Journal of Computer-Aided Molecular Design, 8 (1994) 83–96
ESCOM

J-CAMD 235

Molecular Lipophilicity Potential, a tool in 3D QSAR: Method and applications

Patrick Gaillard[a], Pierre-Alain Carrupt[a], Bernard Testa[a,*] and Alain Boudon[b]

[a]*Institut de Chimie Thérapeutique, École de Pharmacie, Université de Lausanne, BEP, CH-1015 Lausanne, Switzerland*
[b]*Institut de Recherches Servier, 11 Rue des Moulineaux, F-92150 Suresnes, France*

Received 25 May 1993
Accepted 26 August 1993

Key words: Molecular Lipophilicity Potential; MLP; Quantitative structure–activity relationships; QSAR; Comparative Molecular Field Analysis; CoMFA; α_1-Adrenoceptor ligands

SUMMARY

A new method is presented to calculate the Molecular Lipophilicity Potential (MLP). The method is validated by showing that the MLP thus generated on the solvent-accessible surface can be used to back-calculate log P. Because the MLP is shown to be sensitive to conformational effects, the MLP/log P relation is best sought by taking all conformers into account. The MLP method presented here can be used as a third field in CoMFA studies, as illustrated with two series of α_1-adrenoceptor ligands. In the first series, the steric, electrostatic and lipophilic fields are highly intercorrelated, and taken separately yield comparable models. In the second series of ligands, the best model is obtained with the lipophilic field alone, allowing insights into ligand–receptor interactions.

INTRODUCTION

Interactions between bioactive molecules and receptors are governed by different intermolecular forces, classified as steric, electrostatic and hydrophobic. While steric or electrostatic complementarity between a receptor and its ligand can easily be modelled at the molecular level (standard 3D QSAR with CoMFA), the description of hydrophobic interactions is much more difficult to handle. These difficulties arise not only from the physical complexity of such forces but also from a nonrigorous use of vocabulary.

It should be noted that drug–receptor interactions involve fundamentally the same intermolecular forces as those acting on the partitioning of a solute between water and an immiscible organic phase [1]. Hence, it is not surprising that the molecular parameter describing partition

*To whom correspondence should be addressed.

(log P) can often be correlated with biological activity. In other words, a deeper understanding of the partition behaviour of a compound, i.e., its *lipophilicity*, must shed light on the intermolecular interactions it can elicit. A number of contributions have shown that lipophilicity is a physicochemical property which encodes two major structural contributions, namely a bulk term reflecting cavity formation, hydrophobic and dispersive forces, and a polar term reflecting more directional electrostatic interactions and hydrogen bonds [1–4]. According to our definition, the first contribution is called *hydrophobicity*, and the second *polarity*. As lipophilicity is an equilibrium property, its two contributions describe the balance between all intermolecular solute–solvent interactions in both phases. Because water is highly structured as compared to organic solvents, hydrophobicity mainly expresses water–solute interactions and can be represented by a single molecular parameter of the solute, e.g. its molecular volume or its water-accessible surface area.

In contrast, the polarity part of lipophilicity takes into account the intermolecular interactions in the two phases. Thus, the balance between these forces is mainly dependent on the nature of the organic phase and the physicochemical meaning of the polarity term changes from one organic solvent to another. Based on these considerations, we have recently shown that lipophilicity can be factorized as follows (Eq. 1):

$$\log P = aV - \Lambda \tag{1}$$

where V is the molecular volume and Λ is the polarity parameter [3,4].

The above analysis shows the complexity of lipophilicity and how misleading it is to identify this parameter with a combination of hydrophobic and electrostatic terms only. In particular, the polarity contribution to the partition of a solute between water and an organic phase encodes forces that are poorly described by an electrostatic field. Thus, we have looked upon the use of a quantitative 3D description of lipophilicity as a fruitful additional component in recent 3D QSAR methodologies such as CoMFA.

Pioneering work by Dubost [5], Fauchère [6] and Furet [7] has clearly demonstrated the ability of the Molecular Lipophilicity Potential (MLP) to describe qualitatively the 3D distribution of lipophilicity, either in space or on a molecular surface. Unfortunately, the first uses of MLP in 3D QSAR [8] were not fully convincing, principally because of high correlations between MLP and volume descriptors in the sets of investigated compounds. We now present an approach based on the concept of MLP and integrated into the SYBYL software environment [9].

MLP definition

Lipophilicity is the resultant of steric and polar intermolecular interactions. The MLP defines the influence of all lipophilic fragmental contributions of a molecule on its environment. The MLP value at a point in space is generated as the result of the intermolecular interactions between all fragments in the molecule and the solvent system, at that given point. To calculate the MLP, we therefore need a fragmental system of log P [10–12] and a distance function [6,13]. Thus, the MLP is expressed by the following general equation:

$$MLP_k = \sum_{i=1}^{N} f_i fct(d_{ik}) \tag{2}$$

where k = label of a given point in space, i = label of the fragment, N = total number of fragments in the molecule, f_i = lipophilic constant of fragment i, fct = distance function and d_{ik} = distance between fragment i and space point k.

