

UNIVERSIDADE DE SANTIAGO FACULDADE DE FARMÁCIA

DEPARTAMENTO DE QUÍMICA ORGÁNICA



**MODELOS NEUROINFORMÁTICOS PARA A
DESCOBERTA DE FÁRMACOS NEUROPROTECTORES**

Francisco Xavier Romero Durán

Santiago de Compostela, abril de 2016





MODELOS NEUROINFORMÁTICOS PARA A DESCOBERTA DE FÁRMACOS NEUROPROTECTORES

Memória que para optar ao grau de Doutor apresenta

Francisco Xavier Romero Durán



D. XERARDO GARCÍA MERA, Profesor Titular de Química Orgánica e D. HUMBERTO GONZÁLEZ DÍAZ, Profesor IKERBASQUE do Departamento de Química Orgánica II da Universidade do País Vasco,

CERTIFICAM:

Que a memória titulada *MODELOS NEUROINFORMÁTICOS PARA A DESCOBERTA DE FÁRMACOS NEUROPROTECTORES*, que para optar ao Grau de Doutor apresenta FRANCISCO XAVIER ROMERO DURÁN, tem sido realizada baixo a nossa dirección conjunta no Departamento de Química Orgánica da Faculdade de Farmácia da Universidade de Santiago de Compostela.

Polo que autorizam a presentación da Tese indicada, considerando que reúne os requisitos exigidos no artigo 34 do Regulamento de Estudos de Doutoramento, e que como Directores da mesma nom incorrem nas causas de abstenção estabelecidas na lei 30/1992.

E pola presente declaram que esta Tese é idónea para ser apresentada dacordo com o artigo 41 do Regulamento de Estudos de Doutoramento, pola modalidade de compêndio de ARTIGOS, nos que o doutorando tivo participación no peso da investigación e a sua contribución foi decisiva para levar a cabo este traballo.

E que está en coñecemento dos coautores, todos eles participantes nos artigos, que nengum dos traballos reunidos nesta Tese serán presentados por nengum deles noutra Tese de doutoramento, o que assinamos baixo a nosa responsabilidade.

E para que conste, expedimos a presente certificación en Santiago de Compostela, abril de dous mil e dezasseis.

Asdo.: Xerardo García Mera

Asdo.: Humberto González Díaz



Agradecimentos

Aos meus Pais, Maria Josefa Durán Pereira e Joám José Romero Porto. Por todo e por sempre.

A Silvia, pola sua comprensom.

Aos meus irmaos Joám e Maria, e aos meus sobrinhos Adrián e Brais, por dar mais sentido a todo.

A Aritz polo seu brilhante futuro. A Jon e Zoe pola sua meiguice.

Aos meus Titores Xerardo García Mera e Humberto González, pola sua confianza e paciência.

A Javier Ruiz Ojeda. In memoriam





ÍNDICE



1. Introdução	1
1.1 Contexto.....	3
1.2 Neuroprotecçom e Alvos Moleculares.....	5
1.3 A Relaçom Quantitativa Estrutura-Actividade.....	6
1.4 Duas Culturas Estatísticas.....	8
1.5 Análise Discriminante.....	9
1.6 Redes Neurais Artificiais.....	10
1.7 Análise de Redes.....	11
1.8 Síntese e validaçom de compostos sintetizados.....	12
1.9 Neuroinformática e Método indutivo.....	16
1.10 Apresentaçom dos diferentes trabalhos.....	17
2. Objectivos	19
3. Discussom	23
3.1 Modelos QSAR.....	25
3.2 Eficiência dos diferentes descritores moleculares.....	26
3.3 Análise Discriminante.....	27
3.4 Modelos de Redes Neurais.....	27
3.5 Análise de Redes.....	30
3.6 Comparativa de achados experimentais e teóricos.....	31
4. Conclusom	33
5. Resumo. Abstract. Resumen	38

6. Abreviaturas.....	43
7. Bibliografía Geral.....	47
8. Artigos publicados.....	51





1. INTRODUÇOM



1. INTRODUÇÃO

1.1 Contexto.

Para quem convive com um familiar ou pessoa próxima diagnosticada de um distúrbio neurológico degenerativo, é infortunadamente desnecessário expor a relevância que essa circunstância tem no afectado e todo o seu contexto social. As qualidades mais nobres que nos definem como humanos: a memória, a orientação, o juízo, o comportamento social, a personalidade, a autoconsciência e a autonomia pessoal, som destruídas progressiva e inexoravelmente.

Mas, para além disto, o impacto destes trastornos está longe de representar só um drama individual, constituindo um importante problema de saúde pública a nível mundial. O Informe Mundial do Alzheimer do ano 2015 calcula quase 50 milhões de pessoas afectadas por demência a nível global. Umha cifra que se multiplicará por dous em 20 anos. Aproximadamente o 60% das pessoas com demência vivem em países de ingressos baixos ou meios com sistemas sanitários deficientes. Cada 3 segundos há um novo caso de demência no mundo, uns 10 milhões ao ano. Os custos estimados som do 1.09% do produto interior bruto mundial. Outras doenças neurológicas como as enfermidades cerebrovasculares som a causa mais frequente de incapacidade e dependência nos países desenvolvidos (Prince et al., 2015). A Organização Mundial da Saúde considera que as doenças neurológicas som umha importante causa de mortalidade no mundo, e explicariam até o 12% das mortes totais.

Mais de 100 anos transcorridos desde que Alois Alzheimer descrevera as características patológicas da doença bautizada com o seu nome, e graças a inúmeros avances técnicos, tenhem-se acumulado ingentes quantidades de conhecimentos acerca dos mecanismos de dano neurológico. Porém, na imensa maioria de distúrbios neurodegenerativos as controvérsias, incertidumes, dilemas e inconsistências dominam ainda o marco conceitual. Isto visualiza-se melhor sabendo que desde o ano 2002 a 2012 se realizaram 413 ensaios clínicos com 244 compostos para detectar benefícios sintomáticos ou fármacos modificadores do curso da doença de Alzheimer. O 99.6% dos ensaios deram resultados negativos, e ainda as medicações aprovadas nom passam de oferecer um discretíssimo e inconstante benefício sintomático (Cummings et al., 2014). As perspectivas nom som mais optimistas para outros trastornos neurodegenerativos como os diversos Parkinsonismos, enfermidades da motoneurona ou atáxias degenerativas.

Todo isto tem a ver com a inerente complexidade do sistema nervoso. Até um terço dos 20000 genes que compreende o genoma humano som expressados principalmente no cerebro. Nengum outro órgão humano exhibe um número comparável. Como consequência, o dano ao sistema nervoso inicia intrincados processos que abrangem múltiplas interacções entre diferentes dianas moleculares e vias metabólicas. Estas perturbações som responsáveis de

processos como a inflamação, estresse oxidativo, ruptura da barreira hemato-encefálica, excitotoxicidade, apoptose ou autofagia (Jain, 2011).

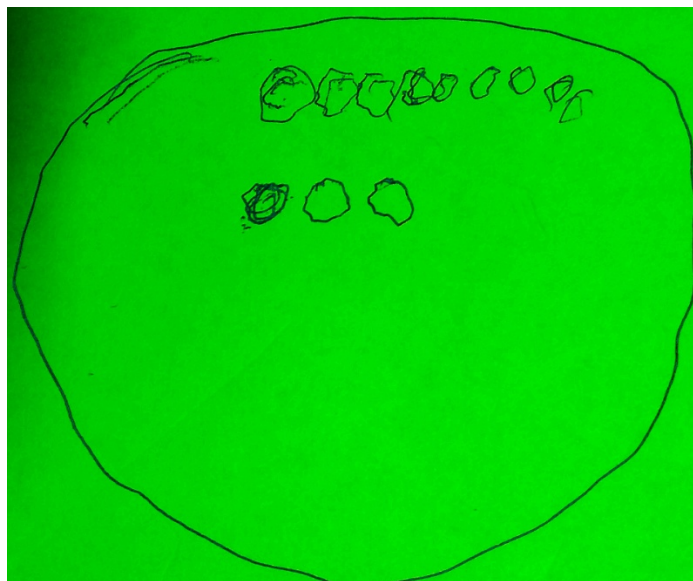


Figura 1. Um relógio marcando as 11:10 desenhado por um paciente com demência neurodegenerativa.

A investigação empírica pré-clínica neste campo mostra dificuldades em modelos animais: diferenças entre espécies ou variabilidade intraespécie que limitam a extrapolação dos resultados, preparações com anestésicos que podem interferir sobre o mecanismo dos fármacos estudados, desenho dos ensaios sem suficiente poder de estatístico ou consistência, e condições experimentais muito mais controladas do que num cenário real. As dificuldades para extrapolar modelos animais entendem-se melhor conhecendo que o chimpanzé, o nosso parente evolutivo mais próximo, compartilha a imensa maioria dos genes, mas até o 80% das proteínas são diferentes em ambas espécies (Glazko et al., 2005). A investigação clínica afronta também muitos obstáculos, tal como não ter em conta a frequentemente limitada janela terapêutica desses fármacos, ou que os ensaios sobre humanos incluem demasiadas vezes um amplo rango de lesões ou níveis patogénicos.

Como consequência disto, a capacidade de intervenção sobre estes quadros segue sendo no melhor dos casos excepcionalmente limitada, constituindo uma fonte de frustração quotidiana para a comunidade científica, assim como para o pessoal sanitário responsável da dispensação de cuidados.

Uma possível abordagem para tentar desbloquear esta complexa situação é o uso de técnicas matemáticas e computacionais para a predição de fármacos e alvos neurofarmacológicos. Neste sentido, os modelos teóricos são indispensáveis para extrair conhecimento útil de grandes conjuntos de dados experimentais disponíveis em bases de dados de acesso público (Nurisso et al., 2012). No nosso trabalho utilizamos como fonte de dados empíricos uma dessas grandes bases de dados, a oferecida pelo Instituto Europeu de

Bioinformática, ChEMBL, com mais de 11,420,000 dados de actividade para mais de 1,715,667 compostos e 10,774 dianas (Gaulton et al., 2012).

O objectivo principal desta tese é modelar matematicamente a relação entre a actividade de diversas moléculas contra alvos implicados em procesos de neuroprotecção em diferentes circunstancias experimentais (**Fig. 2**). Por multiplexado devemos entender o ensaio dumha pluralidade de compostos y dianas, tendo presente estas diversas condições empíricas que modulam essa relação. Ademais, sintetizamos novos derivados de rasagilina e de análogos de rasagilina, umha molécula associada à modificação do curso da doença de Parkinson, e medimos experimentalmente a sua capacidade neuroprotectora ante diferentes estresantes neurais, tratando logo de reproduzir os resultados com a ajuda do nosso modelo .

Ainda que no momento actual constam na literatura científica modelos semelhantes ao nosso, em todos os casos vam dirigidos a umha diana concreta ou partem da avaliação dum número mais limitado de moléculas testadas, e nom recolhem a condição de ter a capacidade de predizer o resultado da actividade em diversas circunstancias empíricas.

A continuação se introduzem os conceitos unificadores mais importantes sobre os que orbita esta tese.

