Journal of Computer-Aided Molecular Design, 5 (1991) 371–380
ESCOM

J-CAMD 122

Automatic search for maximum similarity between molecular electrostatic potential distributions

Francesc Manaut, Ferran Sanz, Jaume José and Massimo Milesi

Institut Municipal d'Investigació Mèdica, Facultat de Medicina (U.A.B), P.Marítim 25–29, 08003 Barcelona, Spain

Received 7 February 1990
Accepted 21 December 1990

Key words: Molecular electrostatic potential; Computational chemistry software; Molecular similarity; Optimal relative position; Spearman coefficient; DHFR

SUMMARY

A new computer program has been developed to automatically obtain the relative position of two molecules in which the similarity between molecular electrostatic-potential distributions is greatest. These distributions are considered in a volume around the molecules, and the similarity is measured by the Spearman rank coefficient. The program has been tested using several pairs of molecules: water vs. water; phenylethylamine and phenylpropylamine vs. benzylamine; and methotrexate vs. dihydrofolic acid.

INTRODUCTION

It is commonly accepted that two molecules have the same biological activity if they reach the same receptor and interact with it in the same way. Molecular electrostatic potential (MEP) is a common tool for studying these interactions. Under this approximation, two drugs will present a similar affinity for the same receptor if they have similar MEP patterns [1–4]. In many cases it is preferable to have good similarity of MEP distributions than good overlap of the atoms of the molecules [5,6].

In a previous work [7], we developed a computer program for characterizing MEP patterns, localizing MEP minima around the molecule and describing the internal relationships between these minima. In another previous study [5] we defined and successfully used a similarity coefficient between MEP distributions of two molecules. This coefficient was evaluated by placing the molecules in a fixed relative position. In this work we present a computer program which has been developed to automatically obtain the maximum value of an electrostatic similarity coefficient, by moving the relative positions of both molecules.

METHODS

Quantum-chemical computations

MEP can be computed at several approximation levels. Our program can be used from the roughest approximation, which is based on point charges, to the most sophisticated expression, which defines MEP at a point R as:

$$V(R) = \sum \frac{Z_I}{|R-R_I|} - \int \frac{\delta(r)}{|R-r|} dr \quad (1)$$

where Z_I are the nuclear charges, R_I are the nuclear coordinates, and $\delta(r)$ is the electron density distribution of the molecule. Electron density distributions can be obtained both from semiempirical or ab initio wavefunctions.

Equation (1) has been used in this study to compute MEP. Electron density distributions have been obtained from ab initio wavefunctions using the STO-3G basis set. These wavefunctions have been computed using the GAUSSIAN 86 computer program [8] running on a VAX system.

Electrostatic similarity

In addition to the classical visual comparison between the shapes represented by isopotential lines over 2D maps or by isopotential surfaces in the space around the molecules, several methods for analyzing the electrostatic similarity between molecules have been devised [3,5,6,9–12].

If MEP is computed using Eq. (1) it is not possible to compute electrostatic similarity in an analogous way as described by Carbo et al. [13] for the electron density similarity, because there are as many discontinuities as the system has nuclei. Attempts have been made to overcome this limitation by using sets of functions associated with the MEP [9].

Another possibility is to compute and compare MEP values over a surface around the molecules [3,10]. Using this approach it is difficult to quantitatively compare two dissimilar surfaces, and furthermore, the analysis of MEP can be interesting outside this surface.

In the case of the SEA program [11] the alignment of two 3D structures is set using both steric and electrostatic factors. The electrostatic component is computed only by electrostatic repulsion between the partial charges over the nuclei of the two molecules to be compared.

In another group of techniques [5,6,12], similarity coefficients are defined considering a volume around the molecules in order to use more complete electrostatic information.

Our program searches the alignment between pairs of molecules maximizing a similarity coefficient between sets of MEP values computed at the points of a 3D homogeneous grid into a volume around the molecules to be compared. In our electrostatic similarity algorithm, and especially when we use Eq. (1) to compute MEP, steric information is also included because the positions of nuclei are indirectly taken into account by their high MEP values.

Similarity coefficients

There are many coefficients for measuring the similarity between two distributions of a continuous variable represented by a finite set of pairs of values [14]. Coefficients such as Pearson and Cosine cannot be used for comparing MEP distributions because normal distribution of the values is a previous condition for applying these coefficients; MEP does not have this kind of distribution, as it presents a few points close to the nuclei with extremely high MEP values (Fig. 1).