The MLP is therefore a potential of log P, i.e., of relative affinity for two solvents. Unlike for the electrostatic potential, it is not necessary to have a probe charge to reveal the MLP in space, since all necessary information is implicit in log P.

The MLP approach was introduced by Dubost et al. [5] who used first the fragmental system of Rekker [14] and a hyperbolic distance function. Their second version was based on the fragmental system of Broto and Moreau [10] and an exponential distance function on the molecular surface. Fauchère et al. [6] used the fragmental system of Rekker [14] and/or Hansch and Leo [15] and an exponential distance function to study qualitatively the spatial distribution of lipophilicity on the molecular surface. The system of Cohen et al. [7] was based on the fragmental system of Ghose and Crippen [11] and a hyperbolic distance function, the MLP being represented on the molecular surface. Kellog et al. [8] used the fragmental system of Hansch and Leo (multiplying these values by the solvent-accessible surface of each atom) and an exponential distance function (HINT program). They applied their MLP system to 3D QSAR to identify complementarity lipophilic zones in ligand–receptor interactions. Kim et al. [16] used the GRID program to generate a so-called 'hydrophobic' potential with an H_2O probe on all lattice points surrounding the molecule. Partial Least Squares (PLS) calculations generated orthogonal latent variables which were correlated with biological data (QSAR).

In this paper, we present a new MLP approach and two applications, one meant to validate our proposed MLP by using it to recalculate log P, and the second to interpret 3D QSAR.

METHOD

The present MLP is based on the atomic lipophilic system of Broto and Moreau [10] and a modification of the distance function used by Fauchère [6], i.e., Eq. 2 with a distance function of $e^{-d/2}$:

$$MLP_k = \sum_{i=1}^{N_{at}} f_i e^{-d_{ik}/2} \tag{3}$$

The atomic lipophilic system of Broto and Moreau

Moreau et al. decomposed 1868 log P (octanol/water) values into 222 atomic contributions of log P which take into account the nature of the atom and the connected bond types. We have defined a code label for each atomic fragment for automatic attribution of a lipophilicity atomic constant, a code that is close to the one published by Dubost et al. [17].

Distance function

A distance function is needed to describe the decrease in space of the lipophilic atomic contributions (Fig. 1). At the atom core, the MLP value must be maximal and at large distances, it must approach zero.

Fauchère's distance function (e^{-d}) [6] or Audry's distance function (1/(1+d)) [13] (Fig. 1) do not appear to be adequate beyond the Solvent-Accessible Surface (SAS), because at these dis-

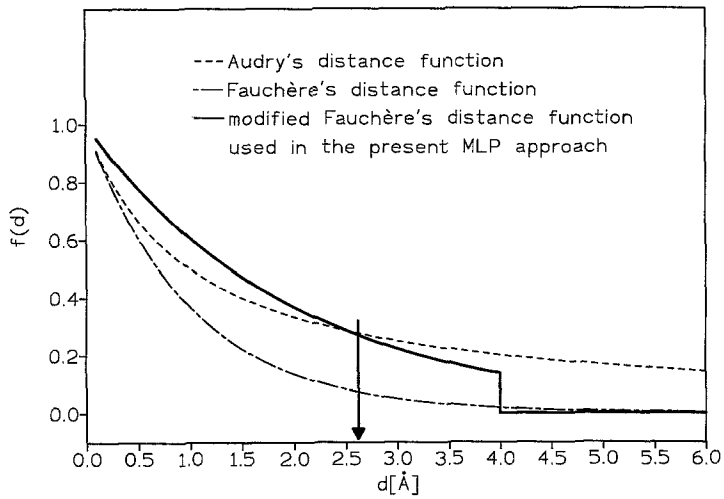

Fig. 1. Graph of three different distance functions characterising the decrease in lipophilicity in 3D space.

tances (i.e., between ca. 2.6 and 4 Å) they vary by less than 10%. A slightly modified function ($e^{-d/2}$) overcomes this limitation. In addition, we have restricted the distance function by a cutoff at 4 Å to avoid influence of too-distant fragments (Fig. 1).

