Suosan derivative 14  | Insípido |
| 340  | Suosan derivative 15  | Insípido |
| 341  | Suosan derivative 16  | Dulce    |
| 342^ | Suosan derivative 17  | Insípido |
| 343  | Suosan derivative 18  | Insípido |
| 344  | Piperine  | Insípido |
| 345  | Methyl-β-D-Fructopyranose   | Dulce    |
| 346^ | 5-deoxy-D-threo-Hexulose  | Dulce    |
| 347  | Oxathiazinon dioxide derivative 1                                 | Dulce    |
| 348  | Oxathiazinon dioxide derivative 2                                 | Dulce    |
| 349  | Oxathiazinon dioxide derivative 3                                 | Dulce    |
| 350^ | Oxathiazinon dioxide derivative 4                                 | Dulce    |
| 351  | Oxathiazinon dioxide derivative 5                                 | Dulce    |
| 352  | Oxathiazinon dioxide derivative 6                                 | Dulce    |
| 353^ | 2-aminobenzoic acid   | Dulce    |
| 354  | N-Carbamoyl derivative 1  | Insípido |
| 355  | N-Carbamoyl derivative 2  | Insípido |
| 356  | 3-anilino-2-styryl-3H-naphthol[1, 2-d]imidazole-5-sulfonate (SSN) | Dulce    |
| 357  | 3-anilino-2-phenyl-3H-naphtho-[1, 2-d]imidazol-5-sulfonate (TSN)  | Insípido |
| 358  | Perillartine  | Dulce    |
| 359  | 1',4',6'-3Cl-sucrose  | Dulce    |
| 360  | 1',4'-2Cl-sucrose   | Dulce    |
| 361  | 4'-Br-4,1',6'-3Cl-sucrose   | Dulce    |
| 362^ | 4'-F-4,1',6'-3Cl-sucrose  | Dulce    |
| 363^ | 4,1',4',6'-4Br-sucrose  | Dulce    |
| 364  | 4,1',4',6'-4Cl-sucrose  | Dulce    |
| 365  | 4,1',4'-3Cl-sucrose   | Dulce    |
| 366  | 4,1',6'-3Br-sucrose   | Dulce    |
| 367^ | 4,1',6'-3F-sucrose  | Dulce    |
| 368  | 4,1',6'-3I-sucrose  | Dulce    |
| 369  | 4,1'-2Cl-sucrose  | Dulce    |
| 370^ | 4,4',6'-3Cl-sucrose   | Dulce    |
| 371  | 4,6,1',6'-4Cl-sucrose   | Dulce    |
| 372^ | 4-Cl-sucrose  | Dulce    |
| 373  | 4-F-1',4',6'-3Br-sucrose  | Dulce    |
| 374  | 4-F-1',4',6'-3Cl-sucrose  | Dulce    |
| 375  | 4-O-n-propyl glycosides DHC                                       | Dulce    |
| 376  | 4β,10α-dimethyl-1,2,3,4,5,10-hexahydrofluorene-4α,6α-             | Dulce    |

---

|      |                                    |       |
|------|------------------------------------|-------|
|      | dicarboxylic acid                  |       |
| 377  | 6,1',6'-3Cl-sucrose                | Dulce |
| 378  | 6-chloro-tryptophan                | Dulce |
| 379^ | Liquiritin                         | Dulce |
| 380^ | N-Acetyl-D-glucosamine             | Dulce |
| 381  | Periandrin I                       | Dulce |
| 382  | Periandrin II                      | Dulce |
| 383  | Periandrin III                     | Dulce |
| 384  | Periandrin IV                      | Dulce |
| 385^ | Abrusoside B                       | Dulce |
| 386  | Abrusoside C                       | Dulce |
| 387^ | Abrusoside D                       | Dulce |
| 388  | Acesulfame potassium_CH3CH2CH2_CH3 | Dulce |
| 389  | Acesulfame potassium_CH3CH2_H      | Dulce |
| 390  | Acesulfame potassium_H_H           | Dulce |
| 391  | Albiziasaponin B                   | Dulce |
| 392  | Arabinose                          | Dulce |
| 393  | Asp 1                              | Dulce |
| 394^ | Asp 2                              | Dulce |
| 395  | Asp 3                              | Dulce |
| 396  | Asp 4                              | Dulce |
| 397  | Asp 5                              | Dulce |
| 398  | Asp 6                              | Dulce |
| 399  | Asp 7                              | Dulce |
| 400^ | Asp 8                              | Dulce |
| 401  | Asp 9                              | Dulce |
| 402  | Asp 10                             | Dulce |
| 403^ | Asp 11                             | Dulce |
| 404  | Asp 14                             | Dulce |
| 405  | Asp 15                             | Dulce |
| 406  | Asp 16                             | Dulce |
| 407^ | Asp 17                             | Dulce |
| 408  | Asp 18                             | Dulce |
| 409  | Asp 20                             | Dulce |
| 410  | Asp 21                             | Dulce |
| 411  | Asp 22                             | Dulce |
| 412  | Asp 23                             | Dulce |
| 413^ | Asp 24                             | Dulce |
| 414  | Asp 52                             | Dulce |
| 415  | Asp 53                             | Dulce |
| 416  | Asp 55                             | Dulce |
| 417^ | Asp 56                             | Dulce |
| 418  | Asp 58                             | Dulce |
| 419  | Asp 60                             | Dulce |
| 420  | Asp 61                             | Dulce |
| 421  | Asp 62                             | Dulce |
| 422  | Asp 63                             | Dulce |
| 423  | Asp 64                             | Dulce |
| 424  | Asp 65                             | Dulce |
| 425  | Asp 66                             | Dulce |

---

---

|      |         |       |
|------|---------|-------|
| 426  | Asp 67  | Dulce |
| 427^ | Asp 68  | Dulce |
| 428^ | Asp 69  | Dulce |
| 429  | Asp 70  | Dulce |
| 430  | Asp 71  | Dulce |
| 431  | Asp 72  | Dulce |
| 432^ | Asp 73  | Dulce |
| 433  | Asp 74  | Dulce |
| 434^ | Asp 75  | Dulce |
| 435  | Asp 76  | Dulce |
| 436^ | Asp 77  | Dulce |
| 437  | Asp 78  | Dulce |
| 438  | Asp 79  | Dulce |
| 439^ | Asp 80  | Dulce |
| 440  | Asp 81  | Dulce |
| 441^ | Asp 82  | Dulce |
| 442  | Asp 83  | Dulce |
| 443  | Asp 84  | Dulce |
| 444  | Asp 85  | Dulce |
| 445  | Asp 86  | Dulce |
| 446  | Asp 87  | Dulce |
| 447  | Asp 88  | Dulce |
| 448  | Asp 89  | Dulce |
| 449  | Asp 90  | Dulce |
| 450  | Asp 91  | Dulce |
| 451  | Asp 92  | Dulce |
| 452  | Asp 93  | Dulce |
| 453  | Asp 94  | Dulce |
| 454  | Asp 95  | Dulce |
| 455^ | Asp 96  | Dulce |
| 456  | Asp 97  | Dulce |
| 457^ | Asp 98  | Dulce |
| 458  | Asp 99  | Dulce |
| 459^ | Asp 100 | Dulce |
| 460^ | Asp 101 | Dulce |
| 461^ | Asp 102 | Dulce |
| 462^ | Asp 103 | Dulce |
| 463  | Asp 104 | Dulce |
| 464  | Asp 105 | Dulce |
| 465^ | Asp 106 | Dulce |
| 466  | Asp 107 | Dulce |
| 467  | Asp 108 | Dulce |
| 468  | Asp 109 | Dulce |
| 469^ | Asp 110 | Dulce |
| 470  | Asp 112 | Dulce |
| 471^ | Asp 122 | Dulce |
| 472  | Asp 123 | Dulce |
| 473  | Asp 124 | Dulce |
| 474  | Asp 125 | Dulce |
| 475  | Asp 126 | Dulce |

