Grasas y Aceites
Vol. 54. Fasc. 2 (2003), 188-193

Technology Information

The new database "Seed oil Fatty Acids" (SOFA)

By **B. Matthäus*[1], K. Aitzetmüller[2]** and **H. Friedrich[3]**

[1]Federal Center for Cereal, Potato and Lipid Research, Institute for Lipid Research, Piusallee 68/76, 48147 Münster, Germany
[2]Federal Center for Cereal, Potato and Lipid Research, former Institute for Chemistry and Physics of Lipids, Piusallee 76, 48147 Münster, Germany
[3]The German Centre for Documentation and Information in Agriculture (ZADI), Villichgasse 17, 53177 Bonn, Germany

RESUMEN

La nueva base de datos "Ácidos Grasos de Semillas Oleaginosas" (SOFA).

Mas de 30 años de información sobre la composición en ácidos grasos de semillas de plantas silvestres ha sido recogida de la literatura farmacéutica, botánica y química por el antiguo Institute for Chemistry and Physics del Federal Center for Cereal, Potato and Lipid Research en Münster. Desde hace varios meses la base de datos está ya disponible sin cargo alguno en Internet (http:\\www.bagkf.de/sofa).

Hasta ahora más de 18.000 tablas diferentes con alrededor de 110.000 datos individuales pueden obtenerse a partir de esa base de datos mediante diferentes tipos de búsquedas. Una panorámica de la aplicación de las diferentes posibles formas de búsquedas así como algunos ejemplos de búsquedas se dan en este artículo.

PALABRAS-CLAVE: Base de datos – Composición en ácidos grasos – Aceites de semillas de plantas.

SUMMARY

The new database "Seed Oil Fatty Acids" (SOFA).

More than 30 years information about the fatty acid composition of wild plant seeds was collected from the appropriate pharmaceutical, botanical and chemical literature by the former Institute for Chemistry and Physics of Lipids of the Federal Center for Cereal, Potato and Lipid Research in Münster. Since it was very difficult to search this unique source of information, the collection was transferred into an electronically searchable database. This plan was supported by financial assistance of the German Ministry of Consumer Protection, Food and Agriculture. Since some month the database is available in the internet (http:\\www.bagkf.de/sofa) free of charge.

Up to now more than 18.000 different tables with about 110.000 individual data are recallable from the database, by different types of search forms. An overview about the application of the different search forms as well as some examples for searches is given in this publication.

KEY-WORDS: Database - Fatty acid composition - Plant seed oils.

1. INTRODUCTION

Normally seeds of plants are rich sources of various kinds of lipidic compounds including fatty acids, tocopherols, triglycerides, phospholipids, sphingolipids and sterols. Depending on the plant, these constituents are present in different proportions in the seeds.

Fats and oils have an outstanding importance, not only in nutrition and pharmacy, but also in the field of renewable resources. The reason is that this raw material is easily available and regarding the fatty acid composition a huge variety exists. Just this great variety enables a wealth of chances for different applications.

In the field of renewable resources the oleochemical industry is interested in seeds with one fatty acid with an amount as high as possible. Such oils can be used as precursors or intermediates for the synthesis of high-quality products without expensive purification of the raw material.

Further on, plant breeders are interested in the knowledge of the occurrence of interesting fatty acids in plants, because additionally to the fatty acid the seeds contain the appropriate enzyme systems, necessary for the production of this fatty acid. By means of the knowledge of the occurrence of interesting fatty acids in plant kingdom and modern biotechnology and genetic engineering it is possible to isolation and transfer the appropriate gene sequences in another easy to grow cultivar. Therefore the knowledge about the fatty acid composition of wild plants is of great interest.

Unfortunately the information about the fatty acid composition of plant seeds is scattered in the literature. One important source for such information is the New Crops database of the National Center for Agricultural Utilization Research (NCAUR) [1] which contains chemical and physical information of 15,738 seed accessions. This database is accessible on the internet since some years. Another important source for information about the fatty acid composition of seed oils is the database of the AOCS [2]. This database is available as printout, but also as electronically searchable database on CD-ROM and contains information of more than 150 more or less common used seed oils.

One disadvantage of both databases is that only usual fatty acids, such as oleic acid or tetracosenoic acid (24:1) are presented. Unusual fatty acids with interesting structure elements are not available. But just these unusual fatty acids are interesting for the use in oleochemical or pharmaceutical industry. Further on it is not possible to search for individual structure elements such as cyclopropen or allyl, which could be interesting for certain questions of chemists or botanists.

