Journal of Computer-Aided Molecular Design, 7 (1993) 61–69
ESCOM

J-CAMD 188

Chiral chromatography and multivariate quantitative structure–property relationships of benzimidazole sulphoxides

Patrick Camilleri, David J. Livingstone, José A. Murphy and David T. Manallack*

SmithKline Beecham Pharmaceuticals, The Frythe, Welwyn, Herts AL6 9AR, U.K.

Received 9 September 1992
Accepted 17 October 1992

Key words: Principal components analysis; Non-linear mapping; Computational chemistry; Unsupervised learning; Pattern recognition

SUMMARY

Various benzimidazole sulphoxides were chirally resolved employing an amylase-based chiral stationary phase. The structure–property relationships of these compounds were investigated using calculated physicochemical properties, molecular modelling and multivariate statistical techniques. A data set of 254 molecular descriptors was used to represent the series of compounds. Analysis of the data set using principal components analysis and non-linear mapping suggested that the separation factor of each enantiomeric pair was associated with nine molecular properties and, in particular, molar refractivity of the Z substituent and the partial charge of atom 6. The separation factor of a sulphoxide not used in the analysis was well predicted thus suggesting that these models may be used to generalize.

INTRODUCTION

Benzimidazole sulphoxides, such as omeprazole, have drawn considerable attention in recent years as therapeutic agents for the treatment of acid-related gastrointestinal disorders [1,2]. The mechanism of action of these compounds has been ascribed to the chemical conversion into the corresponding cyclic sulphenamides which react irreversibly with a cysteine residue on (H^+ + K^+)-ATPase [2].

The benzimidazole sulphoxides possess a chiral sulphur atom. While studies [3] have investigated the HPLC behaviour of racemic sulphoxides, chiral resolution of the compounds examined in this study has not been reported previously. This report gives an account of the chiral resolution of a number of benzimidazole sulphoxides of the general structure shown in Fig. 1 and discusses variations in the chromatographic behaviour of pairs of enantiomers and the separation factor, α, in terms of their chemical structure as described by various physicochemical parameters. The

*To whom correspondence should be addressed.

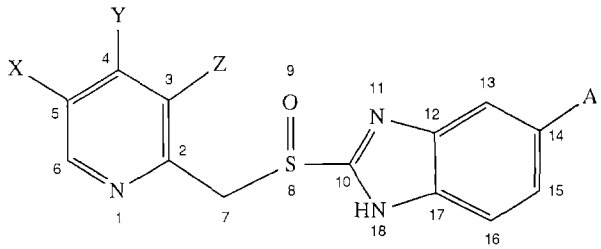


Fig. 1. The numbering system used for atoms and substituents.

Hansch approach to the correlation of biological activity with chemical structure consists of an extension of the Hammett equation into biology/biochemistry [4]. This philosophy can also be applied to any property which relies on chemical structure; in the case of retention on a chromatographic column this might be expected to work well since the interactions involved may be simpler than those which occur between a biologically active molecule and its site of action. Recent reports of the quantitative analysis of biological properties have involved descriptors calculated by theoretical methods and the use of statistical techniques other than regression analysis [5,6]. We have therefore examined these chromatographic data by means of the 'classical' Hansch approach and by using properties derived by the techniques of computational chemistry. This may be regarded as the establishment of quantitative structure–property relationships (QSPR).

EXPERIMENTAL

Materials

n-Hexane (Rathburn) and propan-2-ol (BDH Ltd.) were degassed with helium before use. Benzimidazole sulphoxides **1** to **12** were synthesized at SmithKline Beecham Pharmaceuticals, Welwyn and Byk Gulden Lomberg Chemische Fabrik GmbH, Germany [2,7].