1.2 Neuroprotecção e Alvos Moleculares.

Neuroprotecção é qualquer intervenção terapéutica que evita ou retrasa o dano a células do sistema nervoso central após umha lesão aguda ou crónica degenerativa, interferindo com algum dos passos fisiopatológicos que conduzem à disfunção ou morte neuronal. Neuroprotector é um agente que mostra propiedades neuroprotectoras. En torno a 500 produtos con efeitos neuroprotectores, que podem ser agrupados em 80 categorias, están em estudo em diversas fases. Alguns exemplos destes grupos de fármacos e intervenções neuroprotectoras som: antioxidantes contra radicais libres, agentes anti-excitotóxicos, antidepressivos, antiepilépticos, inibidores da apoptose, anti-inflamatórios, factores neurotróficos, queladores de ions metálicos, moduladores de canais iónicos, ou terapia génica (Jain, 2011). No entanto, os ensaios clínicos sobre estes compostos foram decepcionantes, e só uns poucos tenhem atingido a prática clínica diária. Ainda nestes casos, o seu potencial neuroprotector é controvertido e nom bem compreendido.

A base de dados pública ChEMBL contém um grande número de resultados de ensaios de compostos orgánicos sobre diferentes alvos expressados no sistema nervoso central (SNC) de distintos organismos. Muitos destes alvos están implicados no mecanismo de acção de compostos com actividade neuroprotectora, como se reflite na literatura científica.

Exemplos proeminentes som a proteína SMN (Survival Motor Neuron), com mais de 30000 compostos ensaiados. Esta proteína está relacionada com um grupo de distúrbios neuromusculares caracterizados pola degeneração de células motoras da médula espinal. Provas crecentes indicam um papel chave na regulação da apoptose, e portanto na supervivencia neuronal, e também um rol anti-oxidativo, sendo a sua regulação relevante

num modelo murino de enfermidade da motoneurona. O receptor da Neuroquinina-1 é outra importante diana com mais de 3600 ensaios. Está relacionado com procesos inflamatórios neurogénicos. Também inibe umha via apoptótica demonstrando neuroprotecção em neuronas dopaminérgicas implicadas na patogénese da Doença de Parkinson. Igualmente, com quase 2000 ensaios, o receptor de factor de crecimiento neuronal asociado à tirosina quinasa A (Trk-A NGFR) média os efeitos neuroprotectores do antidepressivo amitriptilina. Outros alvos profusamente ensaiados som também os diversos receptores de acetil-colina, a enzima sintase do óxido nítrico, ou as monoamina oxidases A e B, todas implicadas nos mecanismos patogénicos de procesos neurodegenerativos.

Na dimensom empírica do nosso trabalho centramo-nos no papel neuroprotector da Rasagilina e moléculas relacionadas. A rasagilina é um potente, selectivo, e irreversível inibidor da monoamina oxidase B de segunda geraçom. A sua açom consiste em evitar a desaminaçom das monoaminas dopamina e feniletilamina, incrementando a sua concentraçom na sinapse à vez que reduz a produçom de moléculas oxidantes. O ensaio clínico ADAGIO, com um desenho especial, sugeriu a sua capacidade para modificar o curso da Doença de Parkinson para além do seu efeito sintomático. Sem embargo no ano 2011, um comité da FDA (Federal Drug Administration) rejeitou essa indicaçom pola inconsistência dalguns resultados e nom considerar suficientemente definitivas as provas (Rascol et al., 2011).

Para além destas controvérsias, é umha das moléculas que mais perto se achegou ao conceito de neuroprotecção nos últimos anos. Portanto, a parte empírica do nosso trabalho consistiu na síntese e validaçom experimental de diversos grupos de derivados e análogos de rasagilina. Estos compostos foram ensaiados analisando a viabilidade de neurónios de embriom de rato ante a ausência e presença de agressons químicas como peróxido de hidrogênio e glutamato.

1.3 A Relaçom Quantitativa Estrutura-Actividade.

Umha das finalidades primordiais da química subjaz no desenvolvimento de substâncias com o perfil desejado de interacçons e comportamento. Aplicando esta lógica à açom sobre os seres vivos, cumpre possuir suficiente conhecimento acerca da natureza do composto para poder potenciar a sua eficácia mentres se reduzem os seus efeitos tóxicos. Desde o século XIX se reconhece a relaçom que existe entre a açom fisiológica de umha substância e a sua composiçom e constituiçom química. Já sobrepasada a metade do século XX começam a criar-se modelos matemáticos de actividades biológicas de compostos usando propriedades fisicoquímicas simples.

O conceito da relaçom quantitativa estrutura-actividade (QSAR) tenta inferir umha relaçom entre o comportamento biológico ou químico dumha molécula apartir de informaçom derivada das suas características físico-químicas e estruturais (Todeschini and Consonni, 2008). Estas propriedades químicas, chamadas descriptors moleculares, podem representar parámetros deduzidos teóricamente ou bem variáveis experimentais. Ambos som, em último termo, dados numéricos tratáveis matematicamente utilizando diversas técnicas de modelado

estatístico. O modelo é capaz de detectar e sistematizar regularidades nos dados que iluminam a natureza dessa relação. Deste modo, pode-se inferir o comportamento biológico de moléculas não ensaiadas empiricamente com só introduzir os seus descritores moleculares no modelo construído (Roy et al., 2015).

Para construir as variáveis de entrada do modelo, cada um dos compostos é transformado a um valor numérico. Mas os descritores moleculares servem também para incorporar informação relativa a cada uma das circunstâncias experimentais em que esse composto é testado. O propósito dos modelos elaborados é prever a ocorrência ou não de uma interação molecular entre um composto concreto e um alvo molecular baixo uma combinação específica de cada um dos protocolos de ensaio, tipo de medida bioquímica, e organismo. Cada uma destas circunstâncias inclui diferentes variáveis categóricas. Por exemplo, dentro da categoria Organismo podemos achar o ser humano e outros organismos, até 12 subgrupos. Nós desestimamos o uso de uma enorme quantidade de variáveis indicatriz para adicionar toda esta informação categórica. Uma abordagem mais razoável a este problema é o uso de operadores de *médias móveis*. Estes são similares a aqueles utilizados em análises de séries temporais, e quantificam a desvio entre o descritor molecular de um composto particular e a média de um conjunto de compostos medidos baixo as mesmas condições experimentais, isto é, sob os mesmos subgrupos categóricos (Box et al., 2011). Neste sentido, eles são também descritores indirectos dos alvos moleculares, já que apresentam diferentes valores para diferentes alvos.

A vantagem essencial da criação destes modelos é incrementar a probabilidade de achar novos fármacos candidatos permitindo a crivagem teórica de um grande número de moléculas potenciais. Deste modo, priorizam-se os ensaios experimentais mais prometedores poupando tempo e recursos.

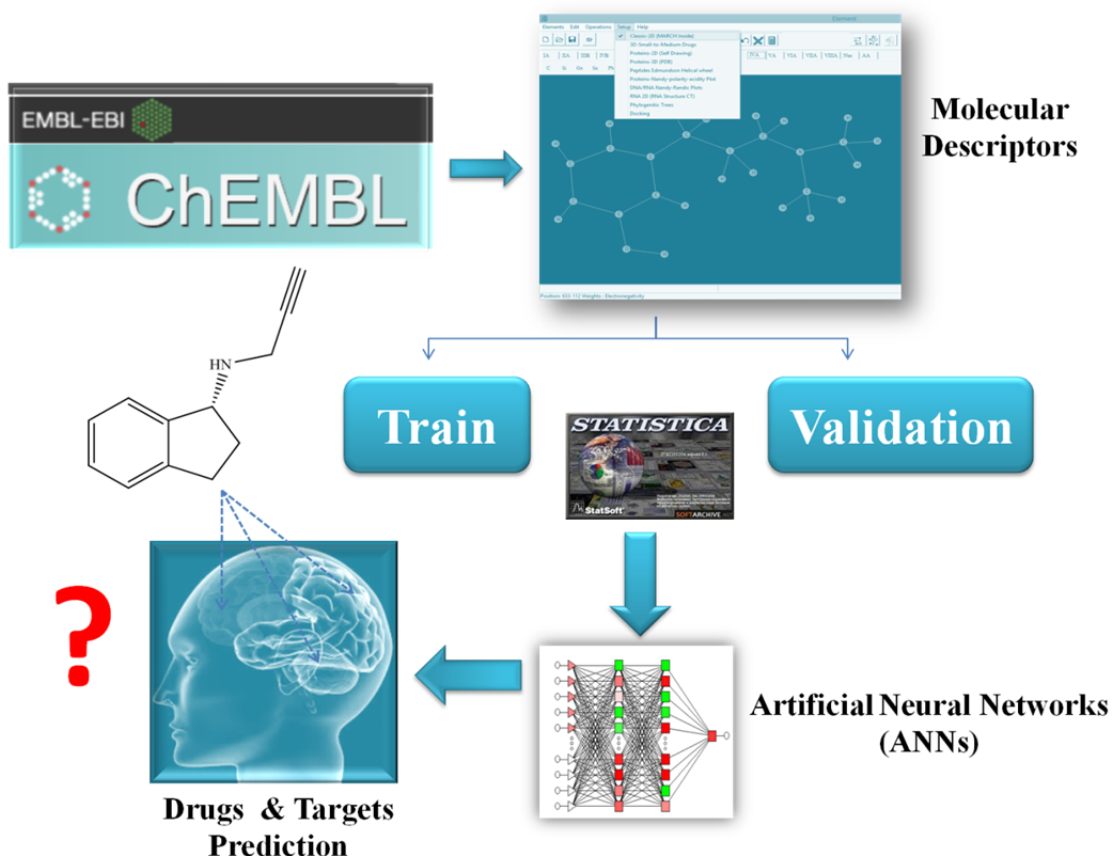


Figura 2. As diferentes fases do método indutivo implementado.

1.4 Duas Culturas Estatísticas.

"Se o único que umha pessoa tem é um martelo, todos os problemas lhe pareceram um cravo."

A análise de dados imagina a natureza como umha função que recebe variáveis de entrada, predictoras ou independentes, e assigna-lhes o valor de variáveis de saída ou resposta. Há duas culturas no uso do modelado estatístico para lograr conclusões úteis apartir da análise dos dados. A primeira assume que os dados som gerados por um modelo estocástico prefixado. A outra usa modelos algorítmicos e trata o mecanismo de produção dos dados como desconhecido.

O primeiro supom a imposição a priori dumhas assunções estreitas sobre os dados. Isto comporta o risco de que as conclusões tiradas do modelo nom se ajustem ao mecanismo natural que as produziu, senom às presunções do próprio modelo. Este risco nom é conjurável apelando só à bondade de ajuste ou à análise de residuais, especialmente em conjuntos de dados complexos e multidimensionais. Poderia dizer-se que o foco recai no modelo e nom nos dados, e as conclusões que oferece podem ser erradas apesar de umha aceitável bondade de ajuste e verificação de residuais.

O modelado algorítmico realiza assunções mais flexíveis e abertas, assumindo o mecanismo de produção dos dados como irrelevante, mas tentando emular com a maior fidelidade os emparelhamentos entre variáveis de entrada e saída que realiza a natureza. Para conjuntos de dados complexos pode-se adaptar melhor às peculiaridades de cada conjunto de dados, e oferece potencialmente maior precisão preditiva. Como contrapartida, os modelos podem ser mais complexos e inescrutáveis. Isto significa focar-se na resolução do problema, em vez de perguntar-se que modelo de dados prefixado pudo ter gerado os resultados (Breiman, 2001). Uma analogia intuitiva seria comparar como se pode ajustar à nossa anatomia um traje comprado de acordo com as medidas padrão em grandes armazéns (modelo estocástico prefixado), a um traje elaborado por um alfaiate após múltiplas medidas do nosso corpo (modelado algorítmico).