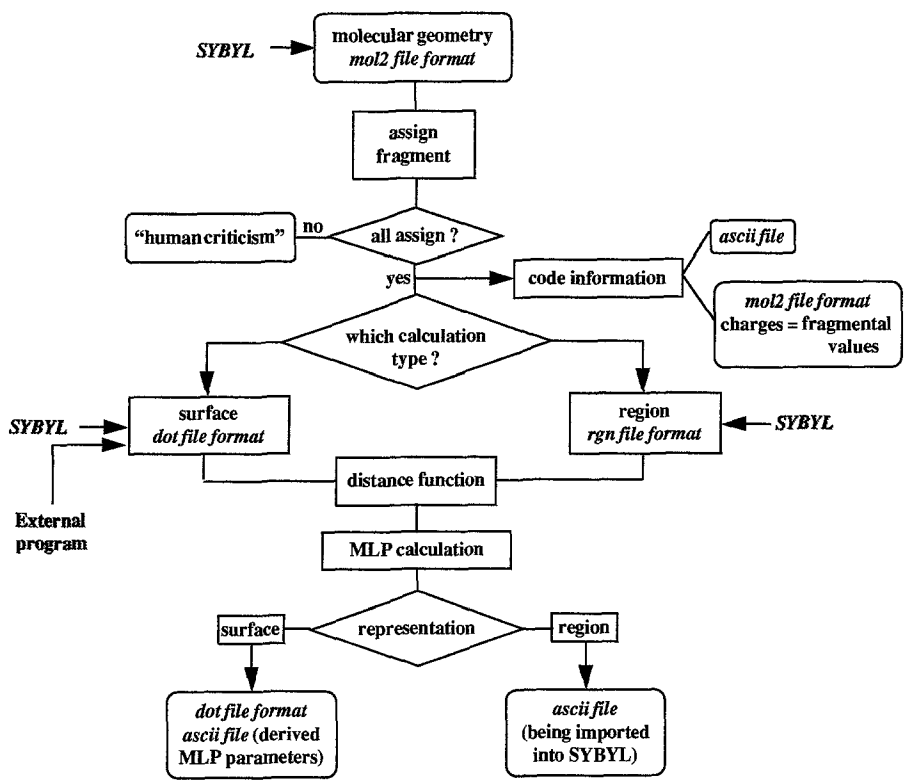

Fig. 2. Algorithm of the MLP program interfaced with the SYBYL software.

TABLE 1
MLP COLOUR SCALE

Region	Actual value	Normalised values	Colour
Polar	$-\infty$	−0.50 to −0.39	Red
		−0.39 to −0.28	Magenta
		−0.28 to −0.17	Orange
		−0.17 to −0.06	Yellow
	0.0	−0.06 to 0.06	White
Nonpolar		0.06 to 0.17	Cyan
		0.17 to 0.28	Green
		0.28 to 0.39	Green-blue
	$+\infty$	0.39 to 0.50	Blue

SYBYL interface

A FORTRAN program was written to generate the MLP and was interfaced with some common molecular modelling software, such as SYBYL. The algorithm of this program and its interface with SYBYL are shown in Fig. 2. The algorithm uses three data files, generated by the molecular modelling software, i.e., a file containing the molecule geometry, a file containing the coordinates of points on the molecular or solvent-accessible surface, and a file describing the tridimensional grid used in CoMFA. According to Eq. 3, the program can calculate the MLP either at each dot on a surface area or at each node of a 3D grid.

SYBYL Programming Language (SPL) macro commands were introduced to integrate the use of MLP into SYBYL. The dot surface is coloured with normalised values in order to obtain a relative scale of MLP, independent of the molecule (Table 1). This allows a qualitative comparison of the 'lipophilic shape' of molecules (Fig. 3).

Validation of the MLP

If spreading out lipophilicity over a given space is valid, the original log P values should be recovered by integration of the MLP. Because the solvent-accessible surface should represent how the molecule is perceived by the environment, it appears to be suited well for integrating the MLP [18,19]. However, a prediction of log P can only be made if a preliminary correlation with experimental lipophilicity is established.

For this purpose, three parameters were derived from the MLP, generated on the solvent-accessible surface of the molecule. The first parameter ($\sum$MLP) is a global parameter, obtained simply by summing all MLP values on the surface. The second and third parameters ($\sum MLP^+$ and $\sum MLP^-$) are the partial summations of the positive and negative MLP values, respectively.

The parameters $\sum MLP^+$ and $\sum MLP^-$ represent the 'lipophilic' and 'hydrophilic' parts of the molecule, respectively, i.e., the regions of the surface where positive and negative atomic values of lipophilicity are expressed. It should be noted that the MLP is dependent on the 3D structure. Thus, the log P generated by the MLP integration is sensitive to intramolecular factors, such as proximity between polar groups and, significantly, molecular conformation.

MLP in 3D QSAR

For each molecule, the MLP values calculated at each point of the 3D grid were imported into

TABLE 2
MLP-DERIVED PARAMETERS FOR AMINOPHENOL

	Exp. log P[a]	ΣMLP	ΣMLP$^+$	ΣMLP$^-$
Ortho	0.52	204	455	−251
Meta	0.16	168	382	−213
Para	0.04	167	378	−211

[a] See Ref. 21.

a SYBYL QSAR table for creating a new field. Thus, three different CoMFA columns now exist in a QSAR table: a first column with the steric field (van der Waals interactions), a second one with the electrostatic field (Coulombic interactions) [20], and a third column with the imported lipophilic field. The energy cutoff value of the first two fields was 30 kcal/mol, while the lipophilic field was taken into account only outside the van der Waals surface.

