Journal of Computer-Aided Molecular Design, 7 (1993) 337–347
ESCOM

J-CAMD 202

MEPSIM: A computational package for analysis and comparison of molecular electrostatic potentials

Ferran Sanz[a,*], Francesc Manaut[a], Jesús Rodríguez[a,b], Estrella Lozoya[a] and Elena López-de-Briñas[a]

[a]*Institut Municipal d'Investigació Mèdica, Facultat de Medicina (U.A.B.), Passeig Marítim 25, E-08003 Barcelona, Spain*
[b]*Departamento de Química Física, Facultade de Química, Universidade de Santiago, E-15706 Santiago de Compostela, Galicia, Spain*

Received 23 October 1992
Accepted 19 January 1993

Key words: Molecular electrostatic potential; Computational chemistry software; Molecular alignment; Molecular similarity; Spearman coefficient

SUMMARY

MEPSIM is a computational system which allows an integrated computation, analysis, and comparison of molecular electrostatic potential (MEP) distributions. It includes several modules. Module MEPPLA supplies MEP values for the points of a grid defined on a plane which is specified by a set of three points. The results of this program can easily be converted into MEP maps using third-parties graphical software. Module MEPMIN allows to find automatically the MEP minima of a molecular system. It supplies the cartesian coordinates of these minima, their values, and all the geometrical relationships between them (distances, angles, and dihedral angles). Module MEPCOMP computes a similarity coefficient between the MEP distributions of two molecules and finds their relative position that maximizes the similarity. Module MEPCONF performs the same process as MEPCOMP, considering not only the relative position of both molecules but also a conformational degree of freedom of one of them. The most recently developed module, MEPPAR, is another modification of MEPCOMP in order to compute the MEP similarity between two molecules, but only taking into account a particular plane. The latter module is particularly useful to compare MEP distributions generated by π systems of aromatic rings. MEPSIM can use several wavefunction computation approaches to obtain MEP distributions. MEPSIM has a menu type interface to simplify the following tasks: creation of input files from output files of external programs (GAUSSIAN and AMPAC/MOPAC), setting the parameters for the current computation, and submitting jobs to the batch queues of the computer. MEPSIM has been coded in FORTRAN and its current version runs on VMS/VAX computers.

INTRODUCTION

Molecular electrostatic potential (MEP) is a common tool for describing the electrostatic

*To whom correspondence should be addressed.

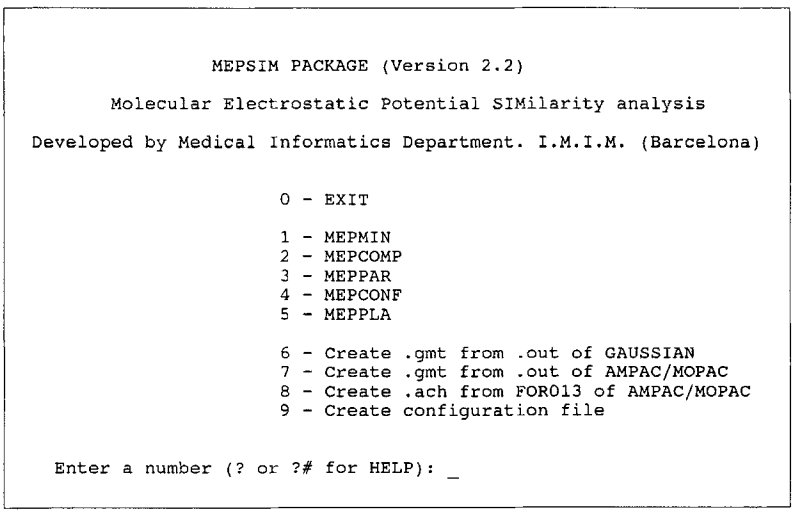

Fig. 1. Main menu of MEPSIM.

interaction potentialities of biomolecules. It is relevant in structure–activity relationship studies, since two compounds would present a similar affinity for the same biological receptor if they have close MEP distributions [1–5]. The ordinary visual comparison between the electrostatic patterns of molecules represented by isopotential lines in two-dimensional maps, or by isopotential surfaces in the space around the molecules, is intuitive but also subjective. For this reason, the comparison of graphical representations of MEP ought to be complemented by an objective quantitative analysis of MEP distributions and their similarities. Our group has developed and tested several algorithms and computer programs to analyze MEP distributions [4], and to compute electrostatic similarity coefficients between molecules finding the positions that maximize their electrostatic similarity [5–8].