Mas não em todas as situações é possível decidir que tipo de aproximação renderia resultados mais úteis. Pelo tanto, neste trabalho tentamos construir modelos QSAR utilizando a análise discriminante como exemplo de primeiro paradigma. Aliás, como exemplo do segundo paradigma servimo-nos de redes neurais artificiais, um subconjunto de algoritmos de aprendizado automático inspirados no funcionamento do sistema nervoso.

1.5 Análise Discriminante.

É um método paramétrico estatístico tradicional que faz parte dos modelos prefixados de dados comentados no ponto anterior.

No contexto do QSAR usa-se em problemas de classificação onde a variável dependente ou resposta é categórica, com tantas categorias como grupos, e os descritores moleculares são as variáveis contínuas predictoras. Matematicamente consiste num procedimento de transformação dos dados para maximizar a variação entre grupos e minimizar a variação intragrupo. Deste modo, obtêm-se um conjunto de fórmulas discriminantes e também funções classificatórias. Há tantas funções classificatórias como número de grupos, e cada variável preditora contribui com um peso diferente para cada uma dessas funções. Cada caso novo é classificado como pertencente ao grupo para o que tem a maior pontuação na sua função classificatória.

As vantagens e desvantagens da análise discriminante são semelhantes às da regressão linear múltipla. Os dados devem cumprir uma série de condições prévias que podem resultar restritivas para alguns conjuntos de dados concretos. Na discussão se detalham alguns destas suposições prévias (Hill et al., 2006).

O desempenho desta técnica estatística tradicional é contrastado com outros métodos estatísticos próprios da aprendizagem automática e inspirados no funcionamento cerebral, como as redes neurais artificiais.

1.6 Redes Neurais Artificiais.

Nos primeiros tempos da inteligência artificial (IA) os computadores eram programados através de uma sequência de expressões matemáticas e lógicas, na assunção de que se pode simular comportamento inteligente através da manipulação de símbolos. Este paradigma da IA obteve sucessos notáveis, como a criação dos Sistemas Expertos, que emulam a capacidade de decisão de expertos humanos em áreas de conhecimento circunscritas. No entanto, tarefas perceptivas tão simples para os humanos como o reconhecimento de objectos eram extraordinariamente difíceis de imitar.

Para confrontar estas deficiências, um novo paradigma denominado *Aprendizado Automático* tomou o relevo nas últimas décadas. O aprendizado automático pode ser definido como o campo de estudo que dá aos computadores a habilidade de aprender da experiência, melhorando o seu desempenho com respeito a alguma tarefa sem ser explicitamente programados. Os métodos de aprendizado automático são os melhores métodos disponíveis quando a tarefa a aprender é demasiado complexa como para desenhar manualmente um algoritmo de aprendizado. Numha tarefa classificatória supervisionada, a experiência consistiria em aprender uma função desconhecida após a apresentação de exemplos junto com a variável de agrupamento ao que pertencem. O algoritmo desenvolvido seria capaz de asignar eficientemente o grupo de pertença de casos novos não utilizados para a construção do modelo.

Uma concreção notável destes métodos são as Redes Neurais Artificiais (RNA). Apesar do seu nome, as redes neurais são formalmente muito diferentes do cérebro em termos de estrutura, mas inspiradas em princípios básicos do funcionamento de circuitos do sistema nervoso. Uma rede neural artificial é uma coleção em paralelo de unidades de processamento não linear capaz de estimar ou aproximar qualquer função desconhecida. O conhecimento da rede subjaz nas conexões entre estas unidades básicas de processamento. Esta propriedade de aproximador universal de funções as converte num instrumento matemático e computacional útil para modelar relações complexas (Haykin et al., 2009). Neste sentido, empregam-se também para descobrir regularidades ocultas no crescente volume de dados empíricos moleculares das neurociências, e para a elaboração de modelos QSAR.

Na última década surgiu um refinamento destes métodos chamado *Aprendizado Profundo*. O aprendizado profundo inspira-se igualmente na biologia, nomeadamente no processo cortical da informação visual. Uma definição simples seria: máquinas que aprendem a representar o mundo de forma integral, de forma que cada estágio na aprendizagem da máquina pode ser treinado (Hinton et al., 2006). Um dos grandes reptos no aprendizado automático tradicional é um processo chamado *extração de características* ou *representações*. Estas representações são regularidades ou abstrações que nos ajudam a entender a variabilidade nos dados. Nas redes neurais não profundas, com apenas uma capa entre as entradas e saídas, o programador precisa dizer ao computador que classe de elementos deveria procurar para ser informativos à hora de tomar uma decisão. Deve

desenhar a mão essas abstrações. Porém, os algoritmos que definem o aprendizado profundo são capazes por eles mesmos de extrair as características adequadas, interpretando os dados numa hierarquia crescente de representações. Desta forma, a medida que os dados são transformados pelas sucessivas camadas da rede cada nova representação está formada por combinações de representações mais simples. Ao final do processo capturam-se características abstractas ocultas nos dados iniciais, muito úteis numa tarefa classificatória por exemplo (LeCun et al., 2015).

1.7 Análise de Redes.

O estudo das redes faz parte da teoria de grafos, um área das matemáticas discretas. A análise de redes permitiu identificar uma série de princípios unificadores e propriedades estatísticas comuns a muitos sistemas do mundo real.

Uma rede é a expressão simplificada dum sistema que captura o padrão de relações entre as entidades que o constituem. Estas entidades podem consistir em objectos tangíveis do mundo real ou ser abstrações sem referente físico. Formalmente, uma rede pode ser representada por um grafo. Um grafo consiste num conjunto de vértices e num conjunto de arestas entre eles. Os componentes do sistema são os nodos ou vértices, e as relações ou enlaces que estabelecem entre eles são as arestas. Um grafo pode traduzir-se a uma matriz para o seu tratamento matemático, permitindo resolver problemas planteados no estudo desses sistemas.

Em toda rede há uma lógica particular que guia o estabelecimento das conexões. Esta disposição das conexões conforma a estrutura topológica da rede, e pode descrever-se matematicamente com umas medidas como o grau, longitude média do caminho entre dois vértices, centralidade, conectividade, ou o coeficiente de agrupamento. À sua vez, desta estrutura emerge uma certa funcionalidade, ou comportamento global da rede, por exemplo a sua robustez ou estabilidade às mudanças (Csermely et al., 2013). Por exemplo, em muitos casos, os vértices com a maior quantidade de conexões jogam papéis importantes na dinâmica do sistema e, portanto, a sua identificação permite focar a nossa atenção nestes elementos cruciais.

Um exemplo interessante de funcionalidade presente em muitos sistemas reais é a chamada rede de Pequeno Mundo, onde a distância entre dois nodos escolhidos aleatoriamente cresce proporcionalmente ao logaritmo do número de nodos presentes na rede. Empíricamente tem-se mostrado que o funcionamento dos neurónios corticais da corteza cerebral reflete propriedades de pequeno mundo, facultando-as para a coexistência de computações globais e locais.

Inspirados nestes princípios, e dado que esta tese pretende iluminar aspectos da neuroquímica cerebral, construímos por primeira vez redes fármaco-alvo a partir de resultados experimentais de ensaios recolhidos na base de dados ChEMBL relacionados com

neuroprotecção. Posteriormente, reconstruímos essa rede apartir das predições do nosso modelo e comparamos os parâmetros topológicos de ambas.

1.8 Síntese e validação de compostos sintetizados.

O nosso trabalho experimental consiste na síntese de moléculas derivadas de rasagilina e do seu enantiómero, o composto TVP1022. A rasagilina é um derivado químico da propargilamina que inibe selectivamente a monomine oxidase tipo B, regulando assi a concentração cerebral de importantes neurotransmissores. Aprovada como um tratamento sintomático efectivo para a Doença de Parkinson Idiopática, umha meta-análise de ensaios clínicos publicada em 2013 confirma a ausencia de efeitos adversos importantes. Mas a razão principal da escolha desta molécula vem dada pola sua implicação em procesos de neuroprotecção. No ano 1985 descreve-se a sua acção limitante da toxicidade de dopamina em modelos animais. No ano 2000 prova-se que prevem a morte celular. Finalmente, as conclusões do estudo ADAGIO publicadas em 2011 sugirem a sua capacidade para modificar o curso da doença parkinsoniana retrasando a necessidade de medicação sintomática. Apesar da cautela dos autores do estudo e certa controvérsia metodológica, esta molécula é sem dúvida um dos melhores exemplos de fármaco neuroprotector que possuímos, portanto a síntese e valoração de compostos derivados apartir da sua estrutura semelha plenamente justificada.

A continuação introduzimos umha descrição superficial destas diversas séries de compostos. Os pormenores técnicos do processo de síntese química e a composição concreta de cada um deles podem achar-se na secção que contem os artigos centrais e no material suplementar acessível por meios electrónicos.

No primeiro artigo (Bioorganic Medicinal Chemistry 2013) descrevemos a preparação de umha série de derivados da rasagilina e de TVP1022, portadores de diferentes substituintes na posição 3 tanto da rasagilina como de TVP1022. A preparação e a estrutura dos compostos preparados fica recolhida no diagrama 1.

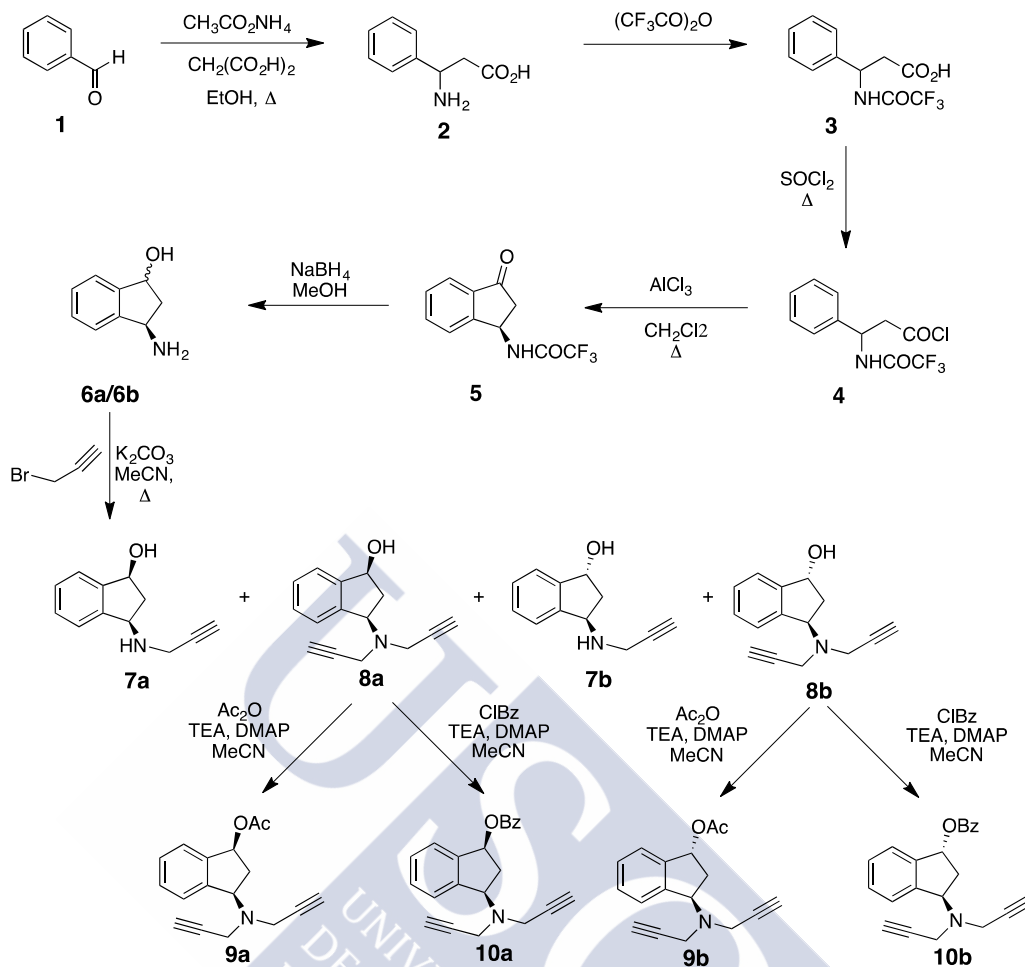


Diagrama 1

No segundo trabalho (ACS Chemical Neuroscience 2013) sintetizam-se derivados da 1,3-rasagilina mono e dipropargilados, que incorporam um resto de carbamato e que apresentam disposição *cis* ou *trans* entre os grupos das posições 1 e 3. Ditos compostos foron sintetizados segundo a estratégia descrita no diagrama 2.