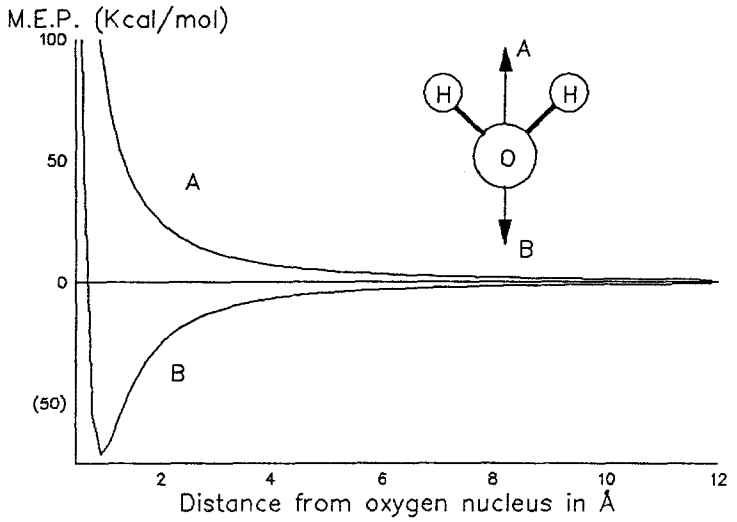


Fig. 1. Variation of MEP values as a function of distance from the oxygen atom of a water molecule. A: in a direction opposite to lone pairs; B: in the direction of lone pairs.

Another similarity coefficient, the Spearman rank correlation coefficient, is scale-invariant and does not require a normal distribution of data. These characteristics make the Spearman coefficient suitable for comparing MEP distributions. In the case where ties are not corrected, the coefficient r_s is computed using the following equation:

$$r_s = 1 - \frac{6\sum d_i^2}{n^3 - n} \tag{2}$$

In this algorithm, MEP values of the two molecules are sorted establishing their ranks. Thus, d_i is the difference between these ranks at the same point i, and n is the number of pairs of points. This coefficient has already been used by other authors for evaluating molecular similarities [10,15,16].

Considered zone

Due to the fact that MEP is a property which takes an infinite value at the nuclei and tends quickly to zero when we move away from them (Fig. 1), the considered volume around each molecule should be limited.

The external shell limiting this volume has been set by the union of spheres with radii proportional to the van der Waals radii of the atoms. Points beyond this limit are also taken into account if their absolute MEP values are higher than a pre-established value, because these points are also important to characterize the electrostatic pattern (Fig. 2). To avoid dealing with excessively high MEP values, it is also possible to create, in the same way, an inner shell close to the nuclei, in order to exclude the internal points (see also Figs. 1 and 2).

After analyzing several values, the external limit of the shell was set at twice the van der Waals radii, and the absolute MEP value for including points outside the limits was set to 5 kcal/mol, a

