Pure & Appl. Chem., Vol. 65, No. 2, pp. 235–244, 1993.
Printed in Great Britain.

Chirbase: A molecular database for storage and retrieval of chromatographic chiral separations

Christian ROUSSEL and Patrick PIRAS

ENSSPICAM, CNRS URA 1410, University Aix-Marseille III
13397 Marseille Cedex 13, France.

Abstract

In order to meet the strong demand for storage and retrieval of chiral separations, we have developed Chirbase a database built on Chembase® from Molecular Design Limited, a very powerful and well spread software. Chirbase allows the selection of the most promising conditions for a given chiral separation by searching and retrieving at the same time molecular fragments issued from the compound and from the stationary phase.

The obtainment of optically pure compounds and the control of their optical purity cover the scope of various very active fields of research.
Enantiomerically pure compounds can be obtained through various routes which involve chiral discrimination (ref. 1):

- -isolation from natural products (chiral pool)
- -asymmetric synthesis
- -enzymatic resolution of racemates or enzymatic reaction on prochiral compounds.
- -crystallization or chromatography of diastereomeric compounds obtained by reaction of a racemate with an optically active compound.
- -preparative chromatography of racemates on chiral stationary phases (CSP) recently reviewed by Francotte et al (ref. 2).

The control of the optical purity of the compounds obtained through the above mentioned techniques can be performed using different methods: NMR experiments, polarimetry or analytical chromatography on a suitable CSP, the later method being far more accurate than the two other (ref. 3).

Chromatography of racemic or scalemic compounds on a CSP (chiral chromatography) is a field of research in expansion since the last 15 years in order to meet the strong demand in chiral issues (ref. 4a, 4b, 4c).

The basic principle of the chromatographic chiral separation rests on the energy difference in the two labile diastereomeric complexes formed between each enantiomer and the chiral framework of the stationary phase. This difference in energy is related to α which is experimentally obtained by the ratio of the capacity factors of the two enantiomers: k'2/k'1. The larger is α, the better is the chiral discrimination between the two enantiomers by the chiral support or in other term the better is the molecular recognition of one enantiomer by the chiral selector. The enantiomer having the smallest affinity with the support will be eluted first.

CSPs are nowadays extremely diverse, for liquid chromatographic applications, we have indexed ca 400 CSP. They are obtained according to different methodology:

- an optically pure compound of known configuration presenting one or several stereogenic centers is bound to an achiral support such as silica. These CSPs are called molecular phases (ref. 5).

- an optically active polymer is coated on or bound to an inactive support which provides mechanical properties suitable for chromatography. In some case the polymer can be mechanically stable enough to be used directly without any additives. Optically active polymers can be obtained by synthesis through polymerisation of an achiral monomer in the presence of a chiral catalyst (ref. 6) or by polymerisation of an homochiral monomer (ref. 7). Polymers can be optically pure by nature, this is the case of the polysaccharides (ref. 8) which can be further derivatized to give extremely useful CSPs. These CSPs are cooperative phases since the chiral selector responsible for the separation cannot be precisely identified.

- a naturally occurring macromolecule such as protein is bound to an inert support. These stationary phases are called protein phases (ref. 9).

The great number of available CSPs as well as the various experimental conditions make the choice of the optimal phase for a given separation a difficult task. These phases are generally expensive and some of them have a poor life time.

CHIRBASE DESCRIPTION

In 1988, we started the project Chirbase on Chembase®, Molecular Design Limited's, graphic-based software, which allows the storage and retrieval of molecular structures and reactions (ref. 10).
Our choice was directed by the unique capability of this software to offer sub-structure querying facilities (vide infra) in a molecule or simultaneously in two molecules linked in a reaction formalism.
Furthermore Chembase® was largely implemented in pharmaceutical industries and in general in all chemical laboratories dealing with relation between structure and activity for the development of in house molecular or reaction databases. Chembase® allows the storage of alphanumerical data which can be retrieved and sorted.

Chirbase is organized into molecular and reaction databases according to the Chembase formalism:

CHIR-SP.DB is a molecular data base which displays all the structures of the CSPs used in Chirbase, with some additional information (Fig. 1): supplier and trade name when they are commercially available, name and synthesis reference. CHIR-SP.DB is used as a source of templates for CSPs, and allows the search of similar constituent in various CSPs. One of the major application of CHIR-SP.DB is to select within the commercially available CSPs those which have some structural features in common with a non commercially available CSP.