---

---

|      |         |       |
|------|---------|-------|
| 476  | Asp 128 | Dulce |
| 477^ | Asp 129 | Dulce |
| 478^ | Asp 130 | Dulce |
| 479  | Asp 131 | Dulce |
| 480^ | Asp 132 | Dulce |
| 481  | Asp 134 | Dulce |
| 482  | Asp 135 | Dulce |
| 483  | Asp 136 | Dulce |
| 484  | Asp 137 | Dulce |
| 485  | Asp 138 | Dulce |
| 486^ | Asp 139 | Dulce |
| 487^ | Asp 140 | Dulce |
| 488  | Asp 141 | Dulce |
| 489  | Asp 142 | Dulce |
| 490^ | Asp 143 | Dulce |
| 491^ | Asp 153 | Dulce |
| 492  | Asp 154 | Dulce |
| 493^ | Asp 155 | Dulce |
| 494  | Asp 156 | Dulce |
| 495  | Asp 157 | Dulce |
| 496  | Asp 159 | Dulce |
| 497  | Asp 160 | Dulce |
| 498^ | Asp 161 | Dulce |
| 499  | Asp 163 | Dulce |
| 500  | Asp 164 | Dulce |
| 501^ | Asp 165 | Dulce |
| 502  | Asp 166 | Dulce |
| 503^ | Asp 167 | Dulce |
| 504  | Asp 168 | Dulce |
| 505  | Asp 169 | Dulce |
| 506  | Asp 170 | Dulce |
| 507  | Asp 171 | Dulce |
| 508  | Asp 172 | Dulce |
| 509  | Asp 173 | Dulce |
| 510^ | Asp 181 | Dulce |
| 511  | Asp 182 | Dulce |
| 512  | Asp 183 | Dulce |
| 513  | Asp 184 | Dulce |
| 514  | Asp 185 | Dulce |
| 515  | Asp 186 | Dulce |
| 516  | Asp 187 | Dulce |
| 517  | Asp 188 | Dulce |
| 518  | Asp 189 | Dulce |
| 519^ | Asp 190 | Dulce |
| 520  | Asp 191 | Dulce |
| 521  | Asp 192 | Dulce |
| 522  | Asp 193 | Dulce |
| 523  | Asp 194 | Dulce |
| 524^ | Asp 197 | Dulce |
| 525  | Asp 198 | Dulce |

---

|      |                          |       |
|------|--------------------------|-------|
| 526^ | Asp 204                  | Dulce |
| 527  | Asp 207                  | Dulce |
| 528^ | Asp 208                  | Dulce |
| 529^ | Asp 211                  | Dulce |
| 530^ | Asp 212                  | Dulce |
| 531^ | Asp 213                  | Dulce |
| 532  | Asp 214                  | Dulce |
| 533  | Asp 215                  | Dulce |
| 534  | Baiyunoside              | Dulce |
| 535  | Benzenepropanoic acid    | Dulce |
| 536^ | Bernardame               | Dulce |
| 537  | cyclamate                | Dulce |
| 538^ | Cyclocarioside A         | Dulce |
| 539  | Gaudichaudioside A       | Dulce |
| 540  | Hesperinin glucoside DHC | Dulce |
| 541  | Homoeohesperidin DHC     | Dulce |
| 542  | Isovanillic 18           | Dulce |
| 543^ | Isovanillic 24           | Dulce |
| 544  | isovanillic 25           | Dulce |
| 545^ | Leucrose                 | Dulce |
| 546  | Lugduname                | Dulce |
| 547  | Mogroside II             | Dulce |
| 548^ | Mogroside IV             | Dulce |
| 549^ | Mogroside V              | Dulce |
| 550^ | Naringin DHC             | Dulce |
| 551^ | Neoeriocitrin DHC        | Dulce |
| 552  | Osladin                  | Dulce |
| 553  | Polypodoside A           | Dulce |
| 554^ | Prunin DHC               | Dulce |
| 555  | Pterocaryoside A         | Dulce |
| 556  | Pterocaryoside B         | Dulce |
| 557  | Rebaudioside F           | Dulce |
| 558  | Rubusoside               | Dulce |
| 559  | Saccharin                | Dulce |
| 560  | Selliguesain A           | Dulce |
| 561  | Siamenoside I            | Dulce |
| 562  | Sorbose                  | Dulce |
| 563  | Telosmoside A            | Dulce |
| 564  | trans-anethole           | Dulce |
| 565^ | trans-cinnamaldehyde     | Dulce |
| 566  | Xylobiose                | Dulce |