The first program, called MEPMIN [4], allows automatic finding of the MEP minima of a molecular system. Another approach to assess the electrostatic similarity is the MEPCOMP program [6]. It was developed to compute a similarity coefficient between the MEP distributions of two molecules and to find their relative positions that maximize the electrostatic similarity. Several succesful applications of this program were obtained [5,6]. The modification of this program, in order to take into account the conformational flexibility of the molecules to be compared, gave rise to the development of the MEPCONF program [7].

This study describes the computational system MEPSIM that integrates the previously described programs into a common interface. MEPSIM includes improved new versions of those programs as well as two other modules (MEPPLA, MEPPAR). The most recently developed module, MEPPAR, computes the similarity between MEP distributions in specific planes and scans relative positions of the two molecules to find those of maximum similarity. In the case of the previously published modules (MEPMIN, MEPCOMP and MEPCONF), the present paper updates their characteristics and describes some unpublished computational peculiarities. MEPPAR is described with major detail because only incomplete preliminary information has been formerly presented [8].

MEPSIM is a computational system that allows an integrated computation, analysis, and comparison of MEP distributions. It has a menu type interface (see Fig. 1) which inlcudes options for running the previously described modules. Options from 1 to 5 generate command files (molname.COM) for running a computation with those modules and proposes to submit these files to a batch queue of the computer. Other options of the main menu of MEPSIM (from 6 to 8) simplify tasks as the creation of input files from outputs of other programs (GAUSSIAN and AMPAC/MOPAC). These input files include the cartesian coordinates of the molecules, and their eigenvectors or atomic charges. The remaining parameters that customize each computation are contained in a configuration file (name.CFG) whose creation is helped by option 9 of the main menu. MEPSIM has been coded in FORTRAN and its current version runs on VAX/VMS computers.

MODULE MEPMIN

This module finds automatically the MEP minima of a molecular system, and supplies the cartesian coordinates of these minima, their values and all the geometrical relationships between them (distances, angles, and dihedral angles) [4]. MEPMIN allows only the use of the exact definition of MEP, which computes separately the contributions of the nuclear charges and that of the electron density distribution. The alternative use of atomic point charges to compute the MEP distribution is not adequate because this approximation is very rough and does not provide the MEP minima corresponding to lone pairs.

In the current version of MEPMIN, as in the rest of MEPSIM modules, the electron density distribution included in the MEP formula, can be obtained from both semiempirical (MNDO and AM1) and ab initio wavefunctions (STO-nG, 3-21G and n-31G, with or without polarization functions).

The input and output files of MEPMIN are schematically presented in Fig. 2. The wavefunction computation method and the dimensions and density of the three-dimensional grid around the molecule are set in the configuration file (name.CFG). Default values for outermost x, y, and z of the grid are the lowest of the molecule minus 3 Å and the greatest plus 3 Å. Default values of the step sizes are 0.5 Å. MEP values in the points of that grid are computed and scanned in order to find the zones with minimum MEP. It should be pointed out that any zone may not be detected when the density of the scanned grid is not sufficient. Inside these zones, precise positions of MEP minima are located by a gradient method. The molname.OUP output file, which contains the MEP values at the points of the grid, is useful as input for producing graphical representations of MEP. Another output file, molname.MIN, contains the cartesian coordinates of minima in a format that is appropriate for incorporating them as dummy atoms in molecular plots.