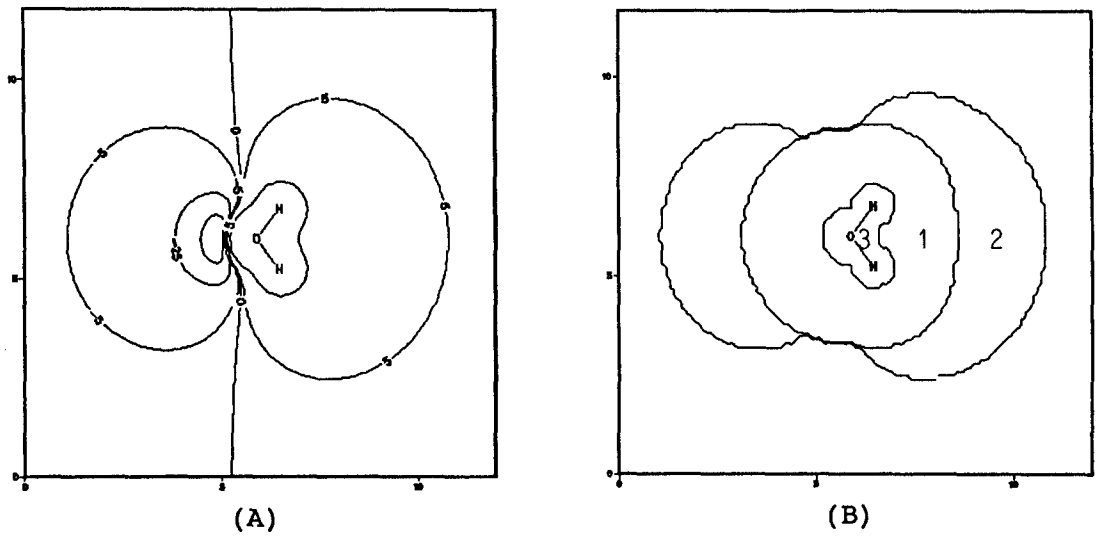


Fig. 2. A: MEP map of a molecule of water in the molecular plane. B: Considered volume around the molecule. 1, Zone selected using distance criteria; 2, zone added because of high absolute MEP values; 3, internal shell that can be optionally excluded because of extremely high MEP values.

convenient value for neutral molecules. Using the Spearman coefficient as the parameter for measuring similarity, there was no need to use the inner shell for excluding points close to the nuclei.

Initial position

Using the gradient method to maximize similarity, the initial relative position of both molecules is highly relevant. A non-convenient initial position can lead to a local maximum, or the process can work very slowly, whereas a good initial position can converge quickly to the maximum.

To avoid the problem of local maxima, previous works proposed the use of the Monte Carlo procedure [11], but in our case it is difficult to use because of the high computational time required. We propose two alternative strategies for selecting relative initial positions. If structures are similar, we can start with a classical matching of the nuclei by the least-squares method. If structures are dissimilar, and because of the nature of the MEP, we can put both molecules with their charge centers at the same position and with the same orientation of their dipole moment vectors.

Optimization procedure

Our computer program starts by placing both molecules in the convenient relative position, following one of the previous criteria. Then, the program computes the volume around each molecule to be taken into account. Next, it computes the MEP values for both molecules at the points of a 3D grid defined into the union of considered volumes.

Using the selected coefficient, the initial similarity between the MEP values of both molecules in the considered volume is then evaluated.

Subsequently, one of the molecules is kept fixed and the other is moved within the six degrees

of freedom (the three translations and the three rotations). Thus, the program begins an iterative process for maximizing the similarity coefficient using a gradient method. Translations are defined along three perpendicular axes placed at the charge center of the fixed molecule, and rotations are defined around these axes.

Movements of one of the molecules imply a loss of superposition between the precomputed points of the two grids and a variation of the union of the considered volumes. Thus, at each step of the maximization process, a new volume is set and a new coincident grid is calculated. The MEP values of some points are determined by linear interpolation from the previous ones in order to decrease the computational time.

The optimization procedure ends when either the variation of the coefficient between two steps or the gradient norm is lower than a pre-established value. Then, the program prints out some parameters of the optimization process, like the final similarity coefficient and the movements necessary to reach this maximum (Table 1 shows an example of printout).

Computational aspects

The computer program has been devised in FORTRAN/VAX using a modular procedure. Each task is accomplished by a separate subroutine. Thus, it is easy to change the similarity coefficient used, the method for computing the MEP or the maximization procedure.

TABLE 1
EXAMPLE OF PRINTOUT FROM MEPCOMP COMPUTER PROGRAM

PROGRAM MEPCOMP

Molecules to be compared :

Benzylamine and Phenylethylamine

	Translations (Å)			Rotations (Degrees)			
Step	X	Y	Z	α	β	δ	r_s
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.592
1	-0.0953	-0.0979	-0.1701	3.8859	-0.9384	-0.0284	0.634
2	-0.1029	-0.2315	-0.4402	7.5148	-1.3169	-2.4892	0.665
3	-0.1210	-0.1252	-0.5671	8.7530	2.2179	-1.0987	0.674
4	-0.1185	-0.1238	-0.5541	8.8778	2.0816	-1.0605	0.674
5	-0.1196	-0.1313	-0.5561	8.8219	2.0816	-1.0605	0.674
6	-0.1196	-0.1313	-0.5561	8.7646	2.0816	-1.0605	0.674