2. AVAILABILITY OF THE DATABASE

The database SOFA is available in the Internet using the address http://www.bagkf.de/sofa without any restriction by use of a password and free of charge.

3. STRUCTURE OF THE DATABASE

The database works with an efficient document database management system (DDBMS) and the data available in the database is filed on three different records (Fig. 1). One record contains information about the plant, its botanical affiliation to a family as well as bibliographical information about the reference. The second record comprises all information regarding more than 600 molecules – fatty acids, sterols, triglycerides and tocopherols, with appropriate information about the IUPAC name, CA registry number as well as some systematic, partially systematic and trivial names for each fatty acid. So the database is able to work with the input of different names for one fatty acid. The third records combine the information available in both of the other records and add the information about the contents of the molecules.

4. Δ-NOTATION

One of the most important aspects of the database is the possibility to search for defined structure elements of fatty acids. For the implementation of this feature a homogeneous syntax, which describes the structural elements of the different fatty acids was established. This syntax, named as "Δ-notation", counts the position of the double bond of a fatty acid from the acid group, and not, as usual for the ω-style from the last methyl group. An example is *arachidonic acid,* called as *20:4*-Δ-5c,8c,11c,14c, according to the definition of the Δ-notation instead of *20:4n-6* in the ω-style. With this term a lot of information was given to the database, which could be used for specific questions. The fatty acid given in the example has 20 carbon atoms and four double bonds with cis configuration in position 5, 8, 11 and 14. All these information are searchable individual or in combination.

This most important tool of the database is suitable to search for chemically interesting structure elements of fatty acids, such as *OH-*, *keto-* or *cyclopentene*, but also *en*yn* for fatty acids containing double and triple bonds in one molecule could be searched. Exact positions of double or triple bonds are searchable, using questions such as *-12-yn* or *Delta-5t*, which search for molecules with a triple bond in position 12 or trans isomers in position 5. Further examples for the rules of the Δ–notations are given in figure 2.

5. SEARCH FORMS OF THE DATABASE

The database allows three different ways of search. In the following these ways are described.

5.1. Search for molecules

Using the form *Molecules-Notation* it is possible to search for information about fatty acids available in the database. In this form information about CA registry number, short name, Δ-notation, systematic name or annotation are searchable. The term is

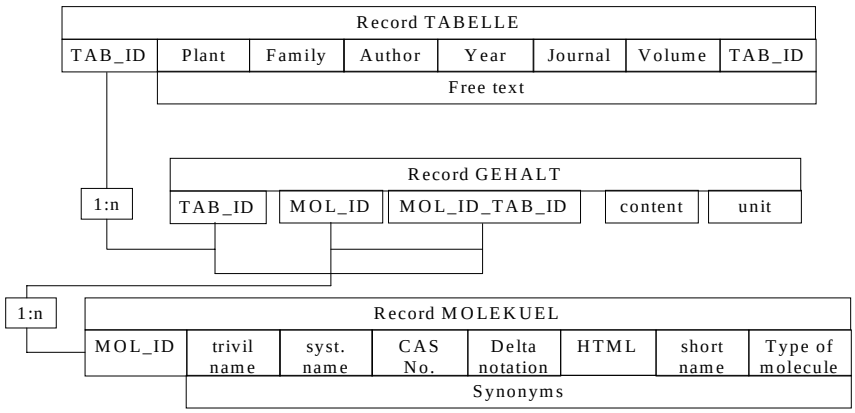

Figure 1
Structure of the database.

Each fatty acid is presented as a computer readable string of characters

Examples:

1. Basic structure of the fatty acids: number in front and behind a colon
2. c (*cis*), t (*trans*), a (*acetylen*) = type of double and triple bond, respectively
3. position of multiple bonds in the molecule: number behind the Δ
4. Further specifications:

-O-	=	epoxy group
-OH	=	hydroxy group
-O=	=	keto group
-cpe-	=	cyclopropene ring
-cpa-	=	cyclopropane ring

Examples for the so-called Δ-notation" of fatty acids:

Linolic acid	=	18:2 Δ 9c,12c
Vernolic acid	=	12,13-O-18:1 Δ 9c
Agonandric acid	=	8-OH-18:2 Δ 9a,11t
Stearidonic acid	=	18:4 Δ 6c,9c,12c,15c
2-Hydroxy sterculic acid	=	2-OH-9,10-cpe-19:1

Figure 2
Rules and examples for the "Δ-Notationen" of fatty acids.

given into the computer either as complete word or as part of a word, truncated by asterisks (*) in front and/or behind. So it is possible to search for groups of fatty acids or for special structure elements of interest, as described before. The result of a search in this form is a list of fatty acids (fig. 3), which is presented either as short list with the fatty acids given in Δ–notation or as a long list with all information, available about the fatty acids, found in this search.