MODULE MEPPLA

This module supplies the MEP values for the points of a grid defined on a plane which is specified by the coordinates of any set of three points. These points can be atoms, MEP minima or any other. Plane dimensions are automatically adjusted to the size of the molecule projection into the plane. The MEP computation can be performed with the point charges approximation or with the expression that considers separately the nuclear charges and the electron density distri-

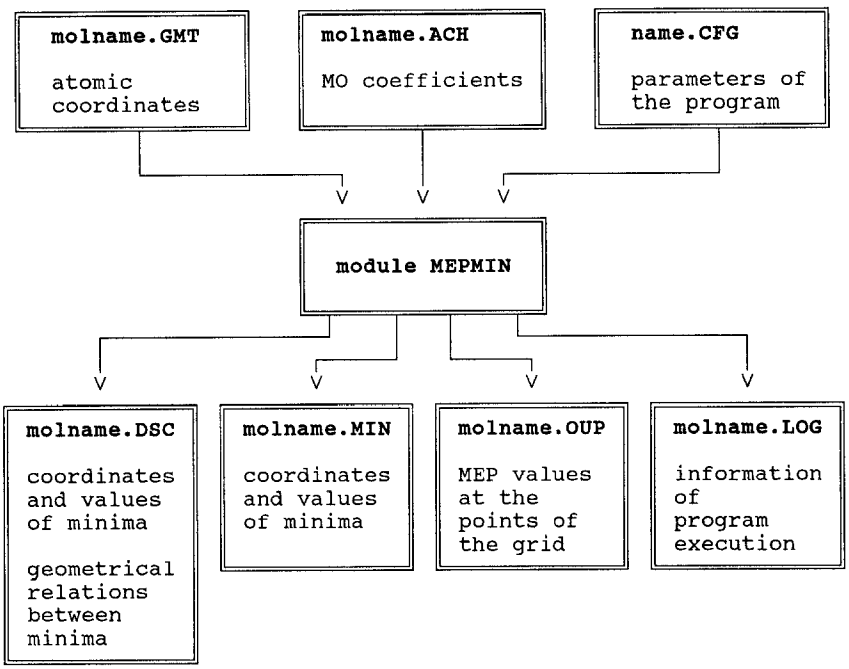

Fig. 2. Input and output files of MEPMIN.

bution. This distribution can be obtained from any of the wavefunctions enumerated in the MEPMIN section. The results of this program can easily be converted into isopotential lines using third parties graphical software. Figure 3 shows an example of a MEP map obtained using

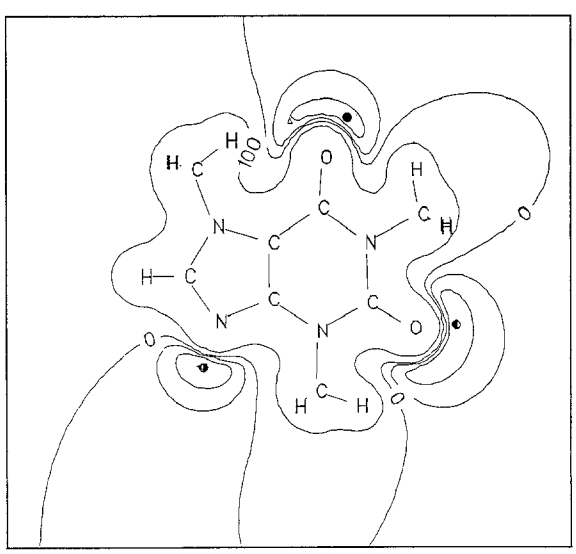

Fig. 3. MEP map of caffeine in a plane containing the three main MEP minima (indicated by dots). This map has been generated from a MEPPLA computation using a 3-21G wavefunction.