APPLICATIONS

Validation of the MLP

Aminophenols

MLP integration by summation over the solvent-accessible surface distinguishes ortho-, meta- and para-disubstituted benzenes in a quantitatively correct manner. Consider, for example, the aminophenols (Table 2). The larger lipophilicity of the ortho-isomer is predominantly due to a larger contribution of ΣMLP^+. To understand better the origin of these interactions, the ΣMLP^+ and ΣMLP^- parameters were calculated for benzene, phenol and aniline (Table 3). As expected, the addition of a polar group (hydroxy or amino) to benzene decreases the ΣMLP^+ and increases the ΣMLP^-. When para to each other, the two groups behave additively, as seen experimentally [21]. For the ortho-isomer, the increase in ΣMLP^+ and ΣMLP^- is due to the so-called OH/NH_2 'ortho effect' [22,23]. In fact, the contribution of the two groups in an ortho-orientation shows that ΣMLP^+ is decreased less than for the para-isomer (−355 instead of −432), while the ΣMLP^- is increased more (+247 instead of +209). The first effect is due to the fact that the ortho-interaction of polar groups increases the solvent-accessible surface having $MLP > 0$. The second effect is due to the fact that ortho-interactions render more negative the $MLP < 0$ points. The balance of the two opposite contributions corresponds to a global increase in lipophilicity, as observed for *ortho*- relative to *para*-aminophenol.

TABLE 3
DIFFERENCES IN MLP VALUES BETWEEN BENZENE AND SUBSTITUTED BENZENES

Compound	ΣMLP$^+$	ΣMLP$^-$
Benzene	0	0
Phenol	−169	+25
Aniline	−257	+183
p-Aminophenol	−432	+209
o-Aminophenol	−355	+247
$OH\cdots NH_2$ interaction	+77	+38

Cyclopeptides

A series of 13 cyclopeptides (11 dicyclopeptides and 2 tetracyclopeptides) was studied, three of which are shown in Fig. 3. Visual inspection of the MLP of cyclopeptides reveals large differences and allows their qualitative ordering by increasing lipophilicity. Clearly, the more lipophilic molecule (M1) has a greater lipophilic surface (coloured in blue), in contrast to the more hydrophilic molecule (M3) where the surface is very hydrophilic (coloured in red).

In a preliminary correlation study between MLP and experimental log P values, the two parameters $\sum MLP^+$ and $\sum MLP^-$ were derived for the most stable conformer of each molecule (Fig. 4A). A good correlation coefficient of 0.95 is seen for this model, where the two classes of cyclopeptides are taken together. It should be noted that the prediction of log P by the program CLOGP [21] (based on the fragmental system of Leo and Hansch) does not allow to mix the two classes of cyclopeptides (not shown). However, cyclopeptides are flexible molecules. We therefore analysed the conformational space of the 13 molecules with a molecular dynamics-based strategy to be described elsewhere [24]. The MLP integration on the solvent-accessible surface was then calculated for all conformers thus generated. The calculated log P variation for all conformers of a single molecule shows that flexible compounds are poorly described by their minimum-energy conformer. This is not surprising, since the gas-phase energy is a very poor descriptor of the conformational behaviour in solution, which may be greatly different from that in the gas phase.

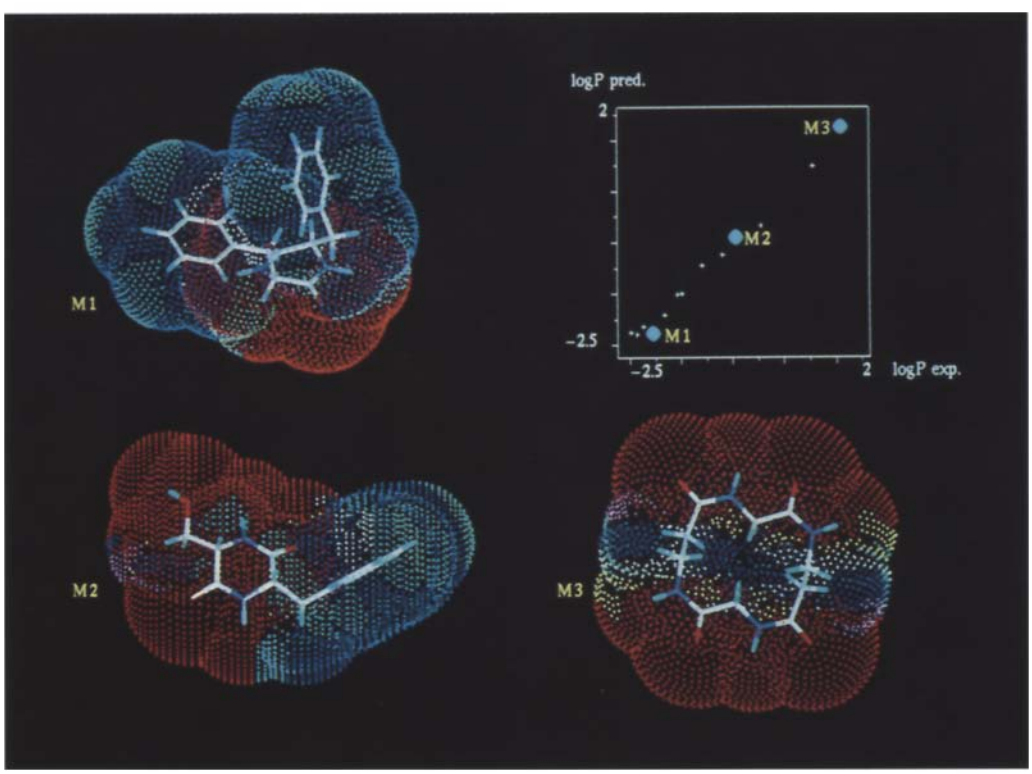

Fig. 3. Illustration of the qualitative order of lipophilicity of the MLP, generated on the solvent-accessible surface of three cyclopeptides.