High-performance liquid chromatography (HPLC)

Compounds **1** to **12** were analysed using a Perkin-Elmer Series 4 liquid chromatograph. A CHIRALPAK AD column (250 mm × 4.6 mm ID) was used as the stationary phase and compounds were eluted with a mixture of *n*-hexane and propan-2-ol at a flow rate of 1 ml min^{-1}. Sulphoxides were detected with a Kratos Spectroflow 783 variable wavelength detector set at 300 nm. The column was operated at ambient temperature.

Molecular modelling

Molecules were constructed and minimized using the COSMIC in-house molecular modelling package [8]. For this procedure, the isomer and conformation used were kept constant for the 12 compounds in this data set. Electronic properties were calculated using the semi-empirical molecular orbital program MOPAC (MNDO Hamiltonian) [9].

Molecular descriptors

The compounds in this data set were classified as either good ($\alpha > 1.40$), moderate ($1.20 < \alpha < 1.40$) or poor separators ($\alpha < 1.20$). A total of 254 molecular properties comprising

whole molecule-, substituent- and atom-based properties were calculated and collated using the program GENPROP [10]. Atom-specific properties were calculated for selected core atoms (see Fig. 1) and included partial atomic charges, electrophilic and nucleophilic frontier electron densities, self-atom polarizabilities, and nucleophilic and electrophilic superdelocalizabilities. Whole molecule properties included the partition coefficient, molar refractivity, moments of inertia, principal ellipsoid axes and axis ratios, dipole moments and vectors, and the energy of the HOMO and LUMO. The properties calculated for each substituent (see Fig. 1) included the partition coefficient, molar refractivity and the maximum and minimum dimensions of the substituent in the x, y and z directions. In addition to these calculated properties, GENPROP also provides access to a look-up table of literature substituent constants [11].

Multivariate analysis

Data handling and analysis were done with RS1 (BBN Software products U.K. Ltd., Staines, Middlesex) and the pattern recognition package, ARTHUR (Infometrix, Inc., Seattle, WA). Calculations were performed on a VAX cluster.

RESULTS AND DISCUSSION

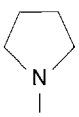
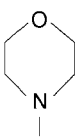
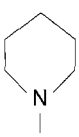
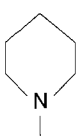
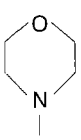
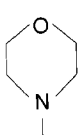
A number of enantiomeric sulphoxides have been resolved using a variety of cellulose-based chiral stationary phases (CSPs) found in CHIRALCEL OF, OB and OD columns with hexane/propan-2-ol as the mobile phase [12]. Since these three cellulose-based CSPs were not available in our laboratory, we examined the use of other cellulose-based columns such as OC and OJ. These did not give satisfactory and consistent resolution of the benzimidazole sulphoxides under study. In contrast, the amylase-based CHIRALPAK AD stationary phase gave excellent separation of the majority of enantiomeric pairs studied.

Table 1 summarizes the chromatography data for compounds **1** to **12**. We have also included in this table the calculated log P obtained using the MEDCHEM software package [13]. A plot of the log of the capacity ratio k′ for the faster eluting enantiomer against Clog P (data not shown) demonstrated that retention of the compounds increases with hydrophilicity as is expected for the behaviour of molecules on a normal phase stationary support. However, longer retention on the CSP does not guarantee resolution of the enantiomers. Figure 2 illustrates the separation of the enantiomers of **8** compared to the lack of resolution of the antipodes of **9** under the same chromatographic conditions, when the antipodes have similar values of log k′.

Resolution of the enantiomers of the benzimidazole sulphoxides appears to depend strongly on the presence of a substituent in position 'Z' and to a lesser extent on the electronic nature of substituent 'A'. In all molecules having a substituent 'Z' the separation factor, α, is greater than 1.20 with the exception of compound **2** where 'Z' is the relatively small fluorine substituent; the molar refractivity (a measure of steric bulk) of a fluorine atom is close to that of hydrogen [14]. Moreover, the electronic nature of 'Z' does not seem to be of importance in determining the ease of resolution of enantiomers. In fact, 'Z' can be either electron withdrawing (e.g. 'Z' = Cl) or electron donating (e.g. 'Z' = CH_3 or OCH_3). Thus, from the data in Table 1 there is evidence to suggest that the Z substituent may influence the orientation of the pyridine ring with respect to the benzimidazole moiety to become a conformation that allows a transient diastereoisomeric complex to occur between the CSP and one of the pair of enantiomers.