CHIR-MOL.DB is a molecular database which displays all the structures of the enantiomers which have been separated (Fig.2). When a given molecule has been separated several times under different experimental conditions, it will appear as many times as it has been separated. CHIR-MOL.DB contains ca 17500 entries dealing with ca 6000 independent molecules.

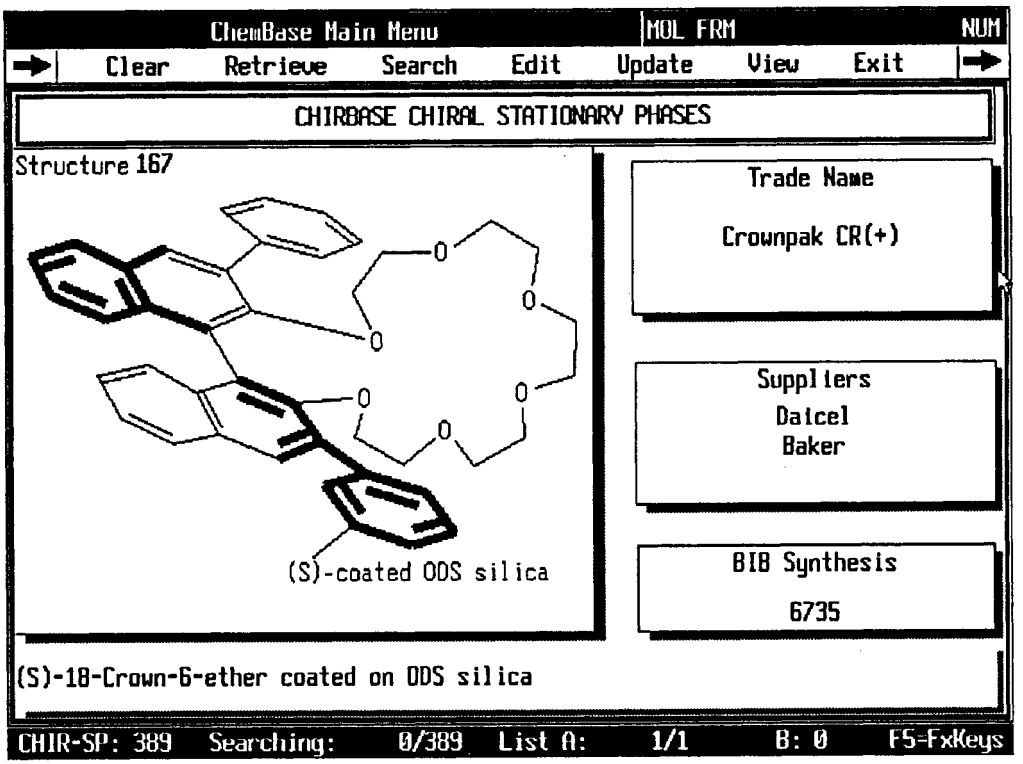


Fig. 1 Data displayed in CHIR-SP.DB

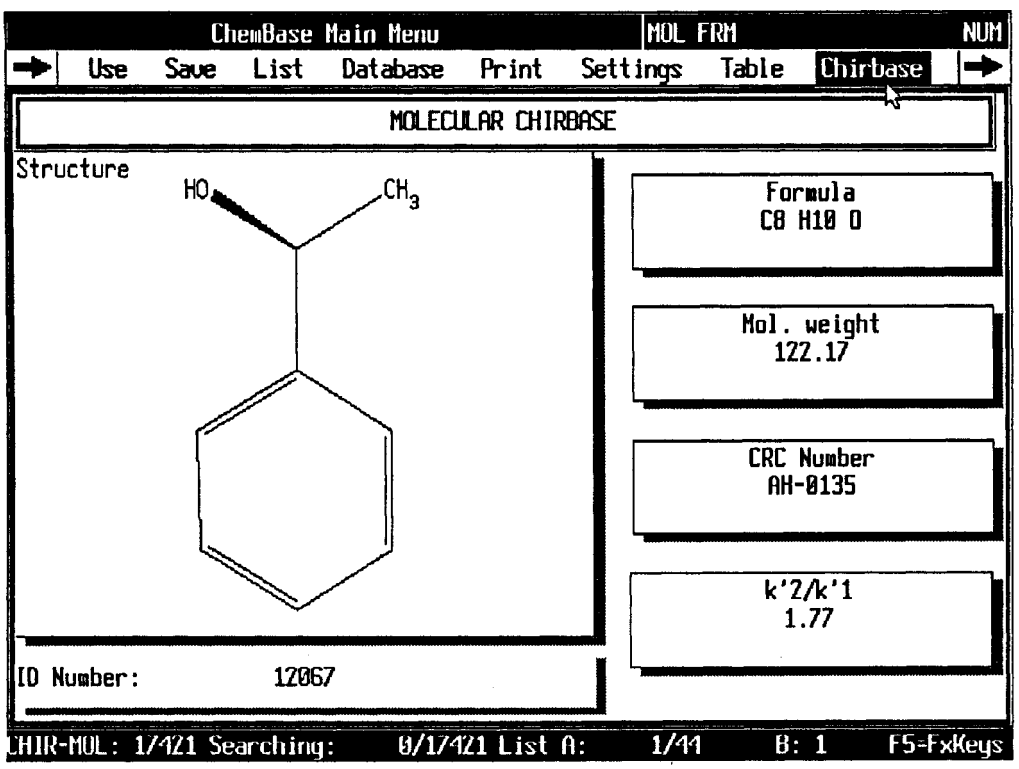


Fig.2 Data displayed in CHIR-MOL.DB

Four databases: CHIR-1.DB, CHIR-2.DB, CHIR-3.DB and CHIR-4.DB are built according to the reaction formalism of Chembase.

The three first DBs contain 5000 entries each, this limitation is imposed by Chembase which limits the size of a database to 8 Mo. CHIR-4.DB is under development.

The database structure contains 30 fields: 1 structure field and 29 alphanumerical fields:

6 fields for references: **authors**, **journal**, **year**, **page**, **volume** and **bib** (a personal reference number for bibliography).

3 fields for compounds: **name**: commercial or chemical name, **code**: identification number specific to each molecule and **chirality**: type of chirality (center, axial, planar ...).

1 field for entry number: **ID**, this entry number is unique and is linked to the entry number of CHIR-MOL.DB. This is the way to link the molecular data base CHIR-MOL.DB with the "reaction" databases CHIR-X.DB.

6 fields for separation data: **first eluted**: sign and/or absolute configuration of the first eluted enantiomer, **second eluted**, **k'1**: capacity factor of the first eluted enantiomer, **k'2**, α : separation factor, **res**: resolution. The data reported in these fields generate a lot of information depending on the user.