^ predicción

**Tabla 10A.** Nombres químicos y respuesta experimental de los 649 compuestos dulces y no dulces

| No. | Nombre                                       | Clase  |
|-----|--|--------|
| 1   | Sucrose                                      | Dulce  |
| 2   | Sucralose / 4,1',6'-Trichloro-galactosucrose | Dulce  |
| 3^  | Alitame                                      | Dulce  |
| 4   | Aspartame/Aspartyl-phenylalanine methylester | Dulce  |
| 5   | Tagatose                                     | Dulce  |
| 6^  | Maltitol                                     | Dulce  |
| 7   | Isomaltulose/Palatinose                      | Dulce  |
| 8   | Trehalose                                    | Dulce  |
| 9   | Acesulfame potassium                         | Dulce  |
| 10  | Sodium cyclamate                             | Dulce  |
| 11  | Neohesperidine dihydrochalcone               | Dulce  |
| 12  | Neotame                                      | Dulce  |
| 13  | Sodium saccharin                             | Dulce  |
| 14^ | Isomalt (Palatinit) / Isomaltitol            | Dulce  |
| 15  | Sorbitol                                     | Dulce  |
| 16^ | Erythritol                                   | Dulce  |
| 17  | Xylitol                                      | Dulce  |
| 18  | Stevioside                                   | Dulce  |
| 19  | Steviolbioside                               | Dulce  |
| 20  | Rebaudioside A                               | Dulce  |
| 21^ | Rebaudioside B                               | Dulce  |
| 22  | Rebaudioside C                               | Dulce  |
| 23  | Rebaudioside D                               | Dulce  |
| 24^ | Rebaudioside E                               | Dulce  |
| 25  | Dulcoside A                                  | Dulce  |
| 26  | Glycyrrhizin                                 | Dulce  |
| 27  | Hernandulcin                                 | Dulce  |
| 28  | Fructose                                     | Dulce  |
| 29  | Glucose                                      | Dulce  |
| 30  | Xylose                                       | Dulce  |
| 31  | Maltose                                      | Dulce  |
| 32  | Lactulose                                    | Dulce  |
| 33  | Lactosucrose                                 | Dulce  |
| 34^ | Raffinose                                    | Dulce  |
| 35^ | Rhamnose                                     | Dulce  |
| 36  | Naringin                                     | Amargo |
| 37^ | Gentiobiose                                  | Amargo |
| 38  | Limonin                                      | Amargo |
| 39^ | Oleuropein                                   | Amargo |
| 40^ | Caffeine                                     | Amargo |
| 41  | Poncirin                                     | Amargo |
| 42  | Neoeriocitrin                                | Amargo |
| 43  | Quinine                                      | Amargo |
| 44  | Divinyl glycol                               | Amargo |
| 45  | Theobromine                                  | Amargo |
| 46  | Picrocrocin                                  | Amargo |

|     |  |          |
|-----|--|----------|
| 47  | Glycerol   | Dulce    |
| 48^ | Saccharin derivative 1   | Dulce    |
| 49  | Saccharin derivative 2   | Dulce    |
| 50  | Saccharin derivative 3   | Insípido |
| 51  | 3-nitro-o-toluidine  | Insípido |
| 52  | 3-nitro-p-toluidine  | Insípido |
| 53  | 2-amino-4-nitro-propoxybenzene   | Dulce    |
| 54  | 2-nitro-4-amino-propoxybenzene   | Insípido |
| 55  | 2,4-dinitro-propoxybenzene   | Amargo   |
| 56  | Dulcin (p-ethoxyphenylurea)  | Dulce    |
| 57  | 4-Ethoxyphenylthiourea   | Amargo   |
| 58  | o-ethoxyphenylurea   | Insípido |
| 59  | Dihydroquercetin 3-acetate 4'-(methylether) (3)                                  | Dulce    |
| 60  | Flavan (4)   | Dulce    |
| 61^ | Compound 5   | Dulce    |
| 62  | 1,3-benzodioxane (6)   | Dulce    |
| 63  | 1,4-benzodioxane (7)   | Dulce    |
| 64  | Bibenzyl derivative (8)  | Dulce    |
| 65  | 1,4-benzodioxin (9)  | Insípido |
| 66^ | Isoflavan or 3-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1benzopyran (10)       | Dulce    |
| 67^ | 2-(3-Hydroxy-4-methoxyphenyl)-4,5-dihydro-1,3-benzodioxepine (11)                | Insípido |
| 68  | 3-(3-Hydroxy-4-methoxyphenyl)-1,5-dihydro-2,4-benzodioxepine (12)                | Insípido |
| 69  | 2-(3-hydroxy-4-methoxyphenyl)-1,3-benzodioxole (13)                              | Dulce    |
| 70  | Compound 23  | Insípido |
| 71  | Compound 24  | Insípido |
| 72^ | Compound 25  | Insípido |
| 73  | 2-(3-Hydroxy-4-methoxyphenyl)-4-H-1,3-benzodithiine (10a)                        | Dulce    |
| 74  | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzoxathiine (10c)                         | Dulce    |
| 75  | 2-(3-Hydroxy-4-methoxyphenyl)-4H-1,3-benzodithiole (10d)                         | Dulce    |
| 76^ | 2-(3-Hydroxy-4-methoxyphenyl)-1,3-benzoxathiole (10e)                            | Dulce    |
| 77  | 7,8-Dimethyl-3-(3-hydroxy-4-methoxyphenyl)-1,5-dihydro-2,4-benzodithiepine (10f) | Insípido |
| 78  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiine (15a)                | Dulce    |
| 79  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzodithiine (15b)                | Dulce    |
| 80^ | 2-(3-Hydroxy-4-methylthiophenyl)-2,3-dihydro-1,4-benzoxathiine (15c)             | Insípido |
| 81  | 2-(3-Hydroxy-4-methoxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol (15d)            | Dulce    |
| 82  | 2-(3-Hydroxy-4-methylthiophenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol (15e)         | Insípido |
| 83^ | 2-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran (18)               | Dulce    |
| 84^ | 3-(3-Hydroxy-4-methoxyphenyl)-3,4-dihydro-2H-1-benzothiopyran (19)               | Insípido |
| 85^ | 2-(3-Mercapto-4-methoxyphenyl)-4H-3,1-benzoxathiine (23)                         | Dulce    |
| 86  | Sulfone  | Dulce    |
| 87  | Haematoxylin   | Dulce    |