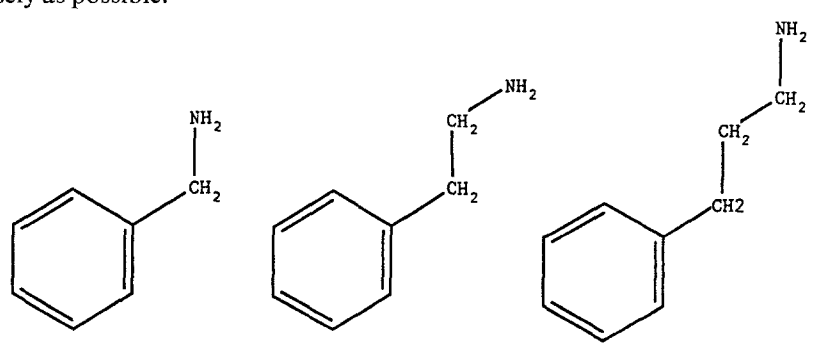
TABLE 2
MOVEMENTS TO RECOVER THE SUPERPOSITION OF TWO MOLECULES OF WATER AFTER TRANSLATIONS OF 1 Å IN EACH CARTESIAN DIRECTION AND ROTATIONS OF 45° AROUND EACH AXIS

Initial position	Final movements						Final r_s
	X	Y	Z	α	β	δ	
1 Å translated in direction							
X	*−0.99*	0.00	0.00	0.0	−0.6	0.1	0.984
Y	0.00	*−1.00*	0.01	−0.3	0.2	0.1	0.999
Z	0.02	0.01	*−1.00*	−0.3	−0.2	−0.6	0.997
45° rotated around axis							
α	−0.02	0.00	−0.01	*−44.9*	0.1	0.2	0.999
β	−0.02	0.00	0.05	0.3	*−44.8*	−0.5	0.979
δ	−0.02	0.00	−0.01	−0.1	0.2	*−42.7*	0.988

RESULTS

To validate the whole computer program, six initial tests were carried out with two molecules of water. Each test consisted in placing the two molecules of water separated by 1 Å in each Cartesian axis or rotated 45 sexagesimal degrees around these same axes as starting positions. In all instances, the program recovers the superposition of both molecules after executing a movement in the opposite direction (Table 2). Small defects in superposition at the end of the processes are due to the finite number of considered points in the grid and also to the linear interpolation of some MEP values.

We have also considered the search for the maximum electrostatic similarity between some interesting neuro-transmitters such as benzylamine (BZA), phenylethylamine (PHEEA) and phenylpropylamine (PHEPA; Fig. 3). All three molecules have been considered in a conformation in which the alkylamine fragment is extended and perpendicular to the aromatic ring. To set the initial position, we matched the aromatic rings and the alkyl chains of both molecules to be compared as closely as possible.



Benzylamine (BZA) Phenylethylamine (PHEEA) Phenylpropylamine (PHEPA)

Fig. 3. Structures of compared neuro-transmitters.

TABLE 3
INITIAL ELECTROSTATIC SIMILARITY IN A GOOD STRUCTURAL SUPERPOSITION AND FINAL SIMILARITY AFTER REACHING THE RELATIVE POSITION OF MAXIMUM ELECTROSTATIC SIMILARITY

	PHEEA/BZA	PHEPA/BZA
Initial position	0.59	0.57
Final position	0.67	0.79

Table 3 shows the results of comparing BZA vs. PHEEA and PHEPA. At the initial position, the similarity is greater between BZA and PHEEA than between BZA and PHEPA. This result is logical because PHEEA is structurally closer to BZA than PHEPA. However, after the maximization process and, once obtained the maximum electrostatic similarity, PHEPA is more like BZA than like PHEEA. This final result is in agreement with general knowledge about chemical reac-