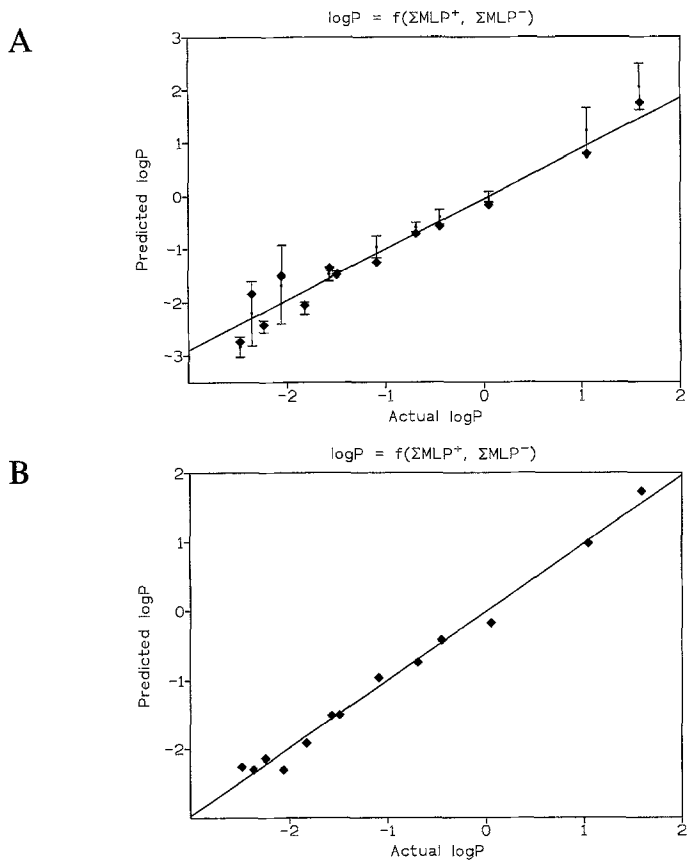

Fig. 4. (A) Linear correlation for 13 cyclopeptides between experimental and estimated log P, derived by integration of the MLP on the solvent-accessible surface of the lowest energy conformers. The range of estimated log P for the entire conformational space of each molecule is represented by vertical bars. (B) Linear correlation between experimental and estimated log P, calculated as the mean of ΣMLP^+ and ΣMLP^- values generated from the whole conformational space of each cyclopeptide.

Therefore, the MLP–log P correlation was refined by averaging for each compound the ΣMLP^+ and ΣMLP^- of all its calculated conformers [25]. The correlation thus obtained is described by Eq. 4 and shown in Fig. 4B:

$$\log P = 6.10 \times 10^{-4}(\pm 3.50 \times 10^{-4})\,\Sigma MLP^- + 3.09 \times 10^{-3}(\pm 3.70 \times 10^{-4})\,\Sigma MLP^+ - 1.61(\pm 0.68) \quad (4)$$

$n = 13,\ r^2 = 0.988,\ q^2 = 0.977,\ F = 439,\ s = 0.151$

This statistical model is a clear improvement over the first one. It is also conceptually better, because it takes conformational behaviour into account. In other words, the gas-phase preferred conformation cannot represent the complex behaviour of a solute in a biphasic system. Moreover, even if one hypothetical conformer could describe these multiple interactions with two solvents, we show here that the mean parameters, generated for all conformers of a molecule, allow a better calculation of log P.

As mentioned above, such an approach of lipophilicity prediction is restricted by the existence

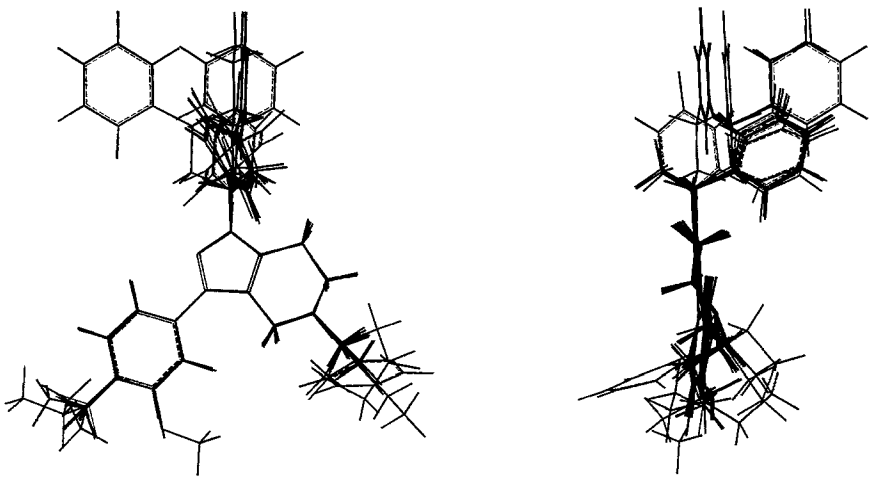

Fig. 5. Orthographic view of the alignment of the 38 derivatives of 3-aryl-4,5,6,7-tetrahydro-1*H*-pyrazolo[4,3*c*]pyridine [26].

of a correlation between the generated parameters (MLP integration) and the experimental lipophilicity measurements. In addition, such an approach will always be limited by the fragmental lipophilic system used and its approximations (for example, the implicit treatment of hydrogen atoms, partially corrected by using the atomic solvent-accessible surface).