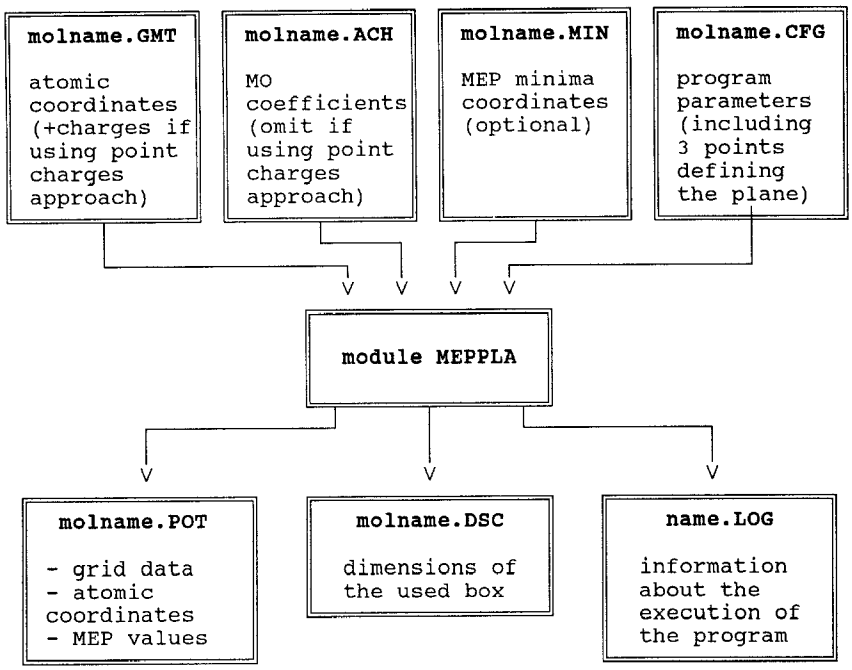

Fig. 4. Input and output files of MEPPLA.

this program and the CA-DISSPLA graphical routines library from Computer Associates. In that example, the plane that includes the three main minima of caffeine is shown.

Figure 4 shows the input and output files of MEPPLA. The main output file is molname.POT, which includes the following data:

– Number of points of the grid along the two axis of the selected plane.

– Step size.

– Number of atoms of the molecule and their cartesian coordinates defined as follows: x and y are the coordinates in the selected plane, z is the distance to the plane.

– When a molname.MIN file has been used as input, the next parameter is the number of minima followed by their coordinates in the selected plane.

– The rest of the file includes MEP values and their corresponding coordinates in the plane.

The information contained in the molname.POT file is the one that is required to produce plots such as depicted in Fig. 3.

ASSESSMENT OF ELECTROSTATIC SIMILARITY

MEPSIM includes three modules (MEPCOMP, MEPCONF, and MEPPAR) to compute the similarity between MEP distributions of two molecules and to find their relative positions that maximize that similarity. Both MEPCONF and MEPPAR are derived from MEPCOMP and share an important part of computational structure and code. The three modules assess similarity by computing the Spearman rank correlation coefficient between the MEP values of both mole-