At this point of the database the user has the possibility to choose one of the fatty acids given in the list, in order to get information, which plants contain these fatty acid. By a click on the button '*plants containing molecule....*' in the field of the appropriate fatty acid the database presents a list with all plants which contain this fatty acid in the seed oil. Additionally, it is possible to limit the list to plants which contain this fatty acid with a certain amount.

This form is particularly useful to get the information which fatty acids are searchable in the database and additionally it is possible to search for plants containing individual fatty acids. The combination of several fatty acids in one search is not possible in this form.

5.2. Search by the Expert Form

The most powerful form of the database is the *Tables-Expert Form*. In this form it is possible to search for all kinds of information available in the database. The different fields of the database can be combined in all conceivable variations by the boolean operators AND, OR and AND NOT. Within

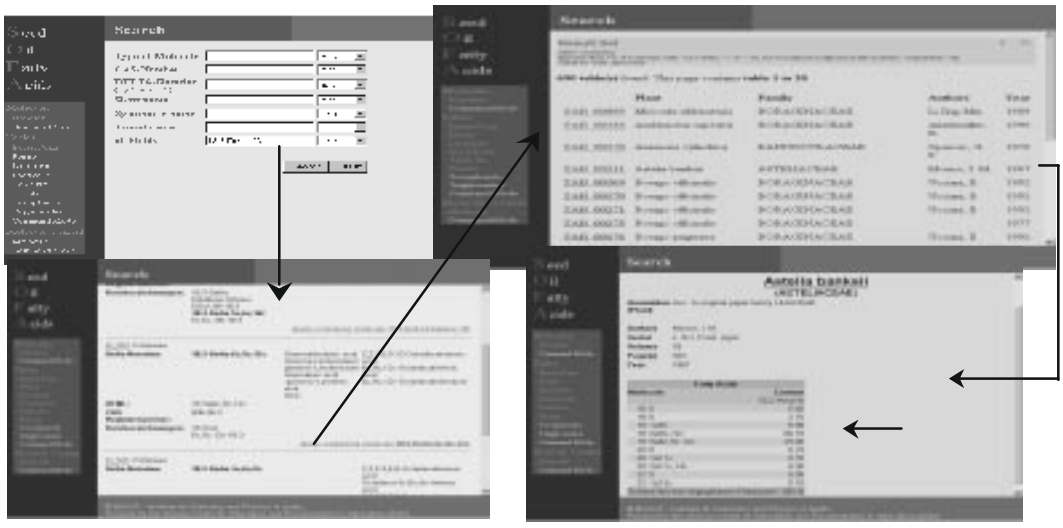

Figure 3
Example for a search for molecules.

most of the fields, e.g. *Author*, *Family* or *Plant* it is also possible to combine several terms by ',' for the boolean operator OR and '&' for the boolean operator AND. So the search can be restricted very precise to get the result as exact as possible. An example is the search for information about plants from the genus Brassica or Arabis, by the term *Brassica*,Arabis** in the field *Plants*.

In detail, up to five different fatty acids or structure elements of fatty acids can be searched together with different chemical characteristic features of the seed oil, botanical information of the plant and bibliographic data regarding the reference, where the information was found. Additionally it is possible to limit the content of the fatty acids upwards or downwards.

The result of such a search is a list of tables which agree with the search criteria. This list contain information about the table number, the name of the plant, the family, the name of the first author, who has published this information in a journal as well as the publication year. At the most 20 different tables can be listed at the same time. If more tables were found in one search it is possible to leaf through the search results by two buttons. From this list the user can choose the table which seems to be interesting for him, in order to get all information available in the database about this item (fig. 4). In the *Expert Form* it is possible to choose sort criteria for the presentation of the list, such as name of plant, family or author, to make the given list clearer and easier to read.

Some fatty acids are not searchable for percentage content, because they are not defined in the molecule list. One reason why these fatty acids were not picked up in the molecule list is, that the names given for these fatty acids in the literature often were neither systematic names nor trivial names. The authors used fantasy names, combination of names, if the content was given as sum for several fatty acids or even unspecific names. Examples are *10:2 Conj, 12:0 + 14:0* or *18:2N-X*. These fatty acids are also searchable by the field *Other Fatty Acids*, but a limitation of the content is not possible. In the tables resulting from a search the content of these fatty acids is also presented.