8 fields for experimental conditions: **Method**: HPLC, LC, SFC, Affinity, **mobile phase**: composition of the mobile phase, **type of chromatography**: analytical, preparative, **amount**: injected in case of preparative or analytical separation; **flow-rate**, **temperature**, **detection**, **type of column**: length, diameter, number of columns.

4 fields for the **CSP**: : chemical name, **trade name**, **supplier**, **particle size**.

1 fields for **comments**: this field is devoted to some information given by the developer. For example, when separation data are missing in a report, one will find in this field key-words like "baseline separation seen on chromatogram", "partial separation" or "estimated alpha value" when we are able to estimate data from retention times reported on a chromatogram.

Chromatographic data and structures appear on a first screen (fig. 3), analytical conditions and reference on a second screen (fig. 4). An other form displaying reference, structures and chromatographic data on the same screen (not shown here) has also been created in order to compare immediately the chromatographic separation with the different sources of information.

The structure field displays the structure of the two enantiomers and the chiral selector of the CSP. Since the Chembase's reaction formalism is intended for taking care of reactions: the first eluted enantiomer is considered as a "reactant", the CSP as an "intermediate" and the second eluted enantiomer as a "product". The stereochemistry of all the displayed structures are reported using the classical conventions. The stereochemistry around stereogenic center can be recognized by Chembase unfortunately axial and planar chirality are not accounted for.