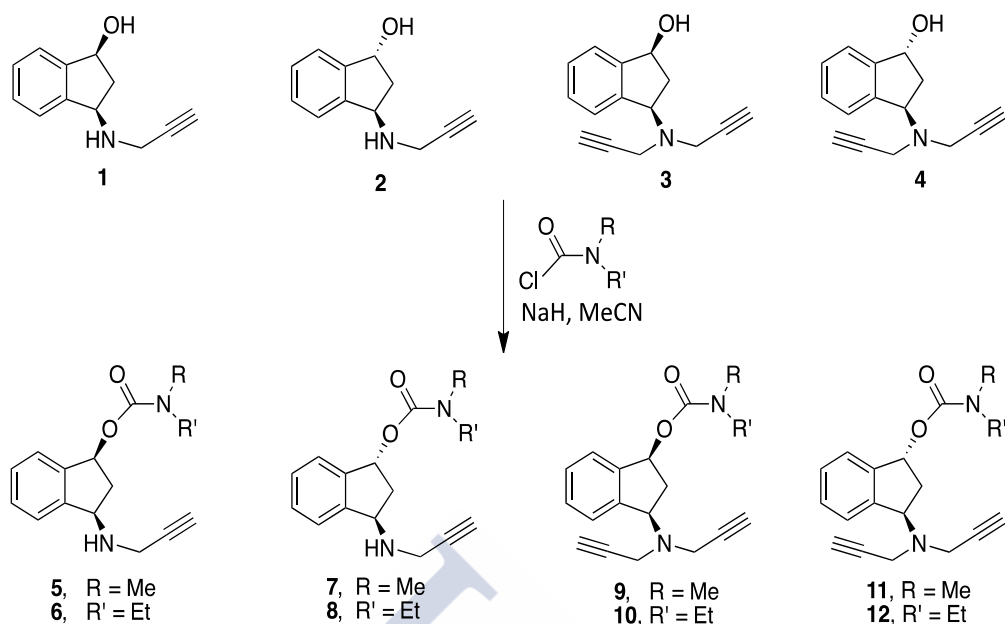


Diagrama 2

No terceiro trabalho (International Journal of Molecular Sciences 2014) e no cuarto traballo (Neuropharmacology 2016), preparamos derivados da rasagilina e do TVP1022 nos que o grupo hidroxilo e os substituíntes que sobre ele se construem están situados na posición 2 do anel ciclopentánico. Som polo tanto isómeros de posición ao respecto dos compostos descritos nos dous anteriores artigos. Outra diferença importante asenta no feito de producir compostos enantiómeros puros. Os compostos preparados nos dous primeiros artigos eran compostos racémicos. Para a súa síntese empregamos como produtos de partida dous aminoálcoois comerciais enantioméricamente puros (1*R*,2*S*)-(+)-1-amino-2-indanol e (1*S*,2*R*)-(-)-1-amino-2-indanol. Apartir deles preparamos os correspondentes ésteres e carbamatos, como se pode apreçar nos diagramas 3 e 4 da páxina seguinte.

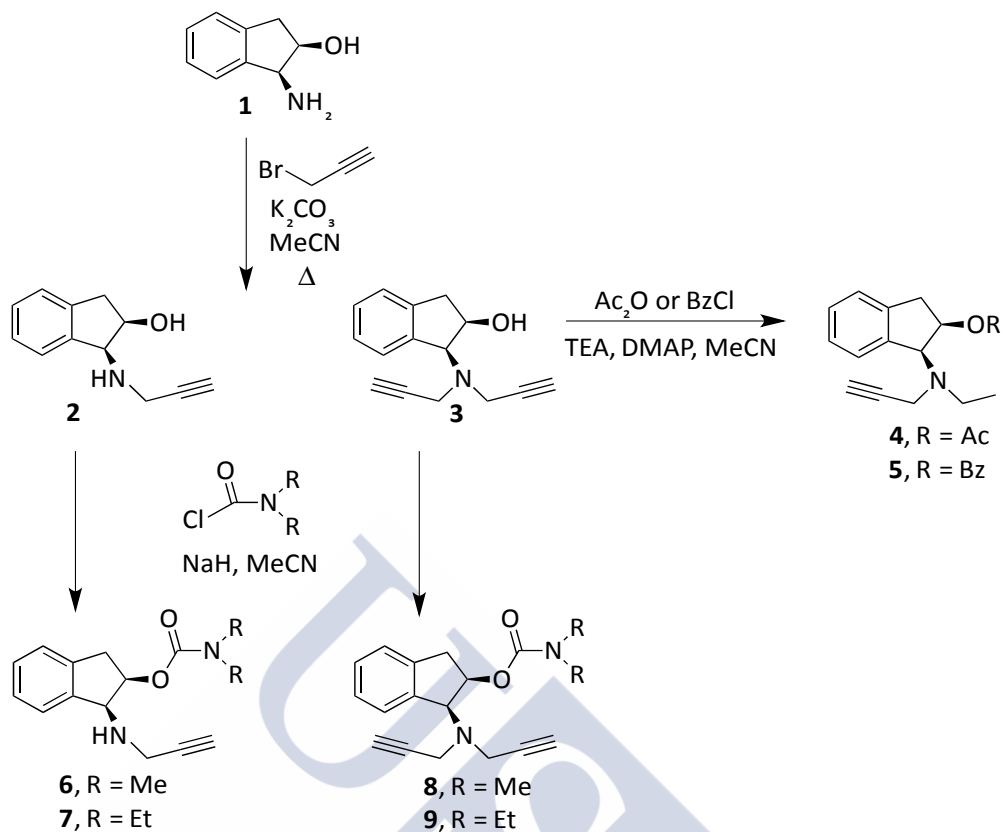


Diagrama 3

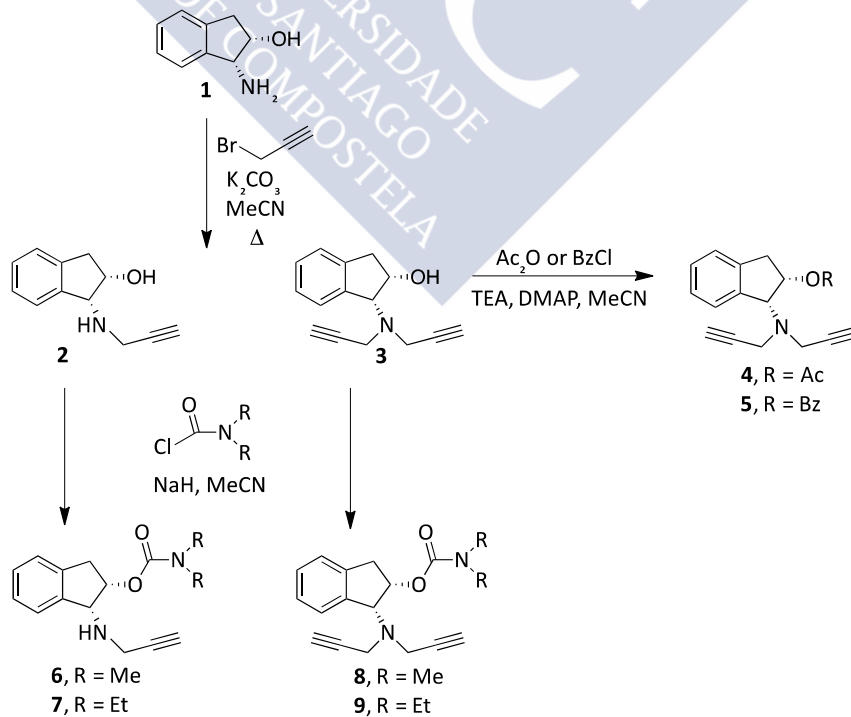


Diagrama 4

A capacidade neuroprotectora de todos estes compostos pondera-se com os mesmos procedimentos. Utilizam-se cultivos de neurónios para examinar a sua viabilidade empregando o método MTT (Bromuro de 3-(4,5- dimetiltiazol-2-ilo)-2,5-difeniltetrazol). Os neurónios provêm da corteza motora de embriões de ratos da linhagem Sprague-Dawley. O método MTT baseia-se numa técnica colorimétrica que permite quantificar a redução metabólica do MTT a um composto colorido (formazan) fazendo uso dum espectrofotómetro. A habilidade dos neurónios para reduzir MTT indica a sua integridade funcional e viabilidade celular. Seguidamente levam-se a cabo os ensaios em três diferentes circunstâncias. Primeiro, estuda-se a habilidade de induzir um efeito neuroprotector em ausência de qualquer estímulo neurotóxico. Logo, a viabilidade neuronal em presença de glutamato, uma substância com propriedades neurotóxicas que conduz à apoptose. Finalmente, examina-se a capacidade dos compostos sintetizados para proteger os neurónios do dano provocado por peróxido de hidrogénio, um neurotóxico implicado na geração de radicais livres e no stress oxidativo.

Ademais da sua síntese e avaliação experimental, as moléculas com um melhor perfil neuroprotector são introduzidas posteriormente nos nossos modelos preditivos para contrastar os seus resultados empíricos com as nossas previsões teóricas.

1.9 Neuroinformática e Método indutivo.

A Neuroinformática é a disciplina que pretende desentranhar a estrutura e funcionamento do SNC e os processos neurais através de técnicas matemáticas e computacionais (De Schutter et al., 2012). Para isto integra, organiza, analisa e elabora modelos a partir de dados derivados das diversas áreas da neurociência. O nascimento desta subespecialidade emergente da neurologia veio motivada inicialmente pela necessidade de criar bases de dados onde poder compartilhar informação obtida na investigação neurocientífica. Com a posta em marcha do Human Brain Project em 1993 introduz-se o termo e reconhece-se a progressiva relevância do campo (Hannawi and Smirnakis, 2013).