The electronic nature of substituent 'A' may influence resolution of compounds **1** to **12**. Indeed, the acidity of the NH group on the benzimidazole moiety may be a key interaction with the CSP. Such a relationship is difficult to assess due to the limited range of substituents in the A position. Enantiomeric resolution is achieved with either an electron donating -OCH_3 group or an electron withdrawing group such as -OCF_2O-. Interestingly, the introduction of a stronger electron withdrawing substituent, -OCF_2CF_2H (compound **9**), did not result in the resolution of each

TABLE 1
HYDROPHOBICITY AND RETENTION CHARACTERISTICS OF BENZIMIDAZOLE SULPHOXIDES

Compound	X	Y	Z	A	Clog P	log k'[a]	α[b]
1	Br		H	OCH_3	3.2	0.33	1.19
2	H		F	OCH_3	1.4	0.70	1.11
3	H		Cl	OCH_3	3.6	−0.06	1.66
4	CH_3		Cl	OCH_3	4.3	−0.06	1.66
5	H	OCH_3	H	OCH_3	1.1	0.37	1.16
6	H	OCH_3	CH_3	OCH_3	1.8	0.34	1.26
7	Br		H	OCH_3	2.1	0.40	1.19
8	H	OCH_3	CH_3	O-CF_2-O	3.3	0.25	1.25
9	H	OCH_3	H	OCF_2CF_2H	2.1	0.27	1.00
10	CH_3	OCH_3	CH_3	OCH_3	2.4	0.06	1.30
11	CH_3	$N(CH_3)_2$	Cl	OCH_3	2.9	−0.04	1.59
12	H		Cl	OCH_3	2.0	0.67	1.22

[a] Capacity ratio for the faster eluting enantiomer.
[b] Separation factor.

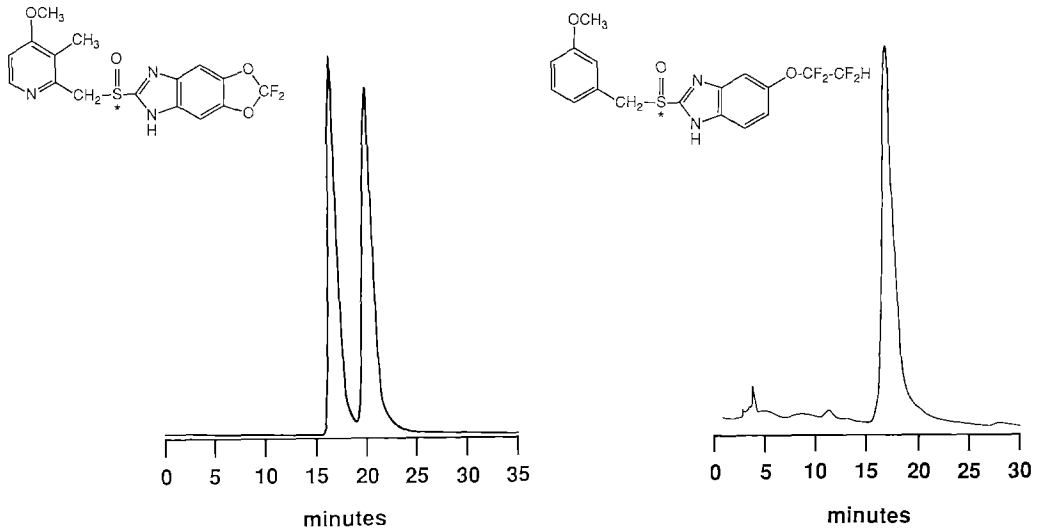


Fig. 2. A comparison of the chromatographic behaviour of compounds **8** and **9** under the same elution conditions.

enantiomer. The possibility remains, however, that this lack of resolution may be due to the increase in conformational space occupied by the $-OCF_2CF_2H$ group. Clearly, the most significant factor affecting the resolution of benzimidazole sulphoxide enantiomers concerns the nature of substituents X, Y and Z.