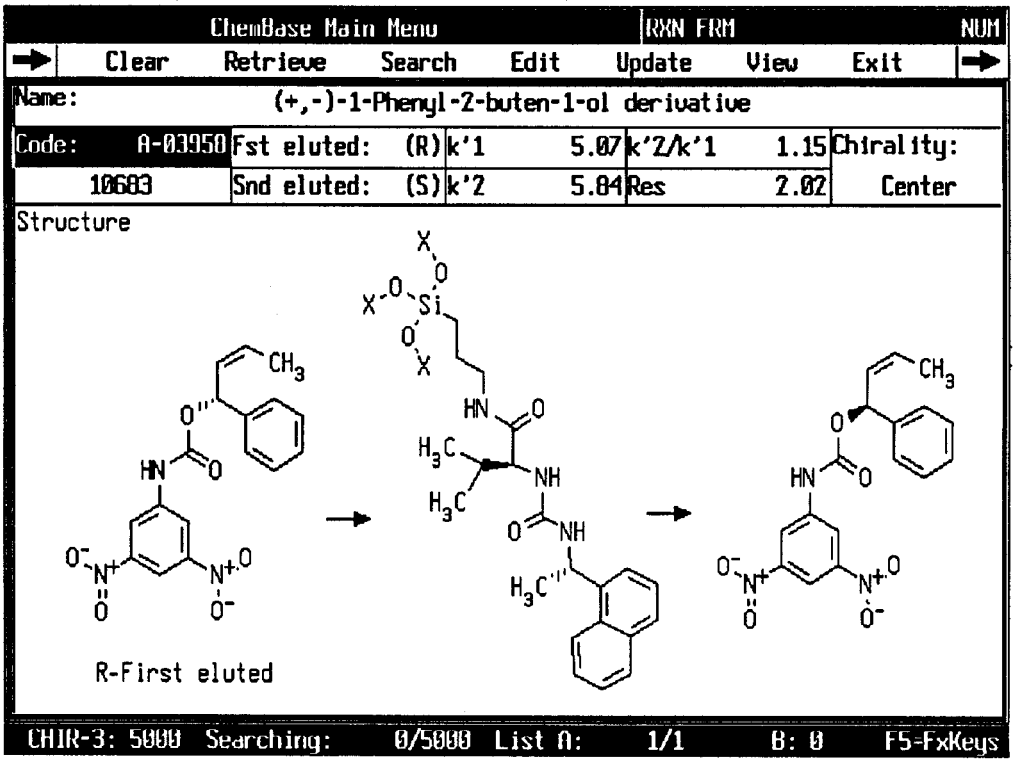


Fig. 3 First screen displaying CHIR-X.DB data

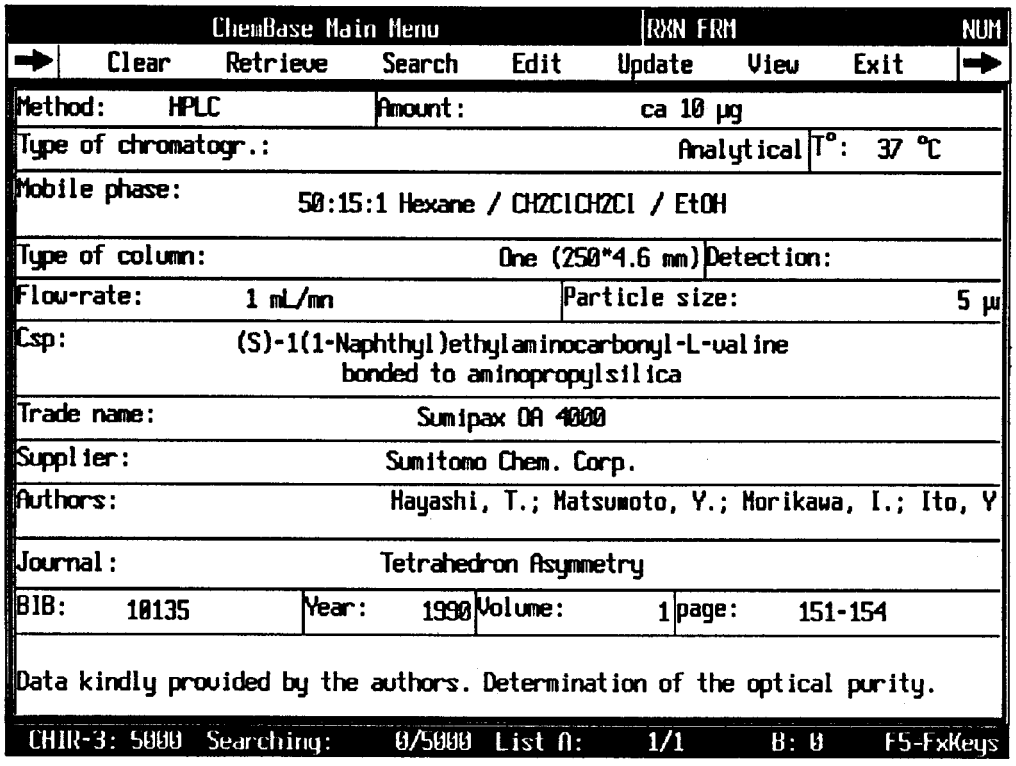


Fig. 4 Second screen displaying CHIR-X.DB data

CHIRBASE USES

It is worth noting that Chirbase is a databank. As it is well documented in information science, the information generated from a databank will go through the specific background of the user and it is clear that the procedure for using the database will be closely related to the user application.

We report in Fig. 5 some typical procedures for selection of the appropriate conditions.

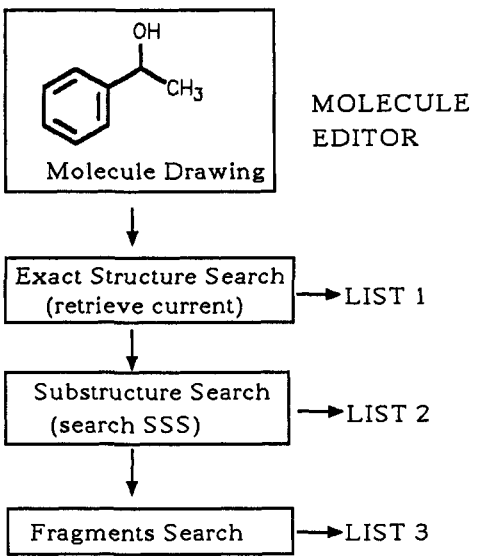


Fig. 5
Description of an usual searching procedure

From the molecule drawing, one can search in CHIR-MOL.DB an exact matching, this exact matching using Chembase Retrieve Current instruction is very fast and list 1 is generated which contains all the separations reported for the given compound. This molecular list may contain separations which are reported in all four CHIR-X-DB. Since CHIR-MOL.DB and CHIR-X.DB are linked through the ID field, List 1 can be manually injected in each CHIR-X.DB for examination. We have developed a macrocommand associated to a button in the molecular editor. This command can be used in different ways: list 1 is automatically generated and injected in all four CHIR-X.DB generating four lists containing all the data, these lists are saved and can be used later or these lists can be automatically merged into a new databank with the same structure as in CHIR-X.DB. All these procedures are very fast and after few seconds on a PC (25 MHz), a new data bank has been created which contains all the data for the initial compound. This new DB has received a name and has been saved.

From the molecule drawing, one can use in CHIR-MOL.DB a substructure matching, this substructure search using Chembase SSS instruction generates list 2. List 2 can be treated as for list 1 by injection in the four CHIR-X.DB. List 2 contains list 1.

Substructure search is particularly well suited for chromatographic chiral separation since in several cases free alcohol, free amine, or carboxylic acid must be derivatized with a suitable achiral reactant. One can choose the best reported derivatization in order to achieve a good separation.

Example: The enantioselective reduction of prochiral methylphenyl ketone to give scalemic 1-phenyl-ethan-1-ol is a model reaction which has been used in several reports to evaluate the chiral discrimination ability of chiral reduction catalysts. The determination of the ee shall be accurate. The exact structure search in CHIR-MOL.DB for the alcohol generates list 1 which contains 44 hints (covering 22 reports).

In order to compare and sort candidate CSPs and analytical conditions, a database corresponding to list 1 is automatically created and the data obtained in list 1 are displayed by Chembase in a table making trends easy to disclose. In table 1, we have selected and sorted different CSPs from the list found in Chirbase for this compound.