O método científico habitual, conhecido como método hipotético-dedutivo, parte da formulação dumha hipótese específica sugerida pela observação dum fenómeno natural. A partir dela desenha-se um experimento controlado, de tal modo que os resultados do experimento possam ser consistentes ou bem inconsistentes com a hipótese. Que umha asserção científica possa ser refutada é umha característica essencial do método científico. No entanto, existe outra possível abordagem no estudo dos sistemas complexos. Na era post-genômica é comum valer-se dumha grande quantidade de dados experimentais para tentar identificar regularidades num sistema complexo, em vez de formular hipóteses concretas sobre um componente do sistema. Estas regularidades servem como fundamento para inferir indutivamente padrões gerais que os expliquem. Ao invés do método tradicional, o ponto de partida são os dados, e o ponto de chegada a geração de novas Hipóteses. Alguns autores falam de *ciência nom baseada em hipótesis* para se referir a esta sequência na que a investigação científica nom é dirigida inicialmente por hipóteses. Em todo caso ambos

métodos, dedutivo e indutivo, som estratégias complementares e iterativas para a compreensão de sistemas complexos (Kell and Oliver, 2004).

Como se explica em secções prévias, este trabalho parte de diversos dados experimentais neuroquímicos na busca dumha síntese que os explique e unifique. Deste modo insere-se sem nengumha dúvida dentro do campo da Neuroinformática. Aliás, este esforço de integração acerca da interacção de moléculas implicadas em processos de neuroprotecção segue umha dinâmica indutiva. Nom há hipóteses específicas a priori que orientem a investigação. Os modelos matemáticos criados som capazes de predizer interacções para outras moléculas e circunstâncias nom testadas experimentalmente por um processo de generalização baseado em casos particulares. Esta generalização logra-se através de técnicas estatísticas tradicionais e, novedosamente, também apartir de novos algoritmos inspirados à sua vez no processo cerebral da informação. O seu valor reside no potencial para gerar múltiplas e mais produtivas hipóteses sobre as interacções moleculares no sistema nervoso.

1.10 Apresentação dos diferentes trabalhos.

A tese compom-se de quatro trabalhos publicados ao longo dos últimos 3 anos.

No trabalho inicial, desenvolvemos o primeiro modelo quimio-informático para ensaios multiplexados dos efeitos neurotóxicos/neuroprotectores de fármacos. Utiliza-se o método TOPS-MODE para calcular os parámetros estruturais dos fármacos.

No segundo, introduz-se o método MARCH-INSIDE para um QSAR multiplexado coa mesma finalidade, calculando os momentos espectrais estocásticos dos compostos.

Em terceiro lugar, apartir do mesmo método anterior elabora-se um QSAR multi-diana que serve como base para criar e comparar duas rede complexas: a observada empíricamente e a predita polo modelo.

O quarto trabalho utiliza diferentes descritores moleculares para comparar o seu desempenho em diversos modelos de redes neurais, introduzindo as redes neurais baseadas no algoritmo de aprendizagem profunda para o proceso dos dados.





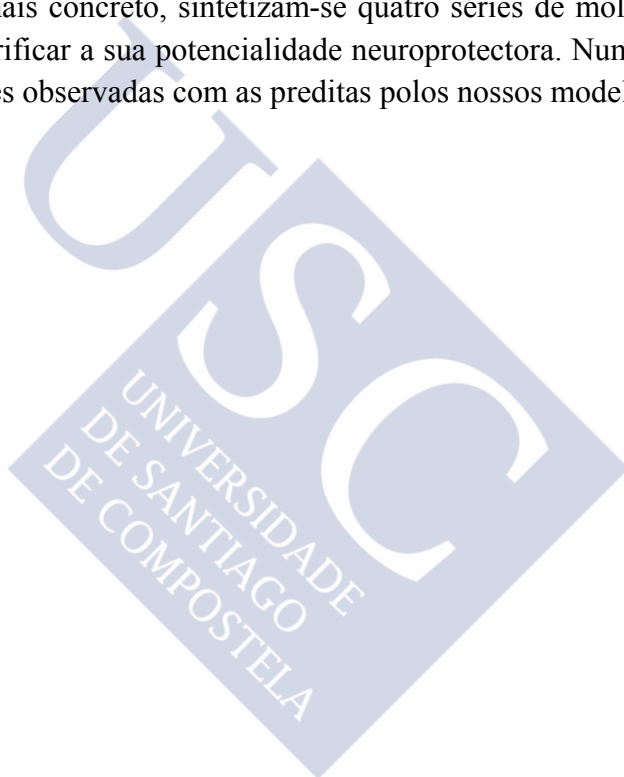
2. OBJETIVOS



2. OBJETIVOS

Como se explica previamente, nom se parte dum paradigma hipotético-dedutivo, polo que a Tese nom tem um objectivo específico trasladável a um enunciado em forma de Hipótese falseável. Pretende-se o uso de métodos estatístico-matemáticos para descobrir regularidades que permitam unificar e entender um grande número de ensaios experimentais. Estes ensaios reflitem interacções moleculares no sistema nervoso associadas fundamentalmente a neuroprotecção. A integraçom de resultados empíricos conduz a um modelo predictivo capaz de gerar, agora si, *hipóteses específicas* plausíveis acerca do comportamento de moléculas em circunstâncias e condições experimentais nom observadas.

Com um objectivo mais concreto, sintetizam-se quatro séries de moléculas relacionadas com a Rasagilina para verificar a sua potencialidade neuroprotectora. Numha segunda fase se contrastam as propriedades observadas com as preditas polos nossos modelos.







3. DISCUSSOM



3. DISCUSSOM

Nesta tese elaboram-se modelos QSAR de multiplexado cuja finalidade é prever interacções entre moléculas implicadas em neuroprotecção no SNC. Isto fai-se, por primeira vez, através de procedimentos de estatística clássica e algoritmos de redes neurais artificiais de última geração.

3.1 Modelos QSAR.

A literatura científica mostra modelos QSAR para prever propriedades neuroprotectoras para uma série de compostos e alvos moleculares mais limitados (Balaji et al., 2014; Bukhari et al., 2014; Rai et al., 2014). O presente trabalho abre esta perspectiva a um conjunto mais plural de condições empíricas e alvos, nom limitando-se a uma só via bioquímica.

Dado que a neurodegeneração implica múltiplas vias bioquímicas interconectadas, há uma vantagem óbvia numa aproximação assi. Uma só molécula candidata poderia ter múltiplos efeitos benéficos, ou uma intrincada combinação de acções opostas. Uma conjectura teórica poderia conduzir a um desenho experimental mais dirigido e fértil. Iguamente, um dos grandes reptos dos ensaios clínicos de neuroprotecção é a grande distancia entre o sucesso dos modelos animais e os decepcionantes resultados nos ensaios sobre humanos. Este desafio translacional poderia ser potencialmente superado polo nosso modelo de multiplexado. Dacordo às nossas predições, uma molécula com uma grande probabilidade de interacção com alvos animais e humanos deveria ser uma melhor candidata para os estudos experimentais. O nosso modelo pode prever os compostos cuja actividade biológica permanece útil em diferentes espécies e condições.

O melhor dos nossos modelos computacionais nom sobrepasa uma precisom do 85%. À hora de julgar este desempenho deveriamos ter em conta que valores superiores ao 75% som considerados aceitáveis nos modelos classificatórios de QSAR. Há diversas fontes de incertidume nos dados e o seu processo, que explicariam as possíveis limitações do nosso modelo:

1º Os dados biológicos som intrínsecamente sujeitos a um erro empírico considerável nas medidas.

2º O conjunto de dados abrange intencionadamente uma grande heterogeneidade de circunstancias empíricas, e isso claramente reduz a possibilidade de achar patrons comuns presentes através de todas as condições estudadas. A variabilidade molecular e genética intra- e interespecies é bem conhecida na dinâmica do sistema nervoso.

3º A pesar das assunções dos modelos QSAR, nalguns casos pequenos cambios na estrutura química das moléculas tenhem um impacto imprezível no mecanismo de acção de um composto. Isto poderia nom ser detectado polo nosso modelo. Devemos ter em conta

que qualquer descriptor molecular nom é mais do que umha versom quantitativa supersimplificada de umha arquitectura química complexa.

4° O dominio químico: O nosso modelo foi construido para alvos e moléculas implicados em interacçõs no sistema nervoso, nom poderia ser legitimamente extrapolado para compostos e interacçõs fora deste dominio concreto. No mesmo sentido, compostos com valores de descritores moleculares fora do rango dos valores numéricos utilizados no nosso modelo estám fora do alcanço do nosso modelo (Cherkasov et al., 2014).

À hora de estabelecer a prediçom, o nosso modelo assume um resultado dicotómico através dum umbral matemático. Isto nom implica automaticamente um significado biológico, nem tampouco o descarta. Umha molécula poderia ter umha probabilidade de interacçom alta com um alvo específico e a sua consequência biológica ser irrelevante, ainda que isto é improvável. Ademais, umha significaçom biológica positiva poderia ser devida a efeitos fisiológicos contrários dependendo do tipo de medida bioquímica concreta utilizada para cada prediçom. Em qualquer caso, o propósito do modelo nom é iluminar todas as subtilezas fisiológicas do interactoma do sistema nervoso, senom poupar tempo e recursos dirigindo os estudos experimentais ao uso de moléculas a priori mais prometedoras.

Às vezes, estabelecem-se correlaçõs falsas em modelos com um grande número de variáveis de entrada em relaçom ao número de observaçõs. Para evitar isto, no nosso modelo utilizamos um máximo de 10 variáveis predictoras (cinco por cada descriptor molecular) para mais de 8300 observaçõs. Tradicionalmente recomenda-se que a ratio de observaçõs versus número de variáveis seja igual ou superior a cinco. Outra precauçom a guardar quando se usam métodos QSAR é o uso de unidades ou pontos de corte inadecuados.

3.2 Eficiência dos diferentes descritores moleculares.

Os diferentes descritores moleculares (momentos espectrais de Markov, entropia de Shannon) nom semelham ter um impacto substancial no desempenho dos modelos baseados na análise discriminante tradicional. Polo contrário, no quarto trabalho objetivam-se diferenças apreçáveis entre os modelos construídos apartir destes dous descritores. Atingem-se melhores resultados com as redes neurais artificiais baseadas na entropia molecular frente às construídas com momentos espectrais (Tenorio-Borroto et al., 2013). É possível que a arquitectura específica das redes neurais, com um número maior de parâmetros que os modelos estatísticos tradicionais, as faga mais sensíveis às diferentes traduções numéricas das estruturas moleculares. Isto poderia ter consequências no futuro desenho destes modelos computacionais. Um tema interessante a explorar em futuros trabalhos seria achar a combinaçom mais produtiva entre os diversos descritores moleculares e cada um dos métodos preditivos considerados.

3.3 Análise Discriminante.

Todos os modelos implementados com técnicas estatísticas tradicionais fazem uso da análise discriminante linear aplicada a dois grupos. Esta é a forma mais simples de análise discriminante. Só é capaz de achar uma combinação linear das variáveis independentes para explicar a pertença dos casos a cada um dos grupos. Ademais, a estrutura dos dados deve satisfazer uma série de requisitos concretos:

- É necessário que haja pelo menos dois grupos, e para cada grupo são necessários dois ou mais casos.
- Nenhuma variável discriminante pode ser combinação linear de outras variáveis discriminantes, nem deve haver uma forte correlação entre estas variáveis explicativas do modelo.
- As matrizes de covariância dentro de cada grupo deve ser aproximadamente iguais.
- As variáveis contínuas devem seguir uma distribuição normal multivariada.
- As probabilidades a priori de pertença a um grupo devem ser identificáveis.
- As relações que se estabelecem entre as variáveis independentes e dependentes são lineares.
- Ademais a análise discriminante apresenta uma sensibilidade grande a valores extremos.

A transgressão destas suposições pode provocar testes de significação inválidos e comprometer a confiabilidade da inferência da função classificatória (Yoon et al., 1993).