Although it appears that there is some explanation of the chromatographic behaviour of these compounds in terms of the 'classical' physical organic descriptors, it is clear that more detailed parameters are required if a quantitative description of the chiral discrimination process is to be obtained. With the procedure described in the experimental section, a data set was created for these 12 compounds containing 254 parameters, which comprised 20 whole molecule-, 108 atom- and 126 substituent-based properties. Experience in the application of this approach to sets of biologically active compounds led us to expect that there would be a considerable degree of redundancy in the data. One way to remove such redundancy is to identify pairs of correlated parameters using a procedure such as the program CORCHOP [16]. Use of this method with a correlation coefficient limit of 0.75 resulted in the removal of 221 descriptors. It should be stressed at this point that the remaining 33 parameters each represent a family of descriptors which contain similar information [6,17].

One way to examine the 33 properties obtained above is to employ multivariate statistical display techniques. Briefly these methods provide a means of reducing large N-dimensional data sets to two- or three-dimensional plots which can easily be displayed and evaluated. Examples of such techniques include principal components analysis (PCA) and non-linear mapping (NLM) [18]. Displays produced using both techniques on the remaining 33 properties demonstrated clustering of compounds possessing similar separation characteristics (data not shown). These clusters were not sharp, however, indicating that the properties used to calculate these plots contain information that is useful for the classification of separation categories but that there is extraneous information not useful for this purpose. A means by which the data set can be

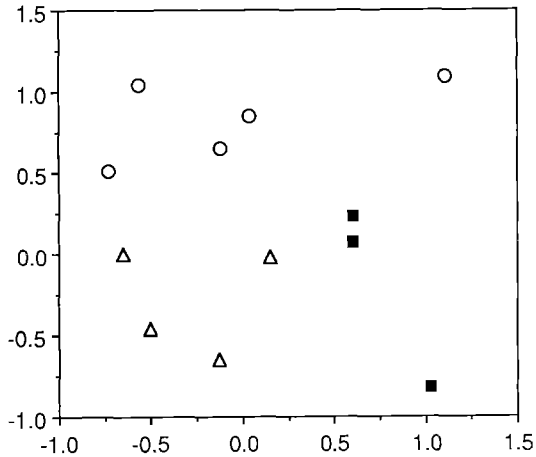


Fig. 3. Non-linear map of compounds **1–12** derived from nine parameters. ■: good separator; △: moderate separator; ○: poor separator.

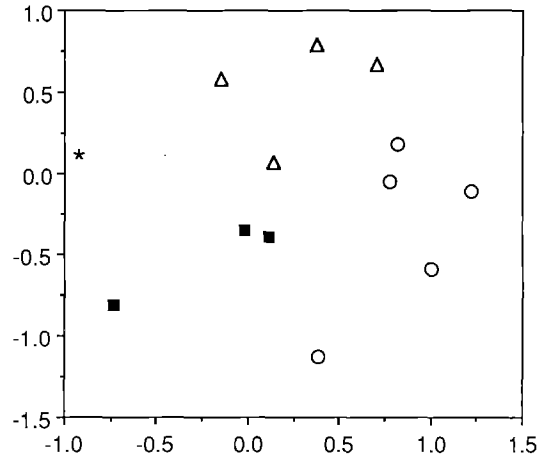


Fig. 4. Non-linear map of compounds **1–12** derived from nine parameters. Included in this calculation was a single test compound which is indicated with a *. ■: good separator; △: moderate separator; ○: poor separator.