|                  |  |          |
|------------------|--|----------|
| 88               | 9-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,10-diol (4)        | Dulce    |
| 89               | 10-Methoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-6a,9-diol (5)        | Insípido |
| 90 <sup>^</sup>  | 3,9,10-Trimethoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-4,6a-diol (6) | Insípido |
| 91 <sup>^</sup>  | 4,9,10-Trimethoxy-7,11b-dihydrobenz[b]indeno[1,2-d]pyran-3,6a-diol (7) | Insípido |
| 92 <sup>^</sup>  | Tetramethylhaematoxylin (8)  | Amargo   |
| 93 <sup>^</sup>  | Compound 8   | Dulce    |
| 94 <sup>^</sup>  | Compound 17  | Insípido |
| 95 <sup>^</sup>  | Compound 18  | Dulce    |
| 96 <sup>^</sup>  | Compound 19  | Insípido |
| 97               | Compound 20  | Dulce    |
| 98               | Compound 21  | Dulce    |
| 99               | Compound 25  | Insípido |
| 100              | Compound 26  | Dulce    |
| 101              | Compound 27  | Insípido |
| 102              | Compound 28  | Insípido |
| 103              | Compound 29  | Insípido |
| 104 <sup>^</sup> | Compound 30  | Insípido |
| 105 <sup>^</sup> | 2-(4-methoxybenzoyl) benzoic acid                                      | Dulce    |
| 106              | Compound 2   | Insípido |
| 107 <sup>^</sup> | Compound 3   | Dulce    |
| 108              | Compound 4   | Dulce    |
| 109              | Compound 5   | Insípido |
| 110              | Compound 6   | Insípido |
| 111 <sup>^</sup> | Compound 7   | Dulce    |
| 112              | Compound 8   | Insípido |
| 113              | Compound 9   | Dulce    |
| 114 <sup>^</sup> | 2-(4-Nitrobenzoyl)benzoic Acid (10)                                    | Amargo   |
| 115              | Compound 11  | Insípido |
| 116 <sup>^</sup> | Compound 12  | Dulce    |
| 117              | Compound 13  | Insípido |
| 118              | 2-(3,4-Dihydroxybenzoyl)benzoic Acid (14)                              | Dulce    |
| 119 <sup>^</sup> | 2-(3-Hydroxy-4-methoxybenzoyl)benzoic Acid (15)                        | Dulce    |
| 120 <sup>^</sup> | 2-(4-Hydroxy-3-methoxybenzoyl)benzoic Acid (16)                        | Insípido |
| 121 <sup>^</sup> | 2-(3,4-Dimethoxybenzoyl)benzoic Acid (17)                              | Insípido |
| 122 <sup>^</sup> | 2-(2,4-Dimethoxybenzoyl)benzoic Acid (18)                              | Amargo   |
| 123              | 2-(4-Hydroxy-2-methoxybenzoyl)benzoic Acid (19)                        | Insípido |
| 124 <sup>^</sup> | Compound 20  | Insípido |
| 125 <sup>^</sup> | Compound 21  | Insípido |
| 126              | Compound 22  | Insípido |
| 127              | 2-(4-Methoxyphenylthio) benzoic Acid S-Oxide (23)                      | Dulce    |
| 128              | Compound 24  | Insípido |
| 129              | Compound 25  | Insípido |
| 130              | 2-Hydroxymethylphenyl-4-methoxyphenylmethanol (26)                     | Insípido |
| 131 <sup>^</sup> | Phthalic Acid Mono(4-methoxyphenyl) Ester (27)                         | Dulce    |
| 132              | 3-(4-Methoxyphenyl)benzo[2]furan-1-one (28)                            | Insípido |
| 133              | Compound 29  | Insípido |

|                  |   |          |
|------------------|---|----------|
| 134              | 3-Methoxy-1,3,5-estratrien-4-ol (4)   | Dulce    |
| 135              | Substituted pterocarpan (5)   | Insípido |
| 136              | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-methoxy-6'-hydroxy-[1H]-2',3'-dihydroindene)] (6) | Dulce    |
| 137              | Spiro[(4-H-1,3-benzodithian)-2,1'-(5'-hydroxy-6'-methoxy-[1H]-2',3'-dihydroindene)] (7) | Insípido |
| 138 <sup>^</sup> | Spiro[(4-H-1,3-benzodithian)-2,1'-(4'-hydroxy-5'-methoxy-[1H]-2',3'-dihydroindene)] (8) | Insípido |
| 139              | 2-Phenoxyethyl-2H-7,8-dihydrobenzopyran (9)   | Insípido |
| 140 <sup>^</sup> | 2-Methyl-2-(3-hydroxy-4-methoxyphenyl)-1,4-dihydrobenzodioxane (12)                     | Insípido |
| 141              | Phyllodulcin  | Dulce    |
| 142              | Compound 4  | Dulce    |
| 143              | Compound 6  | Dulce    |
| 144              | Compound 9  | Dulce    |
| 145 <sup>^</sup> | 3',4'-dihydroxyphenyl-3,1-benzoxathiane (18)  | Dulce    |
| 146              | 2-[3-(4-nitrophenoxy)phenyl]-3,1-benzoxathiane (25)                                     | Dulce    |
| 147              | 2-[3-(4-nitrophenylcarbonylmethoxy)-4-methoxyphenyl]-3,1-benzoxathiane (26)             | Dulce    |
| 148              | 2-[3-(4-nitrophenylcarbamoxyloxy-4-methoxy)phenyl]-3,1-benzodioxane (27)                | Dulce    |
| 149 <sup>^</sup> | Compound 32   | Dulce    |
| 150              | 3-(3-Hydroxy-4-methoxyphenyl)-1-(3-carboxyphenyl)propan-1-one (33)                      | Dulce    |
| 151              | Compound 34   | Dulce    |
| 152              | 2-(3-hydroxy-4-methoxyphenoxy)methylbenzoic acid (40)                                   | Dulce    |
| 153              | Compound 42   | Dulce    |
| 154              | Dihydroquercetin 4'-(methyl ether)  | Dulce    |
| 155              | Dihydroquercetin  | Insípido |
| 156              | Compound II or Hydrangenol  | Insípido |
| 157              | Compound III  | Amargo   |
| 158 <sup>^</sup> | Compound IV   | Insípido |
| 159              | Compound V  | Dulce    |
| 160              | Compound VI   | Insípido |
| 161              | Compound VII  | Dulce    |
| 162 <sup>^</sup> | Compound VIII   | Insípido |
| 163              | Compound IX   | Insípido |
| 164 <sup>^</sup> | Compound X  | Insípido |
| 165              | Compound XI   | Insípido |
| 166 <sup>^</sup> | Compound XII  | Insípido |
| 167              | Compound XIV  | Dulce    |
| 168              | Compound XV   | Insípido |
| 169 <sup>^</sup> | Compound XVI  | Insípido |
| 170              | Compound XVII   | Insípido |
| 171              | Compound XVIII  | Amargo   |
| 172              | Compound XIX  | Amargo   |
| 173 <sup>^</sup> | Compound XX   | Amargo   |
| 174              | Compound XXI  | Dulce    |
| 175              | Compound XXII   | Dulce    |
| 176              | Compound XXIII  | Dulce    |
| 177 <sup>^</sup> | Compound XXIV   | Dulce    |