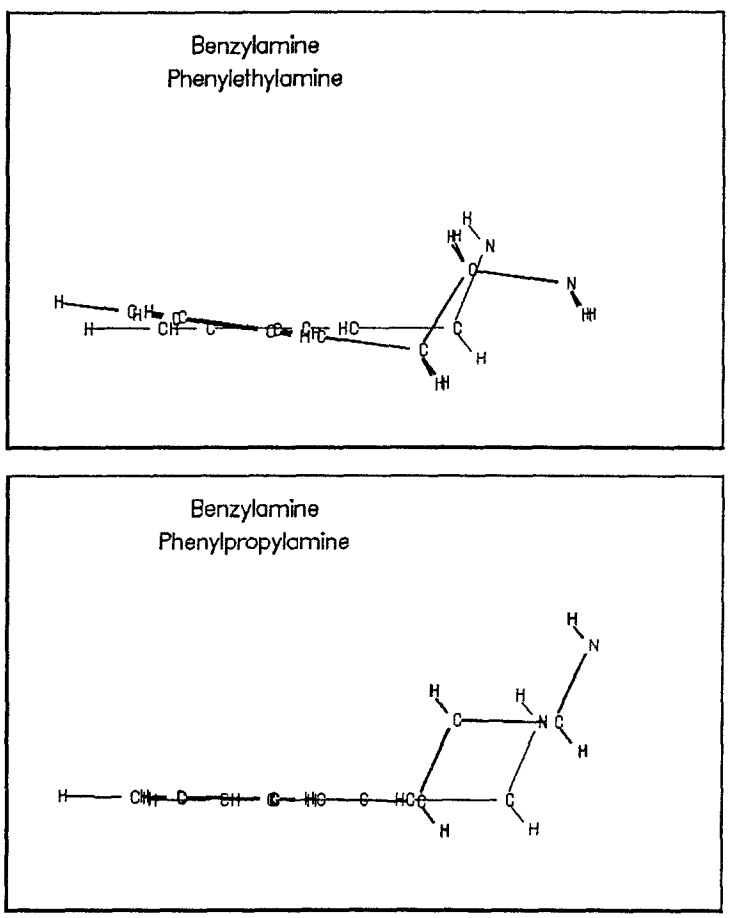


Fig. 4. Positions of maximum electrostatic similarity.

tivity, because alkylamine chains of one carbon (like BZA) and those of three carbons (like PHEPA) have the lone pair of the nitrogen on the same side, whereas those of one carbon and those of two carbons (like PHEEA) have the lone pair on opposite sides. The final relative positions are shown in Fig. 4, in which it can be seen that our program produces a final position which is a con-

TABLE 4
CASE A: SIMILARITY MAXIMIZATION STARTING WITH AN OVERLAP OF THE PTERIDINE RINGS
CASE B: MAXIMIZATION STARTING FROM 180° ROTATED POSITIONS

Molecules to be compared :

Dihydrofolic acid and Methotrexate

	Translations (Å)			Rotations (Degrees)			
Step	X	Y	Z	α	β	δ	r_s
CASE A							
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.105
10	1.2622	-0.1563	-0.3205	2.6284	-19.1721	7.9947	0.176
20	1.5300	-0.0959	-0.7955	50.9957	-21.7463	2.3072	0.318
30	1.9232	0.2175	-0.4378	104.6497	2.7966	17.8894	0.555
40	1.9278	-0.0381	-0.2261	114.2706	3.9718	18.8514	0.587
50	1.9452	-0.0937	-0.0215	141.5839	3.5632	13.0982	0.635
60	1.9066	-0.1273	0.1867	157.0712	5.9159	11.1138	0.669
70	1.8396	-0.0773	0.3630	162.4339	5.9711	10.4631	0.678
80	1.8742	-0.1410	0.4258	166.5949	6.9087	10.0175	0.681
87	1.8654	-0.1521	0.4556	166.7563	6.5470	10.3283	0.682
CASE B							
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.187
5	-0.6693	-1.2693	-0.1448	-12.8675	-8.9477	-40.4613	0.635
10	-1.3392	-1.8772	-0.3007	-9.7692	-4.4645	-38.1654	0.712
15	-1.4235	-1.8464	-0.2885	-9.3506	-3.0164	-37.8246	0.715
18	-1.4462	-1.8489	-0.2910	-9.4698	-3.2440	-37.6561	0.716

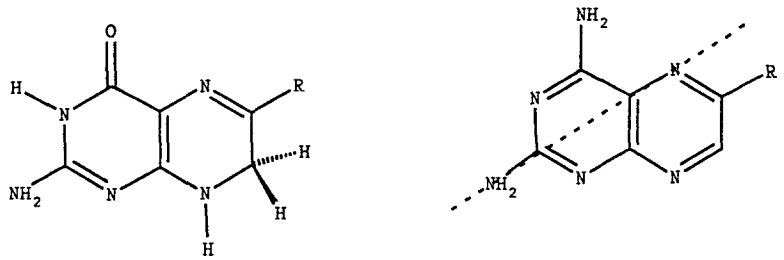


Fig. 5. Analysed fragments of (left) dihydrofolic acid (FH_2) and (right) methotrexate (MTX). The dashed line shows the rotation axis.

sensus between the maximum superpositions of the nitrogen lone pairs, the π systems of aromatic rings, and the backbone of both molecules.