A análise discriminante quadrática é um refinamento desta metodologia, menos exigente quanto à estrutura dos dados, e capaz de desvelar relações não lineares entre as variáveis. No nosso trabalho não nos servimos desta forma mais sofisticada de análise. Preferimos contrastar a exactidão da análise linear discriminante com modelos de aprendizado automático inspirados no funcionamento cerebral, aprofundando desse modo na nossa referência à Neuroinformática.

Nalguns casos, o desempenho dos modelos construídos com esta metodologia estatística tradicional foi aparentemente melhor do que se observou para modelos mais complexos baseados em redes neurais. No entanto, isto só se verificou para bases de dados notoriamente mais simples, com menor número de ensaios e casos incluídos. Quando se estabeleceram comparações legítimas entre métodos lineares e métodos não-lineares baseados na IA, a vantagem foi sempre para estes últimos.

3.4 Modelos de Redes Neurais.

As RNA podem modelar complexas relações não detectadas por métodos estatísticos mais simples, porque elas não estão forçadas por uma associação matemática assumida a priori entre as variáveis, assim que podem determinar por aproximação qualquer função

intrincada nom linear. Isto é de especial valor dada a natureza do nosso estudo, já que a dinâmica das interacções bioquímicas está pragada de fenómenos nom lineares complexos.

Os modelos nom lineares de redes neurais artificiais dam um límite superior de precisom semelhante (85%). Todo isto apesar do diverso software utilizado, arquitectura e metodologia para entrená-las. Umha explicação provável é que todas elas esgotam o potencial para extraer um sinal biológico coerente da base de dados. Esta coerência parece dar umha ideia do robusto da aproximação metodológica.

Ademais das suas múltiplas vantagens, cumpre analisar também as circunstâncias nas que estes modelos podem apresentar dificuldades. Começaremos polo problema da generalização. O sobre-ajuste aparece quando um modelo matemático aprende ruído mais do que a relação latente entre as variáveis, comprometendo a sua capacidade de generalização. Neste caso, os dados utilizados para construir o modelo dam um desempenho optimista e pouco realista, nom corroborável ao utilizar o modelo para predizer novos dados (Lawrence and Giles, 2000). Para evitar o sobre-ajuste, partimos a base de dados inicial em três subconjuntos. Um subconjunto de *treinamento* serve para ajustar os parámetros da rede, quer dizer, os pesos das conexons entre as unidades de processo. Isto fai-se confrontando as entradas com as saídas esperadas e gerando múltiples modelos alternativos. Um subconjunto chamado de *selecçom* escolhe o melhor desses modelos avaliando os hiperparámetros dos modelos em competiçom. Este hiperparámetros som características estruturais das redes como o número de capas ou o número de nodos da rede. Por último, um subconjunto de dados permite a validação externa do modelo. Isto permite conhecer o desempenho real ante dados externos nom utilizados para a construção do modelo. No trabalho publicado pode-se apreçar que o desempenho das redes neurais para cada umha das três subdivisons dos dados originais é muito semelhante, portanto descartando a possibilidade de sobreajuste nos nossos modelos.

Umha das maiores críticas recebidas polas redes neurais artificiais consiste em ressaltar que elas funcionam como caixas-negras e nom podem ser utilizadas para identificar associações causais. A ausência de um conhecimento mecanístico acerca da contribuição individual das variáveis preditivas e a sua interacção pode comprometer a interpretação dos modelos em termos puramente químicos. Isto acentua-se tendo em conta que muitos dos descriptores moleculares utilizados para elaborar modelos QSAR nom temhem relação com a presença ou ausência de certos grupos funcionais concretos, senom que consistem por exemplo numha abstracção topológica que pode negligir propriedades químicas relavantes. Isto é, nom só a relação entre as variáveis predictoras e os resultados se obscurece, senom que as mesmas variáveis som em si mesmas umha abstracção nom fácilmente interpretável. Com respeito à interpretabilidade das redes neurais, no momento actual há várias metodologias disponíveis capazes de superar estas dificuldades (Olden et al., 2004). Deste modo, se o modelo a elaborar tem entre os seus requerimentos umha boa interpretabilidade em termos de grupos funcionais químicos, umha reflexom cuidadosa previa acerca de que descriptores moleculares devem ser escolhidos poderia superar estas dificuldades.

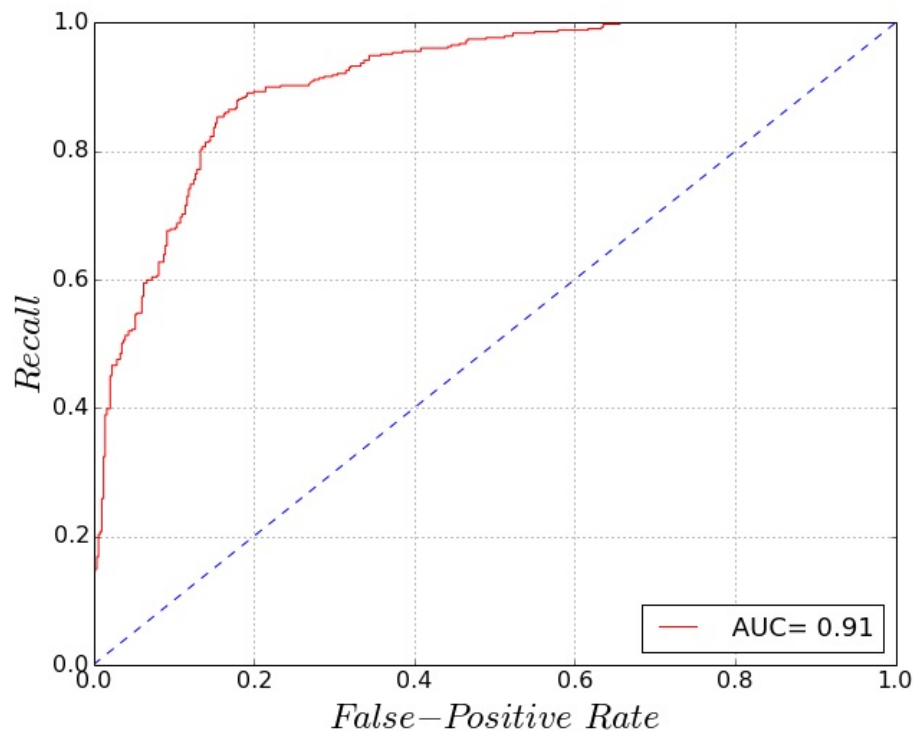


Figura 3. Curva COR (Característica de Operação do Receptor) para o modelo de Aprendizado Profundo

Umha outra consideraçom relevante tem a ver com que um procedimento específico de aprendizado só pode ser óptimo para um conjunto restringido de todos os problemas concebíveis. Em aprendizagem automática, a metáfora do *No free Lunch* sublinha que nengum algoritmo de aprendizado é intrinsecamente e universalmente superior. Se medimos o desempenho sobre toda classe de problemas, todos os procedimentos mostram um desempenho similar. Correlativamente, para um conjunto de dados específico os diversos algoritmos exibem vantagens e deficiências igualmente diversas. As nossas necessidades definirám a prioridade que assignemos em cada momento a diferentes características do modelo como precisom, velocidade, complexidade e interpretabilidade. Isto guiará a escolha entre os diferentes algoritmos. No obstante, a priori nom é possível conhecer como se comportaram os diferentes modelos sobre o nosso conjunto de dados (Ho and Pepyne, 2002). Portanto, na nossa análise fazemos uso de métodos da estatística multivariáve tradicional como a análise discriminante, assi como diversas arquiteturas de redes neurais artificiais.

A análise discriminante e as redes neurais lineares som ambos similares, mais simples, rápidos e interpretáveis, mas mostram um desempenho pior. Os modelos mais complexos como as redes neurais multicapa elevam substancialmente a precisom ao custo de umha grande lentitude no seu treinamento. O modelo baseado nos algoritmos do aprendizado profundo é o que apresenta o melhor compromisso entre as diferentes características: relativamente mais rápido na fase de treinamento e com umha precisom notável.

Os métodos baseados no aprendizado profundo terám um notável impacto no campo da descoberta de fármacos como o tenhem hoje em tarefas perceptivas (Gawehn et al., 2015).

Mas deve-se aprender a refinar o pré-processado de dados. O nosso primeiro passo consistiu na transformação de informação altamente correlacionada da estrutura de compostos químicos a uma forma mais simplificada e numérica representada por descritores moleculares baseados na teoria de cadeias de Markov. Provavelmente esta transformação faz sentido para reduzir os requerimentos computacionais de algoritmos mais simples, mas adicionalmente impede que os novos algoritmos profundos atinjam todo o seu potencial. Como se sublinha em parágrafos prévios, será chave a identificação ou elaboração de descritores moleculares mais adequados para a nova geração das redes neurais artificiais. Idealmente, eles seriam capazes de capturar a informação altamente correlacionada das estruturas moleculares activas no sistema nervoso central, por exemplo grupos funcionais relevantes nos receptores moleculares ou moléculas transdutoras de sinais. Isto requerirá sem dúvida a chegada de conhecimento experto acerca da bioquímica cerebral.

3.5 Análise de Redes.

Outra forma suplementar de comparar os achados experimentais com os nossos modelos parte da análise de redes. A rede de interações fármaco-diana reconstruída a partir do modelo detalhado no terceiro trabalho foi comparada com a rede criada a partir dos resultados experimentais. A conclusão foi que o nosso modelo predizia eficientemente não só o conjunto geral de ligações, mas também os padrões de conexão reflectidos em parâmetros topológicos de centralidade e entropia. A análise de redes é um ramo das matemáticas com múltiplas e crescentes aplicações nas neurociências. Uma das suas aplicações inovadoras consiste em observar diferentes padrões de conectividade cerebral por meio de sofisticadas técnicas de neuroimagem, o que oferece uma perspectiva inédita das patologias neurológicas. No nosso trabalho emprega-se como um método matemático complementar de validação externa das nossas previsões. A representação gráfica ilustra-se na **Figura 4**.

Todas estas aplicações nascem da premissa de que a estrutura de uma rede e o seu padrão de conexões determina a sua funcionalidade. Ainda que esta assumption verifica-se amiúde em trabalhos experimentais, cumpre ser consciente dos seus limites, e não deve ser assumida sem cautela já que não sempre se cumpre (Ingram et al., 2006).