Those to whom this *Expert Form* is to extensive or who only want to search for a few search criteria, can use the forms *Plants*, *Literature* as well as *Seed & Oil*. These forms contain only the search criterion of the specific topic. Here a combination with other search criteria is not possible.

A further interesting point is the search for *Tocopherols*, *Sterols* and *Triglycerides*. The database allows searching for these compounds in three different forms. In these forms it is possible to combine a search for these compounds with a search for interesting fatty acids as well as a search for the plant name. Similar to the *Expert Form* the forms for tocopherols, sterols and triglycerides allow the user to specify the content which should be included in the plant seeds of interest. The result of such a search is again a sorted list of plants agreeing with the given search criteria. From this list it is also possible to display all the information available in the database for a chosen plant.

5.3. Search for content tables

The third possibility for the user to search in the database is given in the *Molecule Content - Molecule* form (Fig. 5). This form can be used to produce lists of fatty acids with the appropriate contents, which

Figure 4
Example for a search in the record Tabelle.

Figure 5
Example for a search for fatty acids with given contents.

fulfil the conditions of a given input. As further information the list shows the plant and family name, the name of the first author and the publication year. In this form, as in the Expert form up to five different fatty acids can be combined with search criteria for the plant and/or the family name. It is possible to get the resulting list sorted ascending or descending by the content of the appropriate fatty acids. By this way the user gets very fast an idea of the contents of unusual or interesting fatty acids in different plant seeds. From this list the user has the possibility to get more information about the appropriate plant by a click on the table number.

6. COMMAND MODE

If the forms available in the database are not sufficient to describe a problem it is possible to use the so-called FQM code. With this query language the user is enabled to search for any conceivable question. There is no limitation regarding the search criteria and the combination of the criteria. This language is not easy to use and therefore not intended for the normal users of computer programs, but this tool can be very helpful for experts who are willing to invest some work with the aim to get out the last from the database. An example for this FQM code is given in figure 6.

Figure 6
Example for a search by the FQM code: Search for fatty acids with an allyl hydroxy group in the middle of a high unsaturated molecule with 18 C atoms with more than 5 %.

7. CONCLUSION

In the moment the database contains more than 18.000 tables with results of the analysis from seeds of more than 7.000 various plant species. The tables contain at least the oil content, but in most cases the fatty acid composition of the seed oil is given. Information about the bibliographical reference of the data is presented for every table. Further on in several tables also some characteristic features, such as iodine value or density are available. In a separate molecule list more than 580 different fatty acids with the appropriate synonyms are linked with the database. For most of these fatty acids the user can find some data in the database. The other fatty acids were added because it may be possible that they could be interesting for plants not included in the database up to now, but probably in the future.

The new database should be very useful not only for renewable resources and "green" chemistry, but also for gene technology, for understanding the enzymes of fatty acid biosynthesis and their mutations during the evolution of plant families and species, in plant chemotaxonomy and in systematic and phylogenetic botany.

The search fields of the database can be combined in any way so that it is possible to recognize not only systematic connections between plants but also the occurrence of interesting fatty acids in plant kingdom. Thus, the database is interesting for food scientists, biochemists, chemists as well as for botanists.

ACKNOWLEDGEMENT

The work has been carried out with financial support from the Agency of Renewable Resources (FNR) (FKZ 97NR193). The authors thank Mrs. Miriam Schmitz-Peters for the technical assistance and Mrs. Birgit Borgmann, Mrs. Bärbel Michels as well as Mrs. Stefanie Enste for their unremitting efforts typing the data from the paper sheets into the computer. In the same way Prof. Dr. J. Heym (Fachhochschule Hof) and Mrs. Irene Santini de Sigel (ZADI) deserve thanks for programming the database.

REFERENCES

Abbot, T.P., Phillips B.S., Butterfield R.O., Isbell T.A., Kleiman R. (1997). On-line Chemical Database for New Crop Seeds. *J. Amer. Oil Chem. Soc.* **74**, 723-726.

Physical and Chemical Characteristics of Oils, Fats and Waxes Section I (1996). In: Official Methods and Recommended Practices of the American Oil Chemists' Society.

Aitzetmüller K., Friedrich H., Heym J. (1999). Einrichtung, Betrieb und Nutzung der web-basierten Faktendatenbank SOFA (Seed Oil Fatty Acids) in *Nutzung von Online-Datenbanken in regionalen Informationsnetzen im Agrarbereich,* ed. 2. Hallescher GIL-Workshop, Universität Halle.

Recibido: Abril 2002
Aceptado: Diciembre 2002