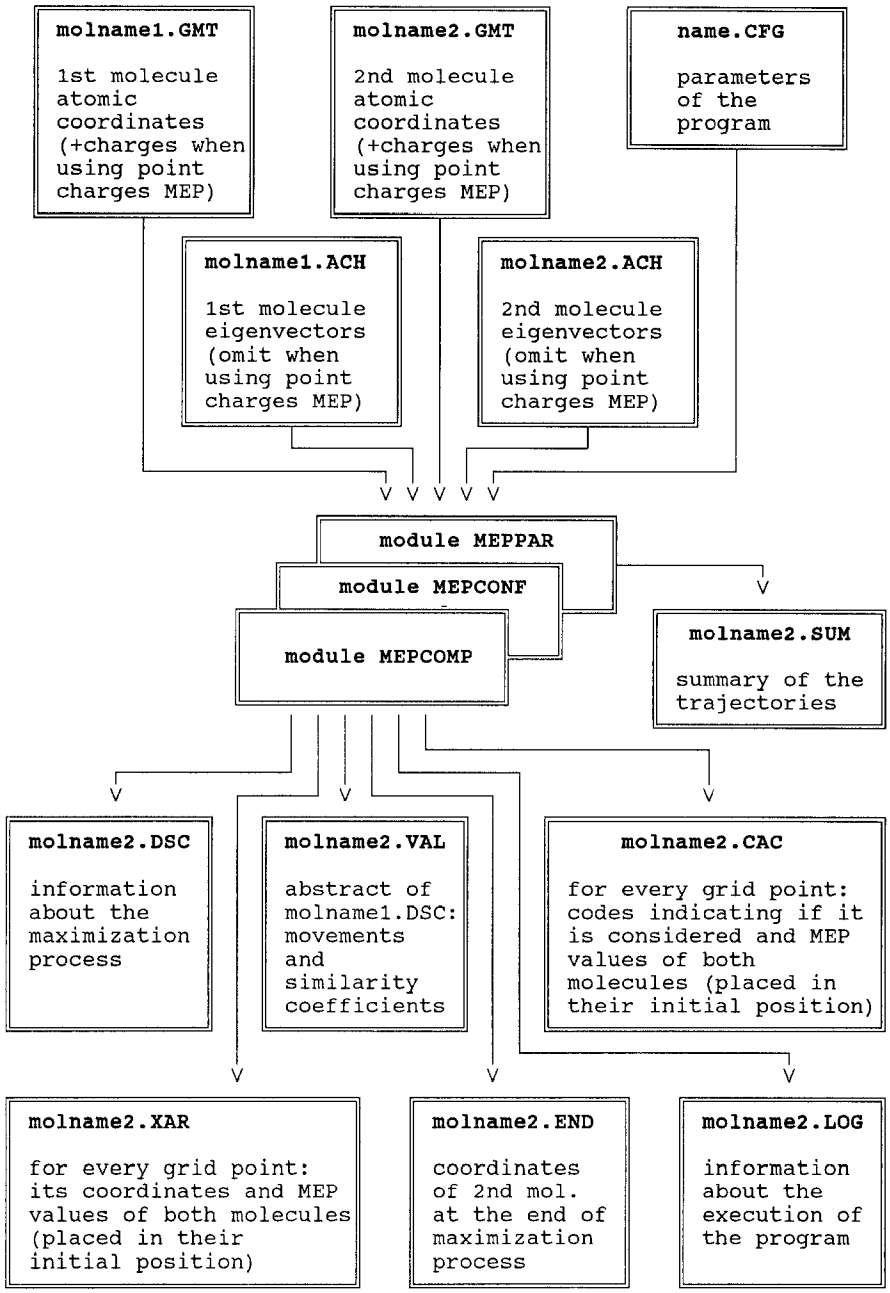

Fig. 5. Input and output files of MEPCOMP, MEPCONF, and MEPPAR.

cules computed in the points of a common grid. The Spearman coefficient, r_s, is computed using the following expression:

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

In the former 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.

The search for the relative positions of two molecules that maximizes their similarity is performed by a gradient method.

Another feature that is common to MEPCOMP, MEPCONF, and MEPPAR modules is their input and output files (see Fig. 5), with the only exception of the molname2.SUM file which is exclusive to the MEPPAR module. The standard output file is the molname2.VAL that includes the relative positions and similarity coefficients along the maximization processes. Another useful output file is molname2.END which contains the coordinates of the second molecule at the end of the maximization process. This information is necessary to produce joint plots of the two molecules in their positions of electrostatic alignment.

The previously published modules, MEPCOMP [6] and MEPCONF [7], have been incorporated into MEPSIM with the only modification that, in the case of MEPCOMP, any usual semiempirical or ab initio wavefunction can now be used. MEPCONF still maintains its limitations, that is, MEP has to be computed with the point charges approximation and only one conformational degree of freedom of one of the molecules can be considered.

MODULE MEPPAR

The most recently developed module, MEPPAR [8], is another modification of MEPCOMP in order to compute the MEP similarity between two molecules, but only taking into account a particular plane. In the current version of this module, the comparison between MEP distribu-

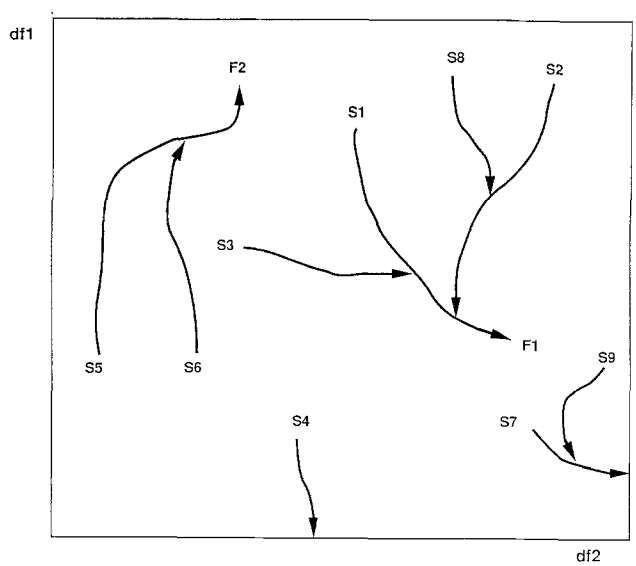

Fig. 6. Scheme of MEPPAR strategy (only two degrees of freedom are depicted). The three possibilities of stopping a trajectory are shown: starting positions S1 reaches maximum F1; S2 and S3 converge with the previous one; S4 stops because it produces a displacement out of range.