CoMFAs with lipophilic field included

To examine the information added by the MLP field to a CoMFA analysis (version 5.5 of the SYBYL software), we analysed the affinity of two series of ligands of the α_1-adrenoceptor [26,27]. In the CoMFA, a grid spacing of 1 Å and a dielectric function of 1/r were used. The maximum steric and electrostatic field value was set to 30 kcal/mol. The charges generating the electrostatic field were calculated by the Gasteiger–Marsili method [28]. For the cross-validated analysis, the minimum sigma value was set to 2.0, with the number of cross-validated groups being equal to the number of compounds. For the final analysis, the minimum sigma value was set to 0.0 and the number of components was chosen to correspond to the first r-squared maximum found in the cross-validated analysis.

Derivatives of 3-aryl-4,5,6,7-tetrahydro-1H-pyrazolo[4,3c]pyridine

The first CoMFA was that of a series of 38 analogues, selected by Singh et al. [26]. These molecules were aligned by superimposing the polycyclic moiety and giving a common orientation to the side chains (Fig. 5). The analysis revealed that the lipophilic field (MLP) did not give more information than the steric field (Table 4 and Fig. 6), suggesting that steric interactions alone explain the variations of α_1 affinity in this series of compounds. This example demonstrates, first, the danger of statistical overfitting of the explanatory variables, and second, the danger of misinterpreting statistical results.

Indeed, the statistical results (Table 4) indicate that each field separately gives a good correlation with the α_1 affinity, suggesting intercorrelation between the three fields. Moreover, the

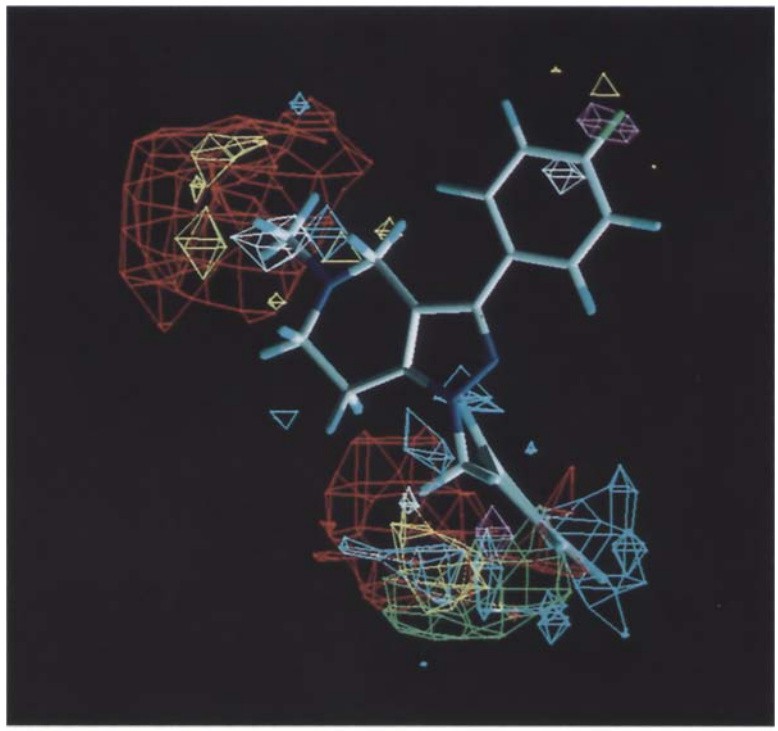

Fig. 6. Graphical representation of the major features (steric, electrostatic and lipophilic) of the 3D QSAR for α_1 affinity of the 38 derivatives of 3-aryl-4,5,6,7-tetrahydro-1*H*-pyrazolo[4,3*c*]pyridine [26]. Green (red) contours surround regions where higher (lower) steric interactions increase α_1 affinity. White (magenta) contours surround regions where higher (lower) electrostatic interactions with a positive charge increase α_1 affinity. Cyan (yellow) contours surround regions where higher (lower) lipophilic interactions increase α_1 affinity. The contour levels are 0.01 for green, white and cyan and –0.01 for red, magenta and yellow. The molecule represented is a good α_1 ligand.