---

|      |                  |          |
|------|------------------|----------|
| 178  | Compound XXV     | Insípido |
| 179  | Compound XXVI    | Insípido |
| 180  | Compound XXVII   | Dulce    |
| 181  | Compound XXVIII  | Insípido |
| 182  | Compound XXIX    | Amargo   |
| 183^ | Compound XXX     | Dulce    |
| 184  | Compound XXXI    | Dulce    |
| 185  | Guaiacol         | Insípido |
| 186  | Compound XXXII   | Dulce    |
| 187  | Compound XXXIII  | Insípido |
| 188  | Compound XXXIV   | Insípido |
| 189^ | Compound XXXV    | Insípido |
| 190  | Compound XXXVII  | Insípido |
| 191  | Compound XXXVIII | Insípido |
| 192^ | Compound XXXIX   | Insípido |
| 193  | Compound XL      | Dulce    |
| 194  | Compound XLI     | Dulce    |
| 195  | Compound XLII    | Dulce    |
| 196  | Compound XLIII   | Dulce    |
| 197  | Compound XLIV    | Amargo   |
| 198  | Compound XLV     | Insípido |
| 199  | Compound XLVI    | Insípido |
| 200^ | Compound XLVII   | Dulce    |
| 201  | Compound XLVIII  | Insípido |
| 202  | Compound XLIX    | Dulce    |
| 203  | Compound L       | Insípido |
| 204^ | Compound LI      | Insípido |
| 205  | Compound LII     | Insípido |
| 206  | Compound LIII    | Insípido |
| 207  | Compound LIV     | Dulce    |
| 208  | Compound LV      | Insípido |
| 209^ | Compound LVI     | Insípido |
| 210  | Compound LVII    | Insípido |
| 211  | Compound LVIII   | Insípido |
| 212^ | Compound LIX     | Dulce    |
| 213  | Compound LX      | Insípido |
| 214  | Compound LXI     | Insípido |
| 215^ | Compound LXII    | Insípido |
| 216  | Compound LXIII   | Insípido |
| 217  | Compound LXIV    | Insípido |
| 218^ | Compound LXV     | Dulce    |
| 219^ | Compound LXVI    | Dulce    |
| 220  | Compound LXVII   | Insípido |
| 221  | Compound LXVIII  | Insípido |
| 222  | Compound LXIX    | Insípido |
| 223  | Compound LXX     | Insípido |
| 224  | Compound LXXI    | Insípido |
| 225  | Compound LXXII   | Amargo   |
| 226  | Compound LXXIII  | Amargo   |
| 227^ | Compound LXXIV   | Amargo   |

---

---

|      |   |          |
|------|---|----------|
| 228  | Compound LXXV                               | Amargo   |
| 229  | Compound LXXVI                              | Amargo   |
| 230  | Compound LXXVII                             | Insípido |
| 231^ | Compound LXXVIII                            | Dulce    |
| 232  | Compound LXXIX                              | Dulce    |
| 233^ | Compound LXXX                               | Dulce    |
| 234  | Compound LXXXI                              | Insípido |
| 235  | Compound LXXXII                             | Insípido |
| 236  | Compound LXXXIII                            | Insípido |
| 237  | Compound LXXXIV                             | Insípido |
| 238  | Compound LXXXV                              | Dulce    |
| 239  | Compound LXXXVIII                           | Insípido |
| 240  | Compound LXXXIX                             | Dulce    |
| 241  | Compound XC                                 | Dulce    |
| 242  | Compound XCI                                | Dulce    |
| 243  | Compound XCII                               | Insípido |
| 244  | Compound XCIII                              | Insípido |
| 245  | Compound XCV                                | Dulce    |
| 246^ | Compound XCVI                               | Dulce    |
| 247  | Compound XCVII                              | Amargo   |
| 248  | Compound XCVIII                             | Insípido |
| 249  | Compound CX                                 | Insípido |
| 250  | Compound C                                  | Insípido |
| 251  | Compound CI                                 | Insípido |
| 252  | Compound CII                                | Insípido |
| 253^ | Nitrobenzene                                | Dulce    |
| 254^ | m-Nitrobenzene                              | Amargo   |
| 255^ | Phenylthiourea                              | Amargo   |
| 256  | Dimethyl urea                               | Dulce    |
| 257  | Urea  | Amargo   |
| 258  | Resorcinol                                  | Dulce    |
| 259^ | Phloroglucinol                              | Dulce    |
| 260  | Glycol                                      | Dulce    |
| 261  | 1,4-Anhydro-ribitol                         | Dulce    |
| 262  | 1,4-Anhydro-mannitol                        | Dulce    |
| 263  | 1,5-Anhydro-mannitol                        | Dulce    |
| 264  | Viboquercitol                               | Dulce    |
| 265  | Allo-inositol                               | Dulce    |
| 266  | Dihydroxyacetone                            | Dulce    |
| 267  | Monohydroxyacetone                          | Dulce    |
| 268  | Methyl dihydroxyacetone                     | Dulce    |
| 269  | Stachyose                                   | Dulce    |
| 270^ | 4-O-Methyl sucrose                          | Dulce    |
| 271  | 6-O-Methyl sucrose                          | Dulce    |
| 272  | 6,6'-di-O-Methyl sucrose                    | Dulce    |
| 273^ | 4,6'-di-O-Methyl sucrose                    | Dulce    |
| 274  | 4,6-di-O-Methyl sucrose                     | Dulce    |
| 275  | 1',6'-di-O-Methyl sucrose                   | Dulce    |
| 276  | Methyl 2-deoxy- $\alpha$ -D-glucopyranoside | Dulce    |
| 277  | Methyl 3-deoxy- $\alpha$ -D-glucopyranoside | Dulce    |

---

|                   |   |          |
|-------------------|---|----------|
| 278               | Methyl 4-deoxy- $\alpha$ -D-gluco-pyranoside  | Dulce    |
| 279               | Methyl 6-deoxy- $\alpha$ -D-gluco-pyranoside  | Dulce    |
| 280               | Methyl- $\alpha$ -D-2,6-dideoxy-gluco-pyranoside                                    | Amargo   |
| 281               | Methyl- $\alpha$ -D-3,6-dideoxy-gluco-pyranoside                                    | Amargo   |
| 282               | Methyl- $\alpha$ -D-4,6-dideoxy-gluco-pyranoside                                    | Amargo   |
| 283^ <sup>^</sup> | 6-Chloro-6-deoxy-D-galactose  | Insípido |
| 284               | 6-Chloro-6-deoxy-D-fructofuranose   | Dulce    |
| 285^ <sup>^</sup> | 1,6-dichloro-1,6-deoxy-D-fructofuranose   | Dulce    |
| 286               | Methyl glucopyranoside  | Dulce    |
| 287^ <sup>^</sup> | 6'-Chloro-sucrose   | Dulce    |
| 288               | 1'-Chloro-sucrose   | Dulce    |
| 289               | 1',6'-Dichloro-sucrose  | Dulce    |
| 290               | 1,6-Di-S-1,6-dithio-D-fructofuranose  | Dulce    |
| 291^ <sup>^</sup> | Alanine   | Dulce    |
| 292^ <sup>^</sup> | Asparagine  | Dulce    |
| 293               | Glutamine   | Dulce    |
| 294               | Glycine   | Dulce    |
| 295               | Histidine   | Dulce    |
| 296               | Methionine  | Dulce    |
| 297               | Serine  | Dulce    |
| 298               | Threonine   | Dulce    |
| 299               | L-Aspartyl-aminomalonic acid diester 1  | Dulce    |
| 300               | L-Aspartyl-aminomalonic acid diester 2  | Dulce    |
| 301               | L-Aspartyl-aminomalonic acid diester 3  | Dulce    |
| 302               | N-(L-aspartyl)-1,1-diaminoalkane 1  | Dulce    |
| 303               | N-(L-aspartyl)-1,1-diaminoalkane 2  | Dulce    |
| 304^ <sup>^</sup> | N-(L-aspartyl)-1,1-diaminoalkane 3  | Dulce    |
| 305^ <sup>^</sup> | N-(L-aspartyl)-1,1-diaminoalkane 4  | Dulce    |
| 306               | N-(L-aspartyl)-1,1-diaminoalkane 5  | Dulce    |
| 307               | N-(L-aspartyl)-1,1-diaminoalkane 6  | Dulce    |
| 308               | N-(L-aspartyl)-1,1-diaminoalkane 7  | Dulce    |
| 309               | N-(L-aspartyl)-1,1-diaminoalkane 8  | Dulce    |
| 310^ <sup>^</sup> | N-(L-aspartyl)-1,1-diaminoalkane 9  | Dulce    |
| 311               | Super aspartame   | Dulce    |
| 312               | Sucrononic acid   | Dulce    |
| 313^ <sup>^</sup> | Cyanoarylurea aspartame   | Dulce    |
| 314               | Aspartic acid fenchyl ester   | Dulce    |
| 315               | Glycine analog of cyanosuosan   | Dulce    |
| 316               | Glycine analog of suosan  | Dulce    |
| 317               | (R)-3-amino-3-phenylpropionic acid analog of N-(4-cyanophenyl)-N' carboxyethyl urea | Insípido |
| 318               | Nitroaniline (2-amino-4-nitrobenzene)   | Dulce    |
| 319               | Methoxy-nitroaniline  | Dulce    |
| 320               | Ethoxy-nitroaniline   | Dulce    |
| 321               | Butoxy-nitroaniline   | Dulce    |
| 322               | Fluoro-nitroaniline   | Dulce    |
| 323^ <sup>^</sup> | Chloro-nitroaniline   | Dulce    |
| 324               | Bromo-nitroaniline  | Dulce    |
| 325               | Iodo-nitroaniline   | Dulce    |
| 326               | Methyl-nitroaniline   | Dulce    |