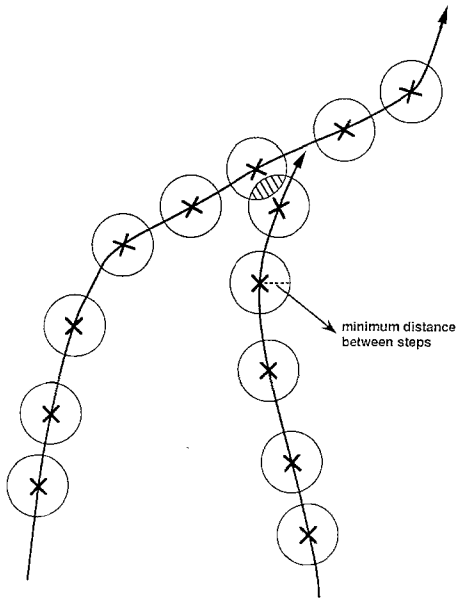

Fig. 7. Scheme of MEPPAR strategy to store steps of a maximization trajectory and to detect convergence.

tions is performed in an xz plane defined by a certain value y that is set in the configuration file (name.CFG). The movements of the molecules in order to search relative positions of maximum electrostatic similarity are restricted to translations along axes x and z, and rotations around an axis y that is placed in the charge center of the first molecule. These features make MEPPAR useful to compare MEP distributions generated by π systems of aromatic rings. The aromatic rings should be placed in the xz plane, and the value y that is set in the configuration file defines a parallel plane where the comparison will be carried out. A meaningful plane where to perform the comparison is that containing approximately the MEP minima generated by the π electrons of the aromatic system. The distance where these minima are located is about 1.5 Å from the aromatic rings, but it can be determined with greater accuracy using the MEPMIN module.

Similarity computation is performed with the points of a two-dimensional grid defined into the union of two areas, one for each molecule. To define these areas of the considered plane, we follow rules analogous to those of MEPCOMP. For each molecule, we consider zones that satisfy one or several of the following criteria:

– Distances to any nuclei lower than its van der Waals radius multiplied by a factor (the default value is 2).

– MEP value lower than a pre-established value (−5 kcal/mol by default).

– MEP value greater than another value (+5 kcal/mol by default).

Criteria to be applied and parameter values are set in the configuration file.

Taking advantage of the reduction of the computational time required to compare planes instead of volumes, and to optimize three degrees of freedom instead of the six degrees of MEPCOMP or the seven of MEPCONF, we have introduced a new strategy with the aim of overcoming the problem of local maxima of similarity that the gradient method sometimes provides. This strategy consists in performing a series of maximizations from several starting