The last test we have considered is a preliminary analysis of the similarity between the pteridine rings of dihydrofolic acid (FH_2) and methotrexate (MTX; Fig. 5). This test is of special interest because there is crystallographic information about the relative position of binding of both molecules to dihydrofolate reductase (DHFR) [17], and it is accepted that the pteridine rings of both molecules are related by a 180° rotation around the axis roughly coincident with the C2-NH_2 bond [18] (Fig. 5).

To apply our program for searching the maximum similarity between FH_2 and MTX, we have chosen two different relative positions to start the computation: the first fits the pteridine rings of both molecules (case A), and the other rotates the pteridine rings as suggested by previous workers (case B).

The similarity at the beginning of the process is higher in case B than in case A, in agreement with experimental data. At the end of the maximization procedure, the program produces minor movements in case B, whereas in case A the program automatically generates a rotation of 167°, exclusively driven by the electrostatic-similarity gradient (Table 4). Although the two final positions present pteridine rings in 'backward' relative positions, they are not coincident and are local maxima. These local maxima could be overcome by using a Monte Carlo method.

CONCLUSIONS

The comparison between biomolecules interacting with a common receptor must be performed in an appropriate relative position, which often could be the position of maximum electrostatic similarity. The program we present has been demonstrated to be a powerful tool to automatically search this relative position. In many cases these positions of maximum electrostatic similarity differ from those of maximal structural similarity, the former being more useful for comparing molecular reactivity.

REFERENCES

1 Dean, P.M., Molecular Foundations of Drug–Receptor Interaction, Cambridge University Press, Cambridge, 1987.
2 Martin, M., Sanz, F., Campillo, M., Pardo, L., Pérez, J. and Turmo, J., Int. J. Quantum Chem., 23 (1983) 1627.

3 Arteca, G.A., Jammal, V.B., Mezey, P.G., Yadav, J.S., Hermsmeier, M.A. and Gund, T.M., J. Mol. Graphics, 6 (1988) 45.
4 Liwo, A., Tempczyk, A. and Grzonka, Z., J. Comput.-Aided Mol. Design, 3 (1989) 261.
5 Sanz, F., Martin, M., Lapeña, F. and Manaut, F., Quant. Struct.–Act. Relat., 5 (1986) 54.
6 Kato, Y., Itai, A. and Iitaka, Y., Tetrahedron, 43 (1987) 5229.
7 Sanz, F., Manaut, F., José, J., Segura, J., Carbó, M. and De la Torre, R., J. Mol. Struct. (THEOCHEM), 170 (1988) 171.
8 Frish, M.J., Binkley, J.S., Schlegel, H.B., Raghavachari, K., Melius, C.F., Martin, R.L., Stewart, J.P., Bobrowicz, F.W., Rohlfing, C.M., Kahn, L.R., Defrees, D.J., Seeger, R., Whitesite, R.A., Fox, D.J., Fleuder, E.M. and Pople, J.A., GAUSSIAN 86, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, PA, 1984.
9 Carbó, R., Lapeña, F. and Suñe, E., Afinidad, 406 (1986) 483.
10 Namasivayam, S. and Dean, P.M., J. Mol. Graphics, 4 (1986) 46.
11 Graham, M.S., Merck Sharp & Dohme, SEA program, QCPE No. 567 (1988).
12 Nakayama, A and Richards, W.G., Quant. Struct.–Act. Relat., 6 (1987) 153.
13 Carbó, R., Leyda, L. and Arnau, M., Int. J. Quant. Chem., 17 (1980) 1185.
14 SPSS Statistical Algorithms, SPSS Inc., Chicago, IL, 1985.
15 Dean, P.M., Callow, P. and Chau, P.-L., J. Mol. Graphics, 6 (1988) 28.
16 Hazdi, D., Koller, J. and Hodosek, M., In Fauchère, J.L. (Ed.) QSAR: Quantitative Structure–Activity Relationships in Drug Design, Alan R. Liss Inc., New York, 1989, pp. 259-263.
17 Bolin, J.T., Filman, D.J., Matthews, D.A. and Kraut, J., J. Biol. Chem., 257 (1982) 13663.
18 Kuyper, L.F., In Perun, T.J. and Propst, C.L. (Eds.) Computer-Aided Drug Design, Marcel Dekker Inc., New York, 1989, pp. 327-369.