---

|      |   |          |
|------|---|----------|
| 327  | Allyl oxy–nitroaniline                    | Dulce    |
| 328  | iso Propoxy–nitroaniline                  | Dulce    |
| 329  | Hydroxy–nitroaniline                      | Dulce    |
| 330  | Saccharine derivative 1                   | Amargo   |
| 331  | Saccharine derivative 2                   | Amargo   |
| 332^ | Saccharine derivative 3                   | Amargo   |
| 333  | Saccharine derivative 4                   | Amargo   |
| 334  | Saccharine derivative 5                   | Dulce    |
| 335  | Saccharine derivative 7                   | Insípido |
| 336^ | Saccharine derivative 8                   | Insípido |
| 337  | Saccharine derivative 9                   | Insípido |
| 338  | Saccharine derivative 10                  | Insípido |
| 339  | Saccharine derivative 11                  | Insípido |
| 340  | Saccharine derivative 12                  | Insípido |
| 341^ | Saccharine derivative 13                  | Dulce    |
| 342  | Saccharine derivative 14                  | Amargo   |
| 343  | Saccharine derivative 15                  | Dulce    |
| 344^ | Saccharine derivative 16                  | Dulce    |
| 345  | Saccharine derivative 17                  | Insípido |
| 346  | Saccharine derivative 18                  | Insípido |
| 347  | Saccharine derivative 20                  | Amargo   |
| 348^ | Saccharine derivative 21                  | Insípido |
| 349  | Saccharine derivative 22                  | Amargo   |
| 350^ | Saccharine derivative 23                  | Dulce    |
| 351  | Saccharine derivative 25                  | Dulce    |
| 352  | Saccharine derivative 27                  | Amargo   |
| 353^ | Thiophenesaccharine                       | Dulce    |
| 354  | DL–4–(2–Chloro–2–propyl)–1–carboxaldehyde | Insípido |
| 355  | Furfuraldehyde oxime                      | Dulce    |
| 356^ | syn–hydroxymethyl Furfuraldehyde oxime    | Insípido |
| 357  | 5–Benzyl–2–Furfuraldehyde oxime           | Dulce    |
| 358  | 1–Cycloheptene–1–carboxaldehyde oxime     | Amargo   |
| 359  | anti–Cyclohexane carboxaldehyde           | Amargo   |
| 360  | Dihydroquercetin 3–acetate                | Dulce    |
| 361  | Hesperitin                                | Dulce    |
| 362  | Neoastilbin                               | Dulce    |
| 363  | ortho–Tolylurea                           | Insípido |
| 364^ | meta–Tolylurea                            | Amargo   |
| 365  | para–Tolylurea                            | Dulce    |
| 366  | 4–Propoxyphenyl urea                      | Dulce    |
| 367^ | Suosan derivative 1 (suosan)              | Dulce    |
| 368  | Suosan derivative 2                       | Insípido |
| 369  | Suosan derivative 3                       | Insípido |
| 370^ | Suosan derivative 4                       | Insípido |
| 371^ | Suosan derivative 6                       | Insípido |
| 372^ | Suosan derivative 7                       | Insípido |
| 373  | Suosan derivative 8                       | Insípido |
| 374  | Suosan derivative 9                       | Insípido |
| 375  | Suosan derivative 10                      | Insípido |
| 376  | Suosan derivative 11                      | Dulce    |

---

---

|      |                                   |          |
|------|-----------------------------------|----------|
| 377  | Suosan derivative 12              | Amargo   |
| 378^ | Suosan derivative 13              | Dulce    |
| 379^ | Suosan derivative 14              | Insípido |
| 380  | Suosan derivative 15              | Insípido |
| 381  | Suosan derivative 16              | Dulce    |
| 382  | Suosan derivative 17              | Insípido |
| 383  | Suosan derivative 18              | Insípido |
| 384^ | Pyrrolidine                       | Amargo   |
| 385  | Pyrrrole                          | Amargo   |
| 386  | Imidazole                         | Amargo   |
| 387  | Pyrazole                          | Amargo   |
| 388  | Piperidine                        | Amargo   |
| 389  | Piperazine                        | Amargo   |
| 390^ | Purine                            | Amargo   |
| 391  | Pyridazine                        | Amargo   |
| 392  | Pyrazine                          | Amargo   |
| 393^ | Pyridine                          | Amargo   |
| 394  | Pyrimidine                        | Amargo   |
| 395  | Strichnine                        | Amargo   |
| 396  | Brucine                           | Amargo   |
| 397  | Piperine                          | Insípido |
| 398^ | Solanine                          | Amargo   |
| 399^ | Diterpene derivative 1            | Amargo   |
| 400  | Diterpene derivative 2            | Amargo   |
| 401  | Diterpene derivative 3            | Amargo   |
| 402  | Diterpene derivative 4            | Amargo   |
| 403  | Denatonium chloride               | Amargo   |
| 404  | Denatonium chloride derivative 1  | Amargo   |
| 405^ | Denatonium chloride derivative 2  | Amargo   |
| 406  | Denatonium chloride derivative 3  | Amargo   |
| 407  | Denatonium chloride derivative 4  | Amargo   |
| 408  | Denatonium chloride derivative 5  | Amargo   |
| 409  | Denatonium chloride derivative 6  | Amargo   |
| 410  | Denatonium chloride derivative 7  | Amargo   |
| 411  | Denatonium chloride derivative 8  | Amargo   |
| 412  | Denatonium chloride derivative 9  | Amargo   |
| 413  | Denatonium chloride derivative 10 | Amargo   |
| 414  | Denatonium chloride derivative 11 | Amargo   |
| 415^ | Denatonium chloride derivative 12 | Amargo   |
| 416^ | Denatonium chloride derivative 13 | Amargo   |
| 417  | Methyl-β-D-Fructopyranose         | Dulce    |
| 418  | 5-deoxy-D-threo-Hexulose          | Dulce    |
| 419  | Oxathiazinon dioxide derivative 1 | Dulce    |
| 420^ | Oxathiazinon dioxide derivative 2 | Dulce    |
| 421  | Oxathiazinon dioxide derivative 3 | Dulce    |
| 422^ | Oxathiazinon dioxide derivative 4 | Dulce    |
| 423  | Oxathiazinon dioxide derivative 5 | Dulce    |
| 424  | Oxathiazinon dioxide derivative 6 | Dulce    |
| 425  | 3-amino-4-chlorobenzoic acid      | Amargo   |
| 426  | 2-aminobenzoic acid               | Dulce    |