cycle	step	tran x	tran z	rot	coeff	convergence criterium
1	7	0.334	0.247	-7.7	0.5781	opt. completed-no gradient
2	1	6.696	3.669	148.9	-0.6886	translation out of range
3	11	0.050	0.025	7.5	0.5205	conver. cycle 1 step 1
4	19	-0.293	0.710	187.0	0.4493	opt. completed-no gradient
5	6	0.437	0.029	351.5	0.5627	conver. cycle 1 step 2
6	5	-0.731	3.097	201.3	0.0051	conver. cycle 4 step 8
7	24	0.139	0.828	292.9	0.7134	opt. completed-no gradient
8	9	-0.266	0.882	182.7	0.4551	conver. cycle 4 step 11
9	14	0.228	0.408	283.4	0.6785	conver. cycle 7 step 10
10	17	0.141	0.329	278.5	0.6658	conver. cycle 9 step 11
11	4	-3.669	1.662	286.6	0.0094	conver. cycle 7 step 3
12	9	0.700	0.057	-13.3	0.5599	opt. completed-no gradient
13	29	1.452	0.090	128.1	0.7744	opt. completed-no gradient
14	10	-6.178	-3.348	134.0	-0.8550	opt. completed-no gradient
15	10	0.786	-0.023	180.3	0.3522	conver. cycle 8 step 5
16	1	-0.334	-0.369	345.9	0.3126	conver. cycle 5 step 6
17	30	0.880	0.126	136.7	0.7101	conver. cycle 13 step 9
18	12	4.828	-4.837	251.6	-0.7446	opt. completed-no gradient
19	17	1.346	0.641	35.0	0.3113	conver. cycle 3 step 6
20	5	0.679	0.347	128.8	0.6830	conver. cycle 13 step 9
21	2	-6.737	2.571	356.6	-0.8233	translation out of range
22	12	1.353	0.281	124.5	0.7544	conver. cycle 13 step 13
23	10	-0.426	-1.888	205.3	0.0969	conver. cycle 17 step 9
24	17	-0.094	0.369	277.0	0.6533	conver. cycle 9 step 11
25	18	0.951	0.646	218.8	0.3563	opt. completed-no gradient
26	14	1.422	0.068	119.0	0.7537	conver. cycle 13 step 12
27	34	0.214	0.917	295.1	0.7145	opt. completed-no gradient
28	20	1.454	0.043	127.4	0.7731	conver. cycle 13 step 11
29	8	0.327	-0.140	150.3	0.5488	conver. cycle 17 step 16
30	25	0.415	0.359	286.9	0.6651	conver. cycle 7 step 10
31	16	2.300	0.277	118.2	0.6873	conver. cycle 26 step 13
32	16	-0.333	2.145	200.6	0.2367	conver. cycle 4 step 9
33	6	-1.884	1.962	311.6	0.2805	conver. cycle 27 step 16
34	9	0.766	0.157	278.6	0.6285	conver. cycle 30 step 13
35	1	-5.038	0.068	263.2	-0.2143	conver. cycle 24 step 4
36	8	-0.976	1.311	294.3	0.5342	conver. cycle 27 step 17
37	11	0.280	-1.973	177.5	0.1788	opt. completed-no gradient

Fig. 8. A fragment of a molname2.SUM output file from MEPPAR.

relative positions. Each position is defined choosing randomly a rotation within 0° and 360° and two translations from an interval, whose limits are minus and plus a certain value that is equal to the longest dimension of both molecules (along any axis) multiplied by a factor. This factor is specified in the configuration file and its default value is 0.5.

For each randomly chosen starting position, a gradient driven maximization procedure is performed. In order to save computational time the program detects convergence of a maximization trajectory with any of those previously considered and bypasses it. With the same purpose, the program examines if a maximization trajectory generates a translation outside the allowed interval (see previous paragraph) and, if so, neglects the current trajectory (see Fig. 6). The procedure to control convergence between trajectories is shown schematically in Fig. 7 and is the following:

– Coordinates of the second molecule are stored in certain steps along each maximization

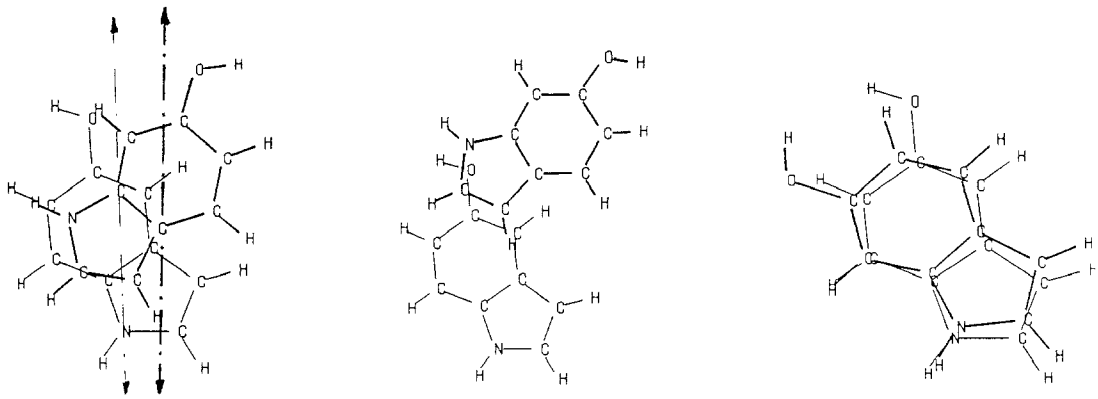

Fig. 9. Optimized relative positions with R>0.5 provided by MEPPAR when comparing 5-hydroxyindole (thin line) and 6-hydroxyindole (thick line). Electrostatic vectors are shown in the position of highest similarity.

process driven by the gradient method. Maximization trajectories are represented by arrows in Fig. 7 and the stored steps are depicted by crosses.