improved is to select those properties which are most useful for the classification of separation categories. To achieve this, the ARTHUR routine SELECT [19], which is related to regression, was used. One drawback is that SELECT is a supervised learning method and there is a danger that variables may be selected by 'chance'. It should be noted, however, that subsequent use of these parameters employed unsupervised learning methods which do not require knowledge of the class membership of the compounds. After selection, nine properties were chosen (see Table 2), and application of NLM and PCA display techniques to these parameters resulted in sharper clustering of separation categories (Figs. 3–5).

Inspection of the NLM (Fig. 3) shows that compounds with good separation characteristics fall close together, an indication that the chemical properties used to describe these compounds possess information that is relevant for the discrimination of separation classes. Compounds whose inter-point distances in N-dimensions are small will retain these characteristics in two dimensions. While it is not possible to perform the opposite operation of going from two-dimensional space to N-dimensional space for interpretation of the plot, NLM has the advantage that it may be used for predictive purposes through association rather than correlation. In addition to the 12 compounds shown in Table 1, results were obtained for another analogue whose structure cannot be revealed at present. Because a non-linear map is calculated by minimization of an error function relating two-dimensional inter-point distances to N-space inter-point distances it is not possible to directly plot this extra compound on Fig. 3. Instead, the entire map must be recalculated and this is shown in Fig. 4. It can be seen from this figure that this compound lies in a region of space which is not occupied by any of the other compounds in the set. However, its nearest neighbours belong to the good and moderate separators and it is a long way from the poor separators in the figure. A prediction of separation characteristics for this compound would be good ($\alpha > 1.4$) or moderate ($1.2 < \alpha < 1.4$). The experimental result for this compound is $\alpha =$

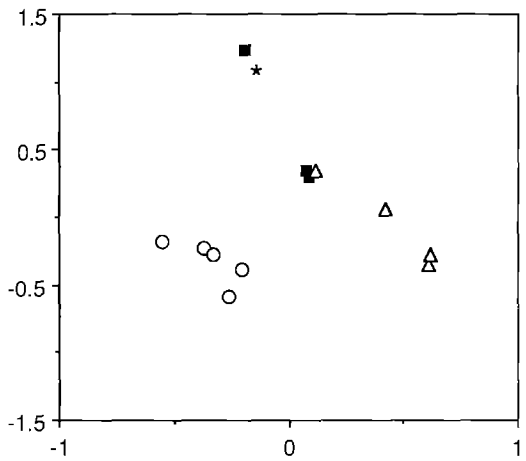


Fig. 5. Plot of compounds **1–12** on the first and third principal components extracted from nine parameters. The * represents the test compound. ■: good separator; △: moderate separator; ○: poor separator.

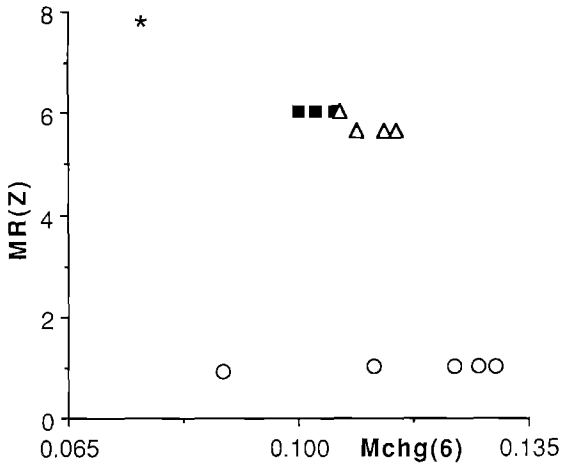


Fig. 6. Plot of MR(Z) against Mchg(6) for compounds **1–12**. The test compound is indicated with a *. ■: good separator; △: moderate separator; ○: poor separator.