---

|      |   |          |
|------|---|----------|
| 427  | 2-nitroaniline  | Amargo   |
| 428  | N-Carbamoyl derivative 1  | Insípido |
| 429  | N-Carbamoyl derivative 2  | Insípido |
| 430  | Germacrolide  | Amargo   |
| 431^ | 3-anilino-2-styryl-3H-naphthol[1, 2-d]imidazole-5-sulfonate (SSN)   | Dulce    |
| 432  | 3-anilino-2-phenyl-3H-naphtho-[1, 2-d]imidazol-5-sulfonate (TSN)  | Insípido |
| 433  | Perillartine  | Dulce    |
| 434  | 1',4',6'-3Cl-sucrose  | Dulce    |
| 435^ | 1',4'-2Cl-sucrose   | Dulce    |
| 436  | 4'-Br-4,1',6'-3Cl-sucrose   | Dulce    |
| 437  | 4'-F-4,1',6'-3Cl-sucrose  | Dulce    |
| 438  | 4,1',4',6'-4Br-sucrose  | Dulce    |
| 439  | 4,1',4',6'-4Cl-sucrose  | Dulce    |
| 440  | 4,1',4'-3Cl-sucrose   | Dulce    |
| 441  | 4,1',6'-3Br-sucrose   | Dulce    |
| 442  | 4,1',6'-3F-sucrose  | Dulce    |
| 443  | 4,1',6'-3I-sucrose  | Dulce    |
| 444  | 4,1'-2Cl-sucrose  | Dulce    |
| 445^ | 4,4',6'-3Cl-sucrose   | Dulce    |
| 446^ | 4,6,1',6'-4Cl-sucrose   | Dulce    |
| 447  | 4-Cl-sucrose  | Dulce    |
| 448  | 4-F-1',4',6'-3Br-sucrose  | Dulce    |
| 449^ | 4-F-1',4',6'-3Cl-sucrose  | Dulce    |
| 450  | 4-O-n-propyl glycosides DHC   | Dulce    |
| 451  | 4 $\beta$ ,10 $\alpha$ -dimethyl-1,2,3,4,5,10-hexahydrofluorene-4 $\alpha$ ,6 $\alpha$ -dicarboxylic acid | Dulce    |
| 452  | 6,1',6'-3Cl-sucrose   | Dulce    |
| 453^ | 6-chloro-tryptophan   | Dulce    |
| 454^ | Liquiritin  | Dulce    |
| 455  | N-Acetyl-D-glucosamine  | Dulce    |
| 456  | Periandrin I  | Dulce    |
| 457  | Periandrin II   | Dulce    |
| 458  | Periandrin III  | Dulce    |
| 459  | Periandrin IV   | Dulce    |
| 460  | Abrusoside B  | Dulce    |
| 461  | Abrusoside C  | Dulce    |
| 462^ | Abrusoside D  | Dulce    |
| 463^ | Acesulfame potassium_CH3CH2CH2_CH3  | Dulce    |
| 464  | Acesulfame potassium_CH3CH2_H   | Dulce    |
| 465  | Acesulfame potassium_H_H  | Dulce    |
| 466  | Albiziasaponin B  | Dulce    |
| 467  | Ammonium glycyrrhizinate  | Dulce    |
| 468  | Asp 1   | Dulce    |
| 469  | Asp 2   | Dulce    |
| 470^ | Asp 3   | Dulce    |
| 471  | Asp 4   | Dulce    |
| 472  | Asp 5   | Dulce    |
| 473  | Asp 6   | Dulce    |

---

|      |        |       |
|------|--------|-------|
| 474  | Asp 7  | Dulce |
| 475  | Asp 8  | Dulce |
| 476  | Asp 9  | Dulce |
| 477^ | Asp 10 | Dulce |
| 478  | Asp 11 | Dulce |
| 479  | Asp 14 | Dulce |
| 480  | Asp 15 | Dulce |
| 481  | Asp 16 | Dulce |
| 482  | Asp 17 | Dulce |
| 483^ | Asp 18 | Dulce |
| 484  | Asp 20 | Dulce |
| 485  | Asp 21 | Dulce |
| 486  | Asp 22 | Dulce |
| 487  | Asp 23 | Dulce |
| 488  | Asp 24 | Dulce |
| 489  | Asp 52 | Dulce |
| 490^ | Asp 53 | Dulce |
| 491  | Asp 55 | Dulce |
| 492  | Asp 56 | Dulce |
| 493  | Asp 58 | Dulce |
| 494  | Asp 60 | Dulce |
| 495  | Asp 61 | Dulce |
| 496^ | Asp 62 | Dulce |
| 497  | Asp 63 | Dulce |
| 498  | Asp 64 | Dulce |
| 499  | Asp 65 | Dulce |
| 500^ | Asp 66 | Dulce |
| 501  | Asp 67 | Dulce |
| 502  | Asp 68 | Dulce |
| 503  | Asp 69 | Dulce |
| 504  | Asp 70 | Dulce |
| 505  | Asp 71 | Dulce |
| 506  | Asp 72 | Dulce |
| 507  | Asp 73 | Dulce |
| 508  | Asp 74 | Dulce |
| 509  | Asp 75 | Dulce |
| 510^ | Asp 76 | Dulce |
| 511^ | Asp 77 | Dulce |
| 512  | Asp 78 | Dulce |
| 513  | Asp 79 | Dulce |
| 514  | Asp 80 | Dulce |
| 515^ | Asp 81 | Dulce |
| 516  | Asp 82 | Dulce |
| 517^ | Asp 83 | Dulce |
| 518  | Asp 84 | Dulce |
| 519^ | Asp 85 | Dulce |
| 520  | Asp 86 | Dulce |
| 521  | Asp 87 | Dulce |
| 522^ | Asp 88 | Dulce |
| 523  | Asp 89 | Dulce |