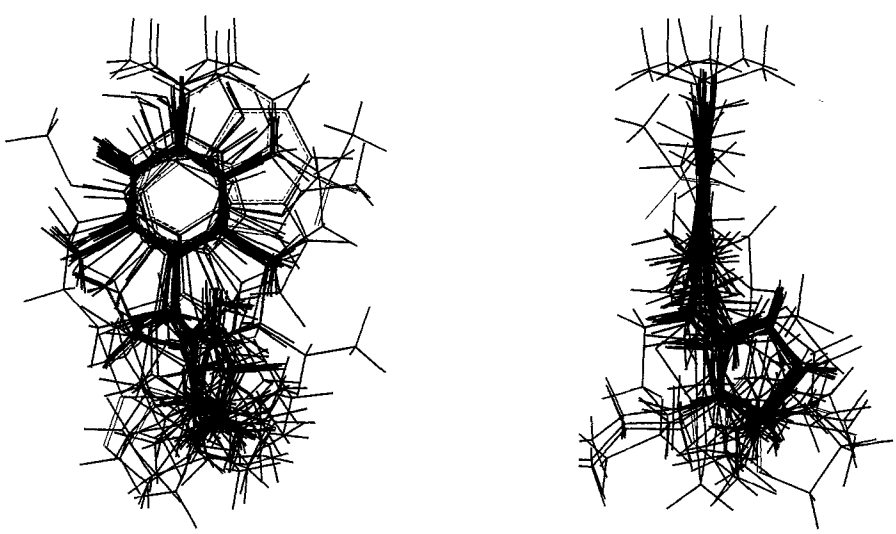

Fig. 7. Orthographic view of the alignment of the 33 miscellaneous α_1 ligands selected by Timmermans et al. [27].

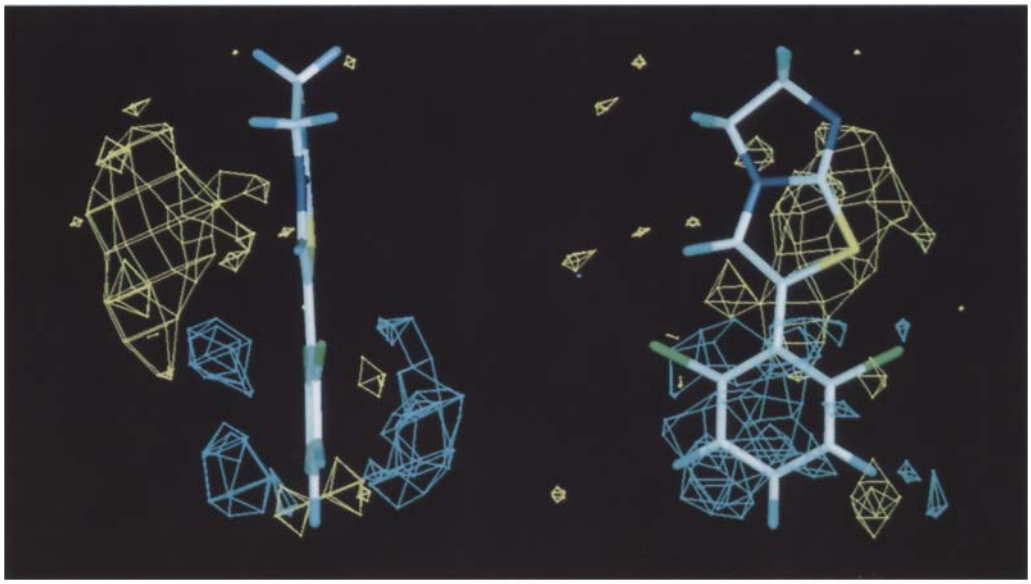

Fig. 8. Graphical representation of the model showing the major lipophilic features of the 3D QSAR for α_1 affinity. Cyan (yellow) contours surround regions where higher (lower) lipophilic interactions increase α_1 affinity. The contour levels are 0.01 for cyan and −0.01 for yellow. The molecule represented is a poor α_1 ligand.