– Steps to be stored are determined by a distance criterion to the previous one. When a step is too close to the former, it is not stored. This distance criterion is represented in Fig. 7 by circles around the crosses. The minimum distance between stored steps is another parameter that can be customized in the configuration file.

– When a new trajectory is tested, the distance between each new step and all the steps stored from other trajectories are computed in order to detect a possible convergence. This is the event that is represented by a dashed zone in Fig 7.

The molname2.SUM output file summarizes the useful information at the end of each trajectory: number of steps, translations and rotation, Spearman similarity coefficient, and stop criterium (maximum reached, convergence with a previous trajectory or translation out of range). Figure 8 shows an example of that file. In that figure we can observe that the proportion of trajectories that supplies new positions of maximum similarity, decreases along the process. This feature can be used to decide when to stop the analysis because the probability of finding a new interesting maximum is low enough.

An example of MEPPAR analysis is the comparison of MEP distributions of 5-hydroxyindole and 6-hydroxyindole in a parallel plane placed at a distance of 1.6 Å from the aromatic rings. Comparisons have been carried out considering points of both molecules with MEP lower than −5 kcal/mol, or with distances to any nucleus lower than two-fold its van der Waals radius. MEP have been computed from STO-3G wavefunctions. Figure 9 shows the three final relative positions of both molecules that have similarity coefficients greater than 0.5. If this information is used to study the relative alignment of those aromatic systems in 5-HT receptors taking into account the model proposed by Weinstein et al. [9], we observe that the absolute similarity maximum (R=0.80) is found at a relative position which causes the overlap of the Weinstein electrostatic orientation vectors (see Fig. 9). Consequently, we are able to identify automatically the relative position of electrostatic alignment that Weinstein et al. [9] have postulated. The high

electrostatic similarity coefficient at this position does not agree with the low serotoninergic activity of 6-hydroxytryptamine because this orientation prevents the interaction of the alkylamine moiety with its complementary region of the active site. There is another electrostatic similarity maximum at a relative position close to the structural matching of the two molecules (R=0.53) that would be more relevant to explain the low activity of that compound.

In conclusion, the results of MEPPAR module permit not only the discovery of the relative position of absolute maximum similarity between two compounds, but also to detect other local maxima that would be interesting in order to explore the alignment of biomolecules in front of a common biological receptor.

People interested in using the MEPSIM software are requested to contact the authors.

ACKNOWLEDGEMENTS

We would like to thank Marta Pulido, M.D. for editorial assistance and copy editing. J.R. is the recipient of a grant from Xunta de Galicia. We also acknowledge financial support from CICYT (FAR- 1991-1072-C03-01) and from CESCA.

REFERENCES

1 Weinstein, H. and Osman, R., In Richards, W.G. (Ed.) Computer-Aided Molecular Design, IBC Technical Services, London, 1989, pp. 105–118.
2 Richards, N.G.J. and Vinter, J.G., J. Comput.-Aided Mol. Design, 5(1991)1.
3 Orozco, M., Canela, E.I. and Franco, R., Mol. Pharmacol., 35(1989)257.
4 Sanz, F., Manaut, F., José, J., Segura, J., Carbó, M. and de la Torre, R., J. Mol. Struct. (Theochem), 170(1988)171.
5 Manaut, F., Lozoya, E. and Sanz, F., In Silipo, C. and Vitoria, A. (Eds.) QSAR: Rational Approaches to the Design of Bioactive Compounds, Elsevier, Amsterdam, 1991, pp. 339–342.
6 Manaut, F., Sanz, F., José, J. and Milesi, M., J. Comput.-Aided Mol. Design, 5(1991)371.
7 Sanz, F., Manaut, F., Sánchez, J.A. and Lozoya, E., J. Mol. Struct. (Theochem), 230(1991)437.
8 Sanz, F., Manaut, F., Dot, T. and López de Briñas, E., J. Mol. Struct. (Theochem), 256(1992)287.
9 Weinstein, H., Osman, R., Green, J.P. and Topiol, S., In Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum Press, New York, 1981, pp. 309–324.