TABLE 1. Choosing a CSP for the separation of 1-phenyl-ethan-1-ol from selected data organized in a table according to α decreasing

CSP	1st Eluted	α	Ref.
Cellulose Tribenzoate (beads)	S	1.77	(11)
Cellulose tris-3,5-Dimethylphenylcarbamate/silica	+R	1.33	(12)
Cellulose tris-(4-methylbenzoate)/silica	(-)	1.17	(13)
Cellulose tris(3-methyl-benzoate) (beads)	(S)	1.15	(14)
Cellulose tris(4-methyl-benzoate) (beads)	(S)	1.13	(14)
Quinine monomer covalently bonded to silica.	(?)	1.06	(15)
(R)-N-(3,5-Dinitrobenzoyl)-1-naphthylglycine	(-)	1.06	(16)
(R)-N-(3,5-Dinitrobenzoyl)phenylglycine	(S)	1.05	(17)
Quinine bonded to mercaptopropyl-silica	(S)	1.00	(18)
Quinidine bonded to mercaptopropyl-silica	(R)	1.00	(18)
beta-Cyclodextrin carbamate bonded to silica	(R)	1.00	(19)
Cellulose tricinnamate/silica	(?)	1.00	(20)
(+)-Hexahelicen-7-ylacetic acid	(?)	1.00	(21)
(+)-Ground-Poly(triphenylmethyl-methacrylate)	(?)	1.00	(22)
Microcrystalline cellulose triacetate	(-)	1.00	(23)
(+)-Poly(triphenylmethyl-methacrylate)	+R	?	(24)

Inspection of this table indicates that baseline resolution can be achieved using cellulose tris(benzoate) coated on silica (Chiralcel OB) or in beads, as well as cellulose tris(3,5-dimethylphenyl) carbamate coated on silica (Chiralcel OD). Poor separations are reported on various other CSPs. It is particularly interesting to note that on Chiralcel OB, the S(-) enantiomer is eluted first whereas the R(+) enantiomer is eluted first on Chiralcel OD .

Since the accuracy of the chromatographic determination of a minor component is always better when it appears first, the user will select the stationary phase according to the expected major enantiomer obtained in the model reaction. The user interested in the preparative obtainment of both enantiomers of labelled 1-phenyl-ethan-1-ol will use cellulose tribenzoate in beads. Since, examination of data in table 1 allows us to eliminate the Pirkle phase type (phenylglycine or naphthylglycine bonded phase) for a preparative purpose. Only recycle technique could be envisaged.

The user can also look at other successful CSPs readily accessible in-house to him or choose an optimal mobile phase for a given CSP among several eluent tested (with also cost and safety considerations).

In case of no satisfactory entries using the exact structure search, one has to apply the substructure search to generate list 2. As an example we have chosen 1-(1-naphthyl)-amino-ethane for which 5 poor separations are obtained using the exact match. Substructure search give list 2: 248 entries. List 2 can be set as the main domain of research for a more precise substructure search by solely allowing derivatization on the NH2. It results in a new list with 119 entries. That list can be used in different ways depending on the requirement of the user. One will select for preparative purposes alpha values larger than 1.5 (36 entries over 119) associated with a derivatization easily removed. One can select on the availability of the CSP in the lab, or on the order of elution or the availability of the derivatizing agent in the lab etc..

The previous example was a good example of the Chirbase possibilities, however in many actual cases list 1 and list 2 are empty since the separation of the exact molecule or any derivatives have not been already reported. In that case the structure must be split into fragments as shown for the azole fungicide derivative in fig. 6. for which the optical purity was determined by HPLC but the phase was not disclosed in the report (ref. 25).