---

---

|      |         |       |
|------|---------|-------|
| 524^ | Asp 90  | Dulce |
| 525  | Asp 91  | Dulce |
| 526  | Asp 92  | Dulce |
| 527  | Asp 93  | Dulce |
| 528  | Asp 94  | Dulce |
| 529  | Asp 95  | Dulce |
| 530  | Asp 96  | Dulce |
| 531  | Asp 97  | Dulce |
| 532  | Asp 98  | Dulce |
| 533  | Asp 99  | Dulce |
| 534  | Asp 100 | Dulce |
| 535  | Asp 101 | Dulce |
| 536  | Asp 102 | Dulce |
| 537  | Asp 103 | Dulce |
| 538  | Asp 104 | Dulce |
| 539^ | Asp 105 | Dulce |
| 540  | Asp 106 | Dulce |
| 541^ | Asp 107 | Dulce |
| 542^ | Asp 108 | Dulce |
| 543^ | Asp 109 | Dulce |
| 544^ | Asp 110 | Dulce |
| 545^ | Asp 112 | Dulce |
| 546  | Asp 122 | Dulce |
| 547  | Asp 123 | Dulce |
| 548^ | Asp 124 | Dulce |
| 549  | Asp 125 | Dulce |
| 550  | Asp 126 | Dulce |
| 551  | Asp 128 | Dulce |
| 552^ | Asp 129 | Dulce |
| 553  | Asp 130 | Dulce |
| 554^ | Asp 131 | Dulce |
| 555  | Asp 132 | Dulce |
| 556  | Asp 134 | Dulce |
| 557  | Asp 135 | Dulce |
| 558  | Asp 136 | Dulce |
| 559  | Asp 137 | Dulce |
| 560^ | Asp 138 | Dulce |
| 561^ | Asp 139 | Dulce |
| 562  | Asp 140 | Dulce |
| 563^ | Asp 141 | Dulce |
| 564  | Asp 142 | Dulce |
| 565  | Asp 143 | Dulce |
| 566  | Asp 153 | Dulce |
| 567  | Asp 154 | Dulce |
| 568  | Asp 155 | Dulce |
| 569  | Asp 156 | Dulce |
| 570^ | Asp 157 | Dulce |
| 571^ | Asp 159 | Dulce |
| 572  | Asp 160 | Dulce |
| 573  | Asp 161 | Dulce |

---

---

|      |   |       |
|------|---|-------|
| 574^ | Asp 163   | Dulce |
| 575^ | Asp 164   | Dulce |
| 576  | Asp 165   | Dulce |
| 577^ | Asp 166   | Dulce |
| 578  | Asp 167   | Dulce |
| 579  | Asp 168   | Dulce |
| 580  | Asp 169   | Dulce |
| 581^ | Asp 170   | Dulce |
| 582  | Asp 171   | Dulce |
| 583  | Asp 172   | Dulce |
| 584  | Asp 173   | Dulce |
| 585^ | Asp 181   | Dulce |
| 586  | Asp 182   | Dulce |
| 587  | Asp 183   | Dulce |
| 588  | Asp 184   | Dulce |
| 589  | Asp 185   | Dulce |
| 590  | Asp 186   | Dulce |
| 591  | Asp 187   | Dulce |
| 592  | Asp 188   | Dulce |
| 593^ | Asp 189   | Dulce |
| 594  | Asp 190   | Dulce |
| 595  | Asp 191   | Dulce |
| 596  | Asp 192   | Dulce |
| 597  | Asp 193   | Dulce |
| 598  | Asp 194   | Dulce |
| 599  | Asp 197   | Dulce |
| 600  | Asp 198   | Dulce |
| 601  | Asp 204   | Dulce |
| 602  | Asp 207   | Dulce |
| 603^ | Asp 208   | Dulce |
| 604  | Asp 211   | Dulce |
| 605  | Asp 212   | Dulce |
| 606  | Asp 213   | Dulce |
| 607  | Asp 214   | Dulce |
| 608  | Asp 215   | Dulce |
| 609^ | Aspartame–acesulfame salt                           | Dulce |
| 610  | Baiyunoside   | Dulce |
| 611  | 3–(3–carbamoyl–2,4,6–tribromophehyl)–propionic acid | Dulce |
| 612  | Bernardame  | Dulce |
| 613^ | Calcium cyclamate                                   | Dulce |
| 614^ | Calcium saccharin                                   | Dulce |
| 615  | cyclamate   | Dulce |
| 616  | Cyclocarioside A                                    | Dulce |
| 617  | Disodium glycyrrhizinate                            | Dulce |
| 618  | Gaudichaudioside A                                  | Dulce |
| 619  | Hesperinin glucoside DHC                            | Dulce |
| 620^ | Homoeohesperidin DHC                                | Dulce |
| 621  | Isovanillic 18                                      | Dulce |
| 622^ | Isovanillic 24                                      | Dulce |
| 623  | isovanillic 25                                      | Dulce |

---

---

|      |                              |       |
|------|------------------------------|-------|
| 624  | Leucrose                     | Dulce |
| 625  | Lugduname                    | Dulce |
| 626  | Mogroside II                 | Dulce |
| 627  | Mogroside IV                 | Dulce |
| 628  | Mogroside V                  | Dulce |
| 629^ | Naringin DHC                 | Dulce |
| 630  | Neoeriocitrin DHC            | Dulce |
| 631  | Osladin                      | Dulce |
| 632^ | Polypodoside A               | Dulce |
| 633^ | Potassium glycyrrhizinate    | Dulce |
| 634^ | Potassium saccharin          | Dulce |
| 635  | Prunin DHC                   | Dulce |
| 636  | Pterocaryoside A             | Dulce |
| 637^ | Pterocaryoside B             | Dulce |
| 638  | Rebaudioside F               | Dulce |
| 639  | Ribose                       | Dulce |
| 640  | Rubusoside                   | Dulce |
| 641  | Saccharin                    | Dulce |
| 642  | Selligueain A                | Dulce |
| 643  | Siamenoside I                | Dulce |
| 644  | Sorbose                      | Dulce |
| 645  | Telosmoside A                | Dulce |
| 646  | trans-anethole               | Dulce |
| 647  | trans-cinnamaldehyde         | Dulce |
| 648^ | Tripotassium glycyrrhizinate | Dulce |
| 649  | Xylobiose                    | Dulce |

---

^ predicción