1.42, which is on the divide between good and moderate separators. This result appears to be quite a reasonable prediction. Figure 4 illustrates another feature of the non-linear mapping technique and that is that the picture has changed due to the inclusion of one extra compound. The relative positions of the poor, moderate and good separators are retained. In other words, the map can be used predictively, but the figure has changed which may be disconcerting to those more accustomed to linear plotting techniques. This can be thought of as another disadvantage of the non-linear mapping method, although the effects are very data set-dependent and may not be seen at all.

The results of the principal components analysis may be visualized by plotting principal component (PC) scores against one another. This is shown in Fig. 5 where the scores for PC1 vs. PC3 indicate the usefulness of the nine properties (Table 2) for the classification of separation categories. The clustering of good and moderate separators is tighter than the clustering seen on the NLM, which may indicate that a *linear* combination of variables is required to describe the separation characteristics of this set. The prediction for the test compound in this case is that it would be a good separator. By definition, PCs are new variables calculated using linear combinations of the starting variables. In addition, each PC is orthogonal with all the other PCs and the first PC contains the largest amount of variance within the data set, with the remaining PCs containing progressively smaller amounts of variance. The individual loadings of each of the nine variables for PC1 and PC3 are summarized in Table 2. While it is difficult to interpret the significance of each PC in terms of chemical structure and separation characteristics, it is possible to determine the relative importance of each property within each PC. In addition, the weightings applied to each variable using the ARTHUR routine SELECT indicated that the partial charge of atom 6 (Mchg(6)) and the molar refractivity of substituent Z (MR(Z)) contained the most useful information for the classification of separation categories (data not shown). Indeed, a plot of Mchg(6) vs. MR(Z) (Fig. 6) demonstrates tight clustering of separation categories. In this plot

TABLE 2
LOADINGS OF EACH VARIABLE FOR PRINCIPAL COMPONENTS 1 AND 3[a]

PC1	Loading	Variance explained
Mchg(18)	−0.517	26.71
ymax	−0.461	21.27
Mchg(9)	−0.457	20.85
MR(Z)	0.376	14.15
Mchg(6)	−0.307	9.45
Malp(3)	0.158	2.51
Malp(9)	−0.138	1.89
c2(Y)	0.135	1.82
xl(Z)	−0.116	1.35
PC3	**Loading**	**Variance explained**
xl(Z)	−0.621	38.59
MR(Z)	0.585	34.18
Mchg(9)	0.353	12.48
Malp(3)	0.284	8.06
Mchg(18)	0.198	3.91
Mchg(6)	0.139	1.94
ymax	0.081	0.66
c2(Y)	0.041	0.16
Malp(9)	−0.009	0.01

[a] The loadings give the magnitude and direction of the correlation of a variable with the principal component. Variance explained is the contribution of that variable to the variance explained by the component. Malp(3) and Malp(9) = self-atom polarizability of atoms 3 and 9; MR(Z) = molar refractivity of substituent Z; ymax = maximum dimension for each molecule in the y direction; Mchg(6) = Mchg(9) and Mchg(18) = partial charge of atoms 6, 9 and 18; xl(Z) = minimum dimension for substituent Z in the x direction; c2(Y) = square of the partial charge on the Y substituent.

the test compound is closest to the good and moderate separators and is well removed from the poor separators. The advantage of this plot is the ease of interpretation of the relationship between each property and separation. As observed qualitatively, a low value of MR for substituent Z resulted in poor separation of enantiomeric pairs. Discrimination of moderate and good separation categories appears to be related to Mchg(6). A note of caution should be made here concerning attempts to assign causality to such relationships; it should be borne in mind that these properties are representatives of families of properties and care must also be used in the interpretation of Fig. 6.