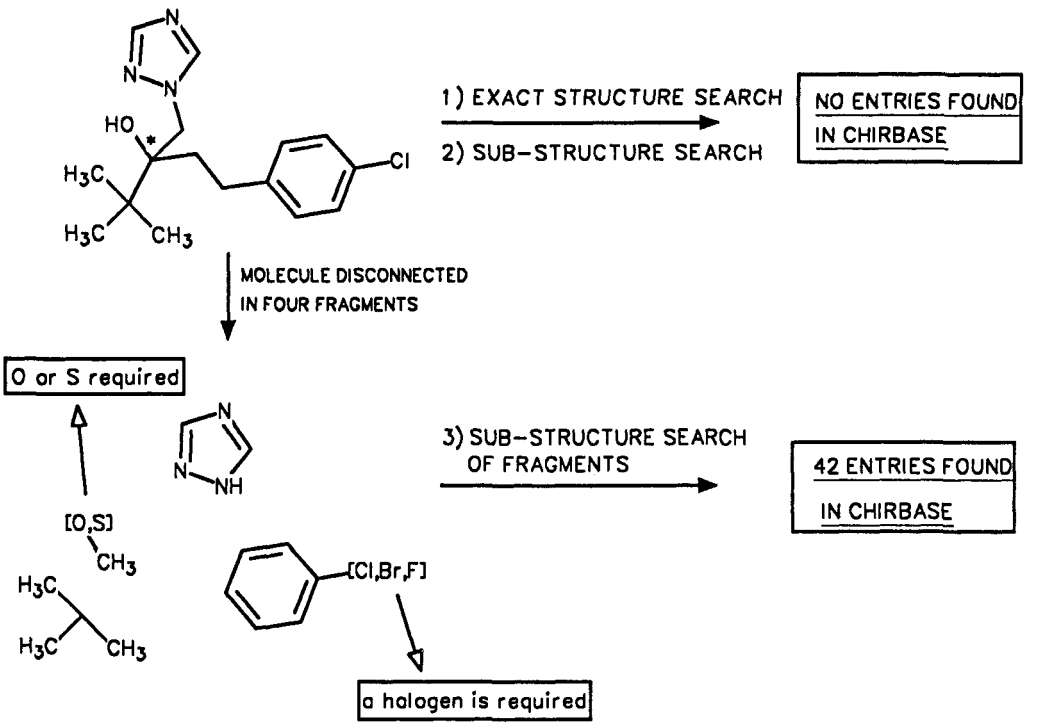
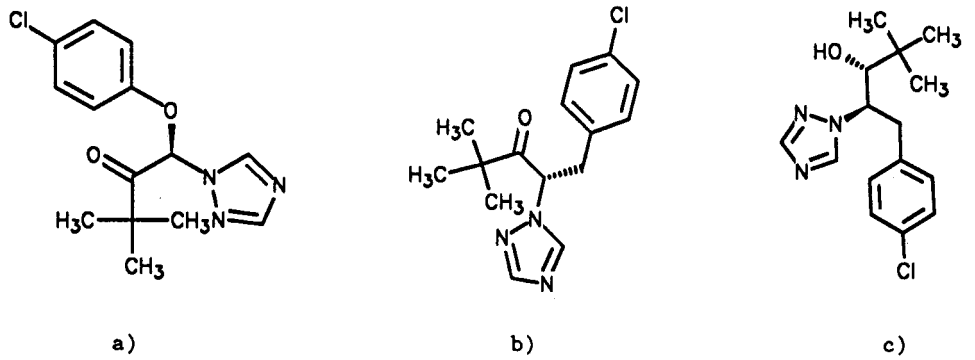


Fig. 6 Fragments searching procedure of an azole derivative

In this way, 42 entries sharing these features were retrieved in Chirbase. Because only some of the structural features of the molecule are involved in chiral recognition, one may expect that the same type of CSPs should be successful for the separation of closely structurally-related compounds. In fig. 7 are shown three examples of compounds found in Chirbase after a fragments search of the azole derivative.



a) b) c)

Fig. 7 Examples of results obtained after a sub-structure search of a combination of structural fragments of an azole derivative.
CSPs reported in Chirbase:
-a): Cellulose tris(4-methylbenzoate) (ref. 13)
-b): Cellulose tris-phenylcarbamate (ref. 26)
-c): Cellulose tris-phenylcarbamate (ref. 27)

The sub-structure search of fragments often permits to isolate a precise column or a family of CSPs. Thus, in the previous example, a cellulose column bearing phenyl groups will be preferred. But it permits also:

-to eliminate the CSPs which gave unsatisfactory results (alpha value equal to 1.00 in Chirbase) and avoid useless experiments.

-to detect the CSPs which are able to recognize the key structural features of a compound and are compatible with them (for example some phases can't afford underivatized NH2 or COOH groups). Useful information often emerge by listing the various substituents and their position in the framework (near or far from an asymmetric centre). Rules can be established by comparing their relative effect on actual separation.

-to choose the optimal conditions for a given CSPs by viewing different mobile phases (modifiers, pH, normal or reverse) or temperature.