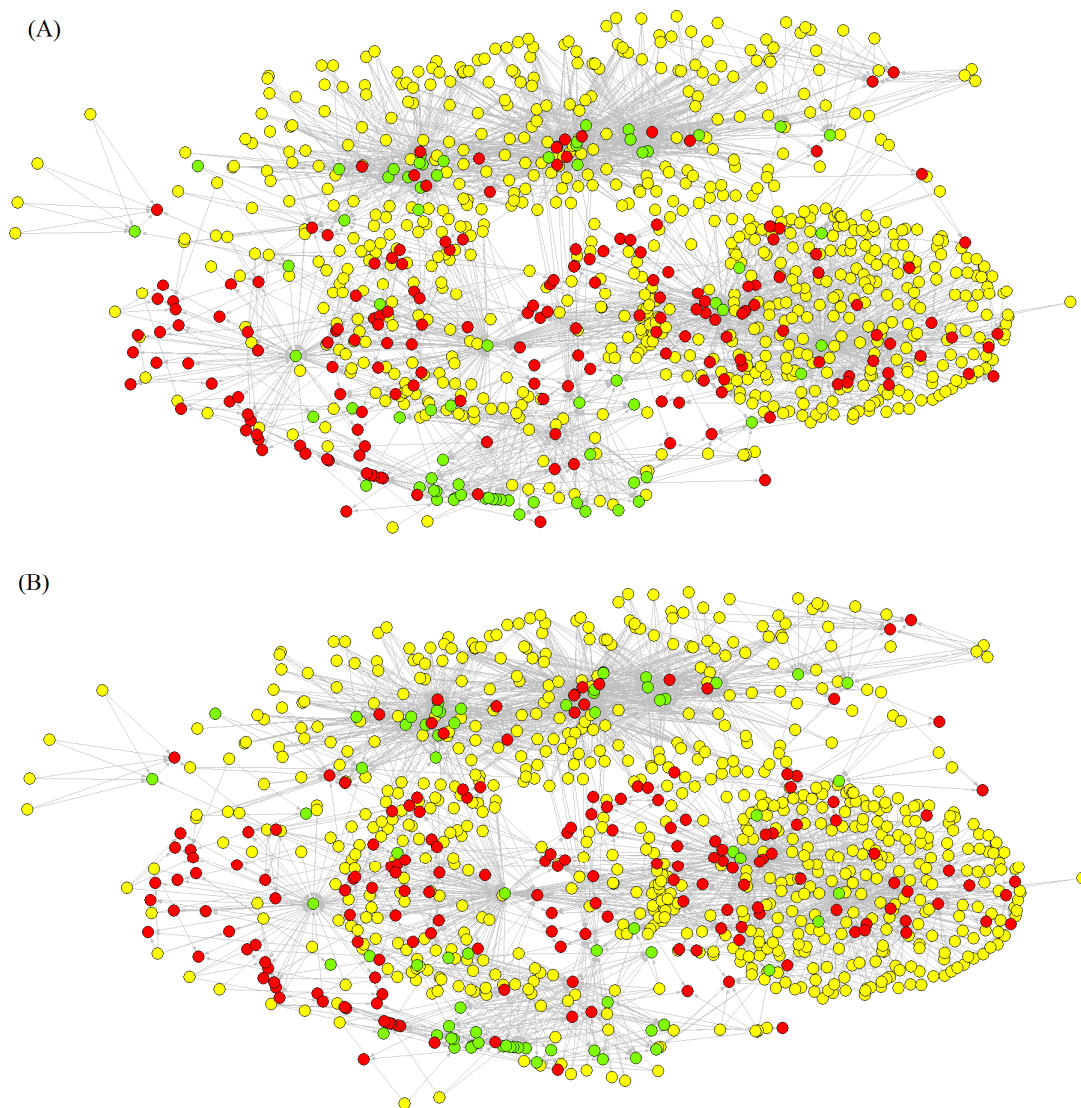


Figura 4. Redes Multi-alvo, Multiescala, e Multi-saída de moléculas implicadas em neuroprotecção (amarelo), alvos (vermelho), e ensaios farmacológicos(verde). Observada (A) vs. Predita (B)

3.6 Comparativa de achados experimentais e teóricos.

A síntese das diferentes séries de compostos relacionados com a rasagilina levou-se a cabo tendo em conta possíveis modificações moleculares com sentido biológico. Por exemplo, o grupo carbamato foi engadido a toda uma série de derivados. Este grupo funcional tem importância reconhecida em fármacos inibidores de acetilcolinesterasa, benéficos na Doença de Alzheimer.

Aliás, exploraram-se diferentes derivados isoméricos, como formas *cis/trans*, e diferentes formas quirais. Desde o desgraçado episódio da Talidomida, a isomeria óptica reconhece-se como um factor fundamental à hora do desenho de fármacos, já que toda a lógica molecular da vida está construída sobre umhas formas enantioméricas específicas. A interacção com

umha molécula, ou a sua forma especular, produz efeitos fisiológicos substancialmente diferentes. Isto também se verifica nas nossas séries quando se estuda o perfil neuroprotector dos diferentes derivados da rasagilina.

Dentro de cada umha das quatro séries de compostos destacam umhas moléculas pola sua maior actividade. Na primeira série, um derivado trans monopropargil mostra umha actividade neuroprotectora alta para os 3 tipos de ensaio comentados na secçom introdutória. O nosso modelo prediz para este composto alta probabilidade de interacçom com receptores de Glutamato, para além de receptores de GABA e acetil-colina. Na segunda série, destaca um derivado dipropargil como activo em cada um dos ensaios, e a nossa prediçom teórica destaca como diana mais provável a sintase de óxido nítrico. Para os compostos com patrom de substituiçom 1,2 da terceira série, os ensaios nom mostram neuroprotecçom relevante ante a adiçom de glutamato. Congruentemente, o nosso modelo teórico prediz actividade importante sobre os receptores 5-HT₃, implicados em neuroprotecçom na literatura científica. Umha hipótese plausível é que a mudança no patrom de substituiçom de 1,3 a 1,2 determine a diferente afinidade por cada um desses receptores. Dentro da série de derivados de TVP1022 (análogo de rasagilina), um derivado carbamato mostrou empíricamente umha actividade superior à média contra a toxicidade induzida por glutamato (**Fig. 5**). Isto de novo está dacordo com as nossas prediçons, um valor maior de $-\text{Log}(\text{IC}_{50})$ implica umha maior inibiçom dos receptores de glutamate reduzindo a excitotoxicidade (Lau and Tymianski, 2010).

Os resultados empíricos e teóricos avalam a estratégia de síntese química, já que dentro da diversidade das moléculas sintetizadas é possível achar perfis igualmente diversos enquanto à sua actividade neuroprotectora. Em termos gerais, as correspondências entre os ensaios experimentais e as nossas prediçons teóricas poderiam indicar a plausibilidade biológica dos nossos modelos. No entanto, nom se deve perder a perspectiva global da finalidade deste trabalho. Cada umha destas plausíveis hipótesis que emerge das nossas prediçons deveria ser testada em experimentos ad hoc numha fase posterior.

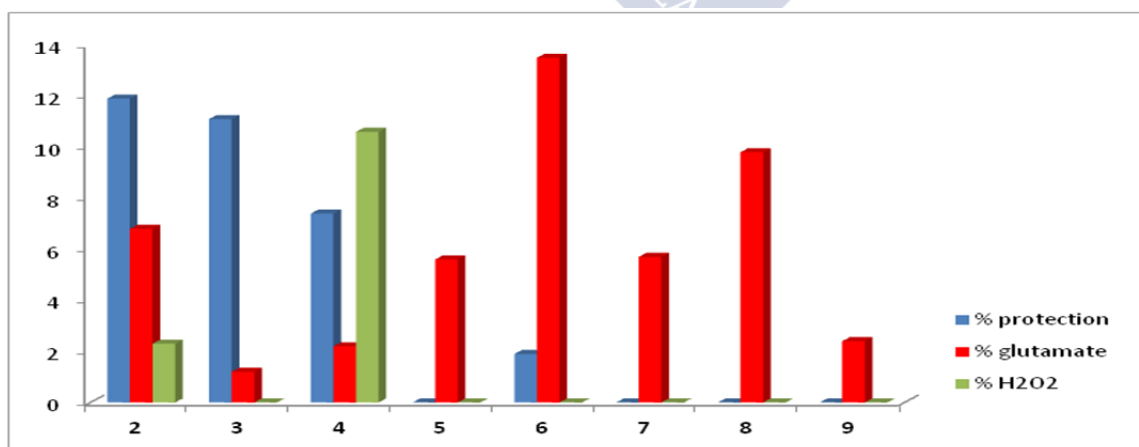


Figura 5. Efeito neuroprotector dos compostos derivados de TVP1022 (análogos-S de rasagilina)



4. CONCLUSOM



4. CONCLUSOM

As patologias neurológicas, nomeadamente enfermidades neurodegenerativas e cerebrovasculares, som condições médicas devastantes e comuns que implicam um reto crítico à saúde pública. Os fundamentos da descoberta de fármacos para outras áreas terapéuticas som igualmente aplicáveis aos distúrbios neurológicos. A química combinatória e a crivagem de alto rendimento, como a automatização e minituarização dos ensaios, tenhem incrementado exponencialmente os dados disponíveis nos modernos programas de descoberta de fármacos. Non obstante, o facto é que a neuroprotecçom é ainda um dos melhores exemplos da divergência entre os avances da investigação básica e a sua traslaçom efectiva à prática clínica.

Neste contexto, enlenteçer a progressom das doenças neurodegenerativas deveria ser um objectivo pragmático a atingir nas próximas décadas. Isto requirirá umha convergencia de disciplinas complementares. Apesar dos incríveis avances nas técnicas de crivagem de alto rendimento, a quantidade masiva de novos dados empíricos no desenho de fármacos nom minimizará a relevancia das aproximaçoms teóricas. Umha das vantagens dos modelos teóricos é a sua capacidade de integrar e dar sentido unificado a numerosos dados empíricos inconexos. Desta forma, possuem o potencial para superar alguns dos problemas planteados na investigação translacional, como as diferenças entre espécies. Um fármaco com um perfil teórico de açom multi-diana, activo em diversas espécies e condições, provavelmente seria um candidato prioritário para avaliar em ensaios empíricos.

O incremento na disponibilidade de dados experimentais associa-se também a um incremento substantivo do poder computacional para o seu processo. O uso de Unidades de Processamento Gráfico especificamente desenhadas para cómputos científicos, e comercializadas a um custo razoável, é já umha realidade. Aliás, novos e potentes métodos de IA, como a aprendizagem profunda, som agora implementados baixo programas de código aberto universalizando o seu potencial.

A interferência molecular das intrincadas perturbaçoms neuroquímicas iniciadas pola neurodegeneraçom e o dano nervoso continuará sendo um objectivo preferente da investigação biomédica. Os métodos neuroinformáticos serám umha ferramenta valiosa para melhorar a nossa aproximaçom a este elusivo alvo.





5. RESUMO



5. RESUMO

Introdução

A compreensão e o tratamento das doenças neurodegenerativas constituem ainda um desafio apesar da pesquisa intensiva. Novos algoritmos computacionais inspirados no processamento cerebral da informação e a disponibilidade pública de dados experimentais permitem uma nova abordagem para este enigma.

Métodos

A partir da base de dados ChEMBL construímos um conjunto de dados unificado de ensaios independentes sobre interações de fármaco-alvo implicadas principalmente na neuroprotecção. Logo transformamos as características moleculares dos compostos neuroquímicos a números através de diferentes descritores moleculares, usando operadores de média móvel de Box-Jenkins para incorporar informação categórica sobre subconjuntos específicos de condições experimentais. Posteriormente, aplicamos a análise discriminante linear e diversos métodos de redes neurais artificiais para inferir modelos que resumem todas estas observações específicas. Ao mesmo tempo, uma nova série de derivados de rasagilina foram sintetizados, validando a sua actividade neuroprotectora empiricamente e teoricamente através do nosso modelo.

Resultados e Discussão

Estes modelos têm o potencial de ultrapassar certas limitações experimentais porque podem prever o resultado provável de uma interacção fármaco-alvo em um conjunto generalizado de condições empíricas. De todos os métodos analisados, os algoritmos baseados na aprendizagem profunda apresentaram o melhor desempenho, com uma precisão de 85%, e um tempo de treinamento mais rápido. Os diferentes derivados de rasagilina mostraram um perfil diferente de actividade neuroprotectora empiricamente, e esta actividade foi congruente com as previsões teóricas deduzidas a partir do nosso modelo.