In conclusion, we have shown that a CHIRALPAK AD column can be used for the resolution of the enantiomers derived from a number of benzimidazole sulphoxides and that the extent of resolution may be explained using a combination of measured and calculated physicochemical properties. The important descriptors, in terms of explanation of chromatographic behaviour, *may* reflect a steric and electronic component of the binding of these compounds to the stationary phase. Chiral analysis by this method can also be suitable for extensive pharmacodynamic and pharmacokinetic studies on the benzimidazole sulphoxides.

REFERENCES

1 Brandstrom, A., Lindberg, P., Junggren, U. and Wallmark, B., Scand. J. Gastroenterol., 21 (Suppl. 118) (1986) 54.
2 Simon, W.A., Keeling, D.J., Laing, S.M., Fallowfield, C. and Taylor, A.G., Biochem. Pharmacol., 39 (1990) 1799.
3 Huber, R., Müller, W., Banks, M.C., Rogers, S.G., Norwood, P.C. and Doyle, E., J. Chromatogr., 529 (1990) 389.
4 Hansch, C., In Chapman, N.B. and Shorter, J. (Eds.) Correlation Analysis in Chemistry: Recent Advances, Plenum Press, New York, 1978, p. 397.
5 Kikuchi, O., Quant. Struct.-Act. Relat., 6 (1987) 179.
6 Ford, M.G. and Livingstone, D.J., Quant. Struct.-Act. Relat., 9 (1990) 107.
7 a. Ife, R.J., Dyke, C.A., Keeling, D.J., Meenan, E., Meeson, M.L., Parsons, M.E., Price, C.A., Theobald, C.J. and Underwood, A.H., J. Med. Chem., 32 (1989) 1970.
b. Sturm, E., Kruger, U., Senn-Bilfinger, J., Figala, V., Klemm, K., Kohl, B., Rainer, G., Schaefer, H., Blake, T.J., Darkin, D.W., Ife, R.J., Leach, C.A., Mitchell, R.C., Pepper, E.S., Salter, C.J., Viney, N.J., Huttner, G. and Zsolnai, L., J. Org. Chem., 52 (1987) 4573.
c. Brandstrom, A., Bergman, N.-A., Lindberg, P., Grundevik, I., Johansson, S., Tekenbergs-Hjelte, L. and Ohlson, K., Acta Chem. Scand., 43 (1989) 549.
d. ZA 8403287/1984.
8 a. Vinter, J.G., Davis, A. and Saunders, M.R., J. Comput.- Aided Mol. Design, 1 (1985) 31.
b. Morley, S.D., Abraham, R.J., Haworth, I.S., Jackson, D.E., Saunders, M.R. and Vinter, J.G., J. Comput.- Aided Mol. Design, 5 (1991) 475.
9 Stewart, J.J.P., J. Comput.- Aided Mol. Design, 4 (1990) 1.
10 Livingstone, D.J., Evans, D.A. and Saunders, M.R., J. Chem. Soc. Perkin Trans. II, (1992) 1545.
11 Van de Waterbeemd, H. and Testa, B., In Testa, B. (Ed.) Advances in Drug Research, Vol. 16, Academic Press, New York, 1987, p. 87.
12 Application Guide for Chiral Column Selection published by Daicel Chemical Industries Limited, 1989.
13 The MEDCHEM program is available from Daylight Chemical Information Systems Inc., Irvine, CA 92713-7821, U.S.A.
14 Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
15 Camilleri, P., Watts, S.A. and Boraston, J.A., J. Chem. Soc. Perkin Trans II, (1988) 1699.
16 Livingstone, D.J. and Rahr, E., Quant. Struct.-Act. Relat., 7 (1988) 103.
17 Livingstone, D.J., Pestic. Sci., 27 (1989) 287.
18 Hudson, B., Livingstone, D.J. and Rahr, E., J. Comput.- Aided Mol. Design, 3 (1989) 55.
19 Kowalski, B.R. and Bender, C.F., Pattern Recognition, 8 (1976) 1.