CONCLUSION

Chirbase is today an effective tool which is implemented in many laboratories. It has already provided direct practical help in meeting the needs of industry or university.
We are now proceeding to the transfer of Chirbase in ISIS (Integrated Scientific Information System), a new available software marketed by Molecular Design Ltd which offer substantial progress in data management (query within multiple remote host databases in a given network) and prediction tools through links with modeling software.
Following our research on chiral discrimination (ref. 28), the ISIS version of Chirbase will provide a number of promising developments and new molecular insights in the selection and the design of CSPs.

REFERENCES

1. G.M. Ramos Tombo and D. Bellus, Angew. Chem. Int. Ed. Engl. 30, 1193-1386 (1991).
2. E. Francotte and A. Junker-Buchheit, J. Chromatogr. Biomed. Applic. 576, 1-45 (1992).
3. A. Mannschreck, Chirality 4, 163-169 (1992).
4a. S. Allenmark, Chromatographic Enantioseparation. Methods and Application, Ellis Horwood, Chichester (1991).
4b. A.M. Krstulovic, Chiral Separations by HPLC, Ellis Horwood, Chichester (1989).
4c. D.W. Armstrong and S.M. Han, C.R.C. Crit. Rev. Anal. Chem. 19, 175-224 (1988).
5. S.R. Perrin and W.H. Pirkle, Chiral Separations by Liquid Chromatography, p. 43-66, ACS Symposium Series 471, S. Ahuja, Washington (1991).
6. Y. Okamoto and K. Hatada, J. Liq. Chromatogr. 9, 369-384 (1986).
7. G. Blaschke, J. Liq. Chromatogr. 9, 341-368 (1986).
8. T. Shibata, I. Okamoto and K. Ishii, J. Liq. Chromatogr. 9, 313-340 (1986).
9. S. Allenmark, J. Liq. Chromatogr. 9, 425-442 (1986).
10. B. Curry-Koenig and C. Seiter, Am. Lab. May (1986).
11. E. Francotte and R.M. Wolf, Chirality 3, 43-55 (1991).
12. Y. Okamoto, M. Kawashima and K. Hatada, J. Chromatogr. 363, 173-186 (1986).
13. T. Shibata, A. Ichida, Y. Fukui and K. Mori, International Symposium on Chiral Separations, Guilford (1989).
14. E. Francotte and R.M. Wolf, J. Chromatogr. 595, 63-75 (1992).
15. H.W. Stuurman, J. Köhler and G. Schomburg, Chromatographia 25, 265-271 (1988).
16. N. Oi, H. Kitahara, Y. Matsumoto, H. Nakajima and Y. Horikawa, J. Chromatogr. 462, 382-386 (1989).
17. W.H. Pirkle and J.M. Finn, J. Org. Chem. 46, 2935-2938 (1981).
18. C. Rosini, C. Bertucci, D. Pini, P. Altemura and P. Salvadori, Chromatographia 24, 671-676 (1987).
19. K. Fujimura, M. Kitagawa, M. Takayanagi and T. Ando, J. Liq. Chromatogr. 9, 607-620 (1986).
20. I.W. Wainer, R.M. Stiffin and T. Shibata, J. Chromatogr. 411, 139-151 (1987).
21. S.A. Matlin, V.E. Stacey and W.H. Lough, J. Chromatogr. 450, 157-162 (1988).
22. Y. Okamoto, I. Okamoto and H. Yuki, Chem. Lett. 835-838 (1981).
23. A. Mannschreck, M. Mintas, G. Becher and G. Stühler, Angew. Chem. Int. Ed. Engl. 19, 469-470 (1980).
24. H. Yuki, Y. Okamoto and I. Okamoto, J. Am. Chem. Soc. 102, 6356-6358 (1980).
25. J. Kaulen, Angew. Chem. Int. Ed. Engl. 28, 462-463 (1989).
26. S. Black, L. Williams, R. Davey, F. Moffatt, R. Jones, D. Mc Ewan and D. Sadler, Tetrahedron 45, 2677-2682 (1989).
27. Z. Wu, D.M. Goodall, D.K. Lloyd, P.R. Massey and K.C. Sandy, J. Chromatogr. 513, 209-218 (1990).
28. C. Roussel, J.-L. Stein, M. Sergent and R. Phan Tan Luu, Recent Advances in Chiral Separations, D. Stevenson and I.D. Wilson, Plenum Press, New-York (1991).