Conclusão

Tirando proveito de métodos computacionais inspirados no processo cerebral da informação construímos modelos do interactoma cerebral. Na nossa opinião, os métodos de aprendizagem profunda terão um impacto notável neste campo num futuro próximo.

Palavras-chave: neuroprotecção, interactoma, neuroinformática, aprendizagem profunda, síntese assimétrica, relação quantitativa estrutura-actividade

Abstract

Background

Understanding and treatment of neurodegenerative disorders are still an elusive challenge despite intensive research. New brain-inspired computational algorithms and publicly available experimental data allow a new approach for this conundrum.

Methods

From the ChEMBL database we build an integrated dataset of unrelated assays about drug-target interactions involved mainly in neuroprotection. In order to make a QSAR model we transformed the molecular features of neurochemicals to numbers through different molecular descriptors, and used Box-Jenkins moving average operators of these values to incorporate categorical information about specific subsets of experimental conditions. Then, we applied linear discriminant analysis and diverse artificial neural networks methods to infer a collection of models that summarize all these specific observations. At the same time, a new series of rasagiline derivatives were synthesized. We validated their neuroprotective activity empirically over neuron cultures, and theoretically implementing our model.

Results and Discussion

These models can predict the probable outcome of a drug-target interaction in a generalized set of empirical conditions not tested yet. It has the potential of overcome some of the limitations of individual clinical trials. Of all the methods analyzed, deep learning algorithms presented the best performance, with an accuracy of 85% and a faster training time. The diverse rasagiline derivatives showed a different profile of neuroprotective activity empirically, and this activity was congruent with the theoretical predictions deduced from our model.

Conclusion

We take advantage of computational methods roughly inspired by brain to illuminate the complicated molecular interactome of brain. In our opinion, these brain-inspired deep learning methods will have a profound impact in this field in the near future.

Keywords: neuroprotection, interactome, neuroinformatics, deep learning, asymmetric synthesis, QSAR

Introducción

La comprensión y el tratamiento de las enfermedades neurodegenerativas constituyen todavía un difícil reto a pesar de intensas investigaciones. Nuevos algoritmos computacionales inspirados en el procesamiento cerebral de la información y la disponibilidad pública de datos experimentales permiten un nuevo enfoque para este rompecabezas.

Métodos

A partir de la base de datos ChEMBL elaboramos un conjunto de datos integrado de ensayos no relacionados sobre interacciones fármaco-diana involucradas principalmente en la neuroprotección. Seguidamente transformamos las características moleculares de los compuestos neuroquímicos a números a través de diferentes descriptores moleculares, utilizando operadores de media móvil de Box-Jenkins para incorporar información categórica sobre subconjuntos específicos de condiciones experimentales. Posteriormente, aplicamos el análisis discriminante lineal y diversos métodos de redes neuronales artificiales para inferir un conjunto de modelos que explican todas estas observaciones específicas. Al mismo tiempo, sintetizamos una nueva serie de derivados de rasagilina, cuya actividad neuroprotectora se validó empírica y teóricamente utilizando nuestros modelos.

Resultados y Discusión

Estos modelos pueden predecir el resultado probable de una interacción fármaco-diana en un conjunto generalizado de condiciones no evaluadas experimentalmente. Ello tiene el potencial de superar algunas de las limitaciones de los ensayos individuales. De todos los métodos analizados, los algoritmos basados en aprendizaje profundo tuvieron el mejor desempeño, con un precisión del 85%, y un tiempo de entrenamiento más rápido. Los diversos derivados de la rasagilina sintetizados mostraron un perfil de actividad neuroprotectora diferente, y esta actividad fue consistente con las predicciones teóricas deducidas a partir de nuestro modelo.

Conclusión

En suma, tomando ventaja de los métodos computacionales inspirados en el proceso cerebral de la información construimos modelos precisos del interactoma cerebral. En nuestra opinión, los métodos de aprendizaje profundo tendrán un impacto sustancial en este campo en el futuro próximo.

Palabras clave: neuroprotección, interactoma, neuroinformática, aprendizaje profunda, síntesis asimétrica, relación cuantitativa estructura-actividad





6. ABREVIATURAS



6. ABREVIATURAS

QSAR: Em inglês, Relaçom Quantitativa Estrutura-Actividade (Quantitative Structure Activity Relationship).

SNC: Sistema Nervoso Central

SMN: Survival Motor Neuron

IA: Inteligência Artificial

RNA: Redes Neurais Artificiais







7. BIBLIOGRAFÍA GERAL



7. BIBLIOGRAFÍA GERAL

- Balaji, B., Hariharan, S., Shah, D. B., Ramanathan, M., 2014. Discovery of potential and selective COX-1 inhibitory leads using pharmacophore modelling, in silico screening and in vitro evaluation. *Eur J Med Chem* 86, 469-480.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., 2011. Time series analysis: forecasting and control. John Wiley & Sons.
- Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science* 16, 199-231.
- Bukhari, S. N., Jantan, I., Masand, V. H., Mahajan, D. T., Sher, M., Naeem-ul-Hassan, M., Amjad, M. W., 2014. Synthesis of alpha, beta-unsaturated carbonyl based compounds as acetylcholinesterase and butyrylcholinesterase inhibitors: characterization, molecular modeling, QSAR studies and effect against amyloid beta-induced cytotoxicity. *Eur J Med Chem* 83, 355-365.
- Csermely, P., Korcsmáros, T., Kiss, H. J., London, G., Nussinov, R., 2013. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacology & therapeutics* 138, 333-408.
- Cummings, J. L., Morstorf, T., Zhong, K., 2014. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther* 6, 37.
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., 2014. QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57, 4977-5010.
- De Schutter, E., Ascoli, G. A., Kennedy, D. N., 2012. Ten years of neuroinformatics. *Neuroinformatics*, 1-2.
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J. P., 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40, D1100-1107.
- Gawehn, E., Hiss, J. A., Schneider, G., 2015. Deep Learning in Drug Discovery. *Molecular Informatics*.
- Glazko, G., Veeramachaneni, V., Nei, M., Makalowski, W., 2005. Eighty percent of proteins are different between humans and chimpanzees. *Gene* 346, 215-219.
- Hannawi, Y., Smirnakis, S. M., 2013. Emerging subspecialties: neuroinformatics. *Neurology* 80, e166-168.
- Haykin, S. S., Haykin, S. S., Haykin, S. S., Haykin, S. S., 2009. Neural networks and learning machines. Pearson Education Upper Saddle River.
- Hill, T., Lewicki, P., Lewicki, P., 2006. Statistics: methods and applications: a comprehensive reference for science, industry, and data mining. StatSoft, Inc.
- Hinton, G. E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18, 1527-1554.

- Ho, Y.-C., Pepyne, D. L., 2002. Simple explanation of the no-free-lunch theorem and its implications. *Journal of optimization theory and applications* 115, 549-570.
- Ingram, P. J., Stumpf, M. P., Stark, J., 2006. Network motifs: structure does not determine function. *BMC genomics* 7, 108.
- Jain, K. K., 2011. *The handbook of neuroprotection*. Springer Science & Business Media.
- Kell, D. B., Oliver, S. G., 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26, 99-105.
- Lau, A., Tymianski, M., 2010. Glutamate receptors, neurotoxicity and neurodegeneration. *Pflügers Archiv-European Journal of Physiology* 460, 525-542.
- Lawrence, S., Giles, C. L., 2000. Overfitting and neural networks: conjugate gradient and backpropagation. *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on. IEEE*, pp. 114-119.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436-444.
- Nurisso, A., Simoes-Pires, C., Martel, S., Cressend, D., Guillot, A., Carrupt, P. A., 2012. How to increase the safety and efficacy of compounds against neurodegeneration? A multifunctional approach. *Chimia (Aarau)* 66, 286-290.
- Olden, J. D., Joy, M. K., Death, R. G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* 178, 389-397.
- Prince, M., Wimo, A., Guerchet, M., Ali, G., Wu, Y., Prina, M., 2015. *World Alzheimer Report 2015. The global impact of dementia. An analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International, London.
- Rai, G., Joshi, N., Jung, J. E., Liu, Y., Schultz, L., Yasgar, A., Perry, S., Diaz, G., Zhang, Q., Kenyon, V., Jadhav, A., Simeonov, A., Lo, E. H., van Leyen, K., Maloney, D. J., Holman, T. R., 2014. Potent and selective inhibitors of human reticulocyte 12/15-lipoxygenase as anti-stroke therapies. *J Med Chem* 57, 4035-4048.
- Rascol, O., Fitzer-Attas, C. J., Hauser, R., Jankovic, J., Lang, A., Langston, J. W., Melamed, E., Poewe, W., Stocchi, F., Tolosa, E., 2011. A double-blind, delayed-start trial of rasagiline in Parkinson's disease (the ADAGIO study): prespecified and post-hoc analyses of the need for additional therapies, changes in UPDRS scores, and non-motor outcomes. *The Lancet Neurology* 10, 415-423.
- Roy, K., Kar, S., Das, R. N., 2015. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Academic Press.
- Tenorio-Borroto, E., García-Mera, X., G Penuelas-Rivas, C., C Vasquez-Chagoyan, J., J Prado-Prado, F., Castanedo, N., Gonzalez-Diaz, H., 2013. Entropy model for multiplex drug-target interaction endpoints of drug immunotoxicity. *Curr Top Med Chem* 13, 1636-1649.
- Todeschini, R., Consonni, V., 2008. *Handbook of molecular descriptors*. John Wiley & Sons.
- Yoon, Y., Swales Jr, G., Margavio, T. M., 1993. A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, 51-60.



8. ARTIGOS PUBLICADOS



TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases

F. Luan, M. N. D. S. Cordeiro, N. Alonso, X. García-Mera, O. Caamaño, **F. J. Romero-Durán**, M. Yañez and H. González-Díaz

Bioorganic & Medicinal Chemistry 2013, 21, 1870–1879

<http://dx.doi.org/10.1016/j.bmc.2013.01.035>

<http://www.sciencedirect.com/science/article/pii/S0968089613000746>

Model for High-Throughput Screening of Multitarget Drugs in Chemical Neurosciences: Synthesis, Assay, and Theoretic Study of Rasagiline Carbamates

N. Alonso, O. Caamaño, **F. J. Romero-Durán**, F. Luan, M. N. D. S. Cordeiro, M. Yañez, H. González-Díaz and Xerardo García-Mera

ACS Chem. Neurosci. 2013, 4, 1393–1403

<http://dx.doi.org/10.1021/cn400111n>

<http://pubs.acs.org/doi/abs/10.1021/cn400111n>

Prediction of Multi-Target Networks of Neuroprotective Compounds with Entropy Indices and Synthesis, Assay and Theoretical Study of New Asymmetric 1,2-Rasagiline Carbamates

F. J. Romero-Durán, N. Alonso, O. Caamaño, X. García-Mera, M. Yañez, F. J. Prado-Prado and H. González-Díaz

Int. J. Mol. Sci. 2014, 15, 17035-17064

<http://10.3390/ijms150917035>

<http://www.mdpi.com/1422-0067/15/9/17035>

Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives

F. J. Romero-Durán, N. Alonso, M. Yañez, O. Caamaño, X. García-Mera and H. González-Díaz

Neuropharmacology 2016, 103, 270e278

<http://dx.doi.org/10.1016/j.neuropharm.2015.12.019>

<http://www.sciencedirect.com/science/article/pii/S002839081530215X>