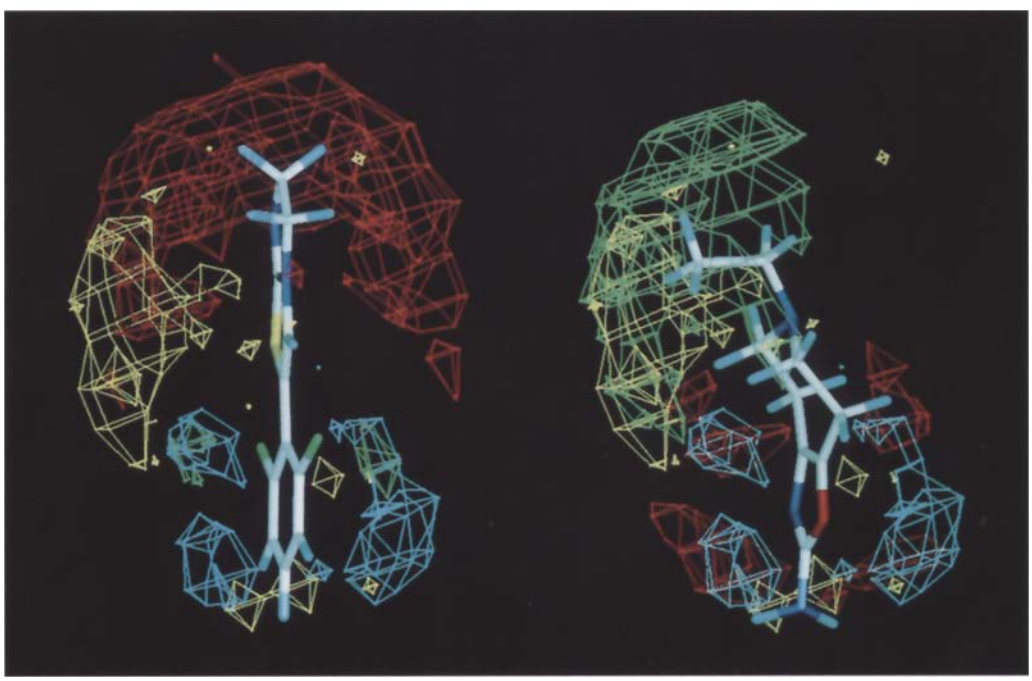

Fig. 9. The same graphical representations as in Fig. 8 for the model, showing (left) the most active molecule and (right) the least active molecule in the series. The contours in green and red show the MLP field of the molecule (green: MLP > 0, red: MLP < 0). Therefore, the best α_1 ligand of this series possesses lipophilic features in the same space regions as for the model. In contrast, the least active molecule of the series presents exactly the opposite features, i.e., lipophilic regions in the polar regions of the model.

TABLE 4
STATISTICAL RESULTS OF THE CoMFA WITH DERIVATIVES OF 3-ARYL-4,5,6,7-TETRAHYDRO-1*H*-PYRAZOLO[4,3*c*]PYRIDINE [26]

Field	q^{2} [a]	n[b]	R^{2}_{final}	Steric[c]	Electrostatic[c]	Lipophilic[c]
Steric only	0.76	5	0.92	100	–	–
Electrostatic only	0.68	4	0.86	–	100	–
Lipophilic only	0.77	7	0.94	–	–	100
Steric and electrostatic	0.73	5	0.91	72	28	–
Steric and lipophilic	0.76	5	0.92	66	–	34
Electrostatic and lipophilic	0.69 (0.74)[d]	3(8)[d]	0.85	–	43	57
Steric, electrostatic and lipophilic	0.72 (0.73)[d]	4(6)[d]	0.90	52	22	26

[a] Cross-validated r-squared measuring the predictability of the model (if $q^{2}<0.4$, the model is considered as non-significant).
[b] Optimal number of PLS components chosen for final analysis.
[c] Relative contributions of each field in the QSAR.
[d] Values in parentheses are defined by SYBYL as 'optimal values'.

models with two fields show an optimal number of components (values in parentheses) that is higher than the chosen number of components. In fact, in the model with the three fields, the first cross-validated r-squared maximum (corresponding to the first local minimum of s, the standard deviation) is found after four components, while the global maximum of the cross-validated r-squared maximum is found after six components.

The CoMFA method in SYBYL defines the optimal number of components by the absolute maximum of the cross-validated r-squared. Here, however, there is danger of overfitting beyond four components, meaning that the additional component contains no pertinent information which would increase the quality of the model. Second, the three fields together give a good model (the predictive quality of a model is derived from the cross-validated r-squared, which is here 0.72), but graphic analysis of this result (Fig. 6) reveals that the regions of space where lipophilic interactions increase affinity are not different from those where a high steric field has the same effect. Therefore, the information described by the lipophilic and steric fields are the same or are represented by the same regions of space. It must be noted that the same is true for the electrostatic field. In fact, the fragments themselves, and not their structural information, are sufficient to explain the variation in affinity of this series of ligands.

Singh et al. [26] proposed a QSAR equation correlating the α_1 affinity with three types of parameters: (a) the van der Waals volume of substituents; (b) the sum of electronic effects (σ Hammet constant) for all aromatic substituents; and (c) the hydrophobic constant π_4 (Hansch et al.) of the para-phenyl substituent. Careful examination of the data set and of the QSAR equation shows that the contributions of σ and π_4 are low and that steric parameters alone explain most of the variance.

Heterogeneous α_1-adrenoceptor agonists

The second CoMFA study was dedicated to a highly heterogeneous series of 33 compounds as selected by Timmermans et al. [27]. These compounds were aligned by superimposing the aromatic ring and the basic nitrogen, the rest of the molecule being geometrically fitted to clonidine (Fig. 7). The CoMFA model obtained revealed large differences between the steric, electro-

TABLE 5
STATISTICAL RESULTS OF THE CoMFA WITH MISCELLANEOUS α_1 LIGANDS [27]

Field	q^{2} [a]	n[b]	R^2_{final}	Steric[c]	Electrostatic[c]	Lipophilic[c]
Steric only	0.23	6	–	–	–	–
Electrostatic only	0.07	3	–	–	–	–
Lipophilic only	0.55	2	0.84	–	–	100
Steric and electrostatic	0.15	4	–	–	–	–
Steric and lipophilic	0.58	7	1.00	34	–	66
Electrostatic and lipophilic	0.55 (0.57)[d]	3(8)[d]	0.93	–	42	58
Steric, electrostatic and lipophilic	0.53 (0.56)[d]	3(8)[d]	0.95	20	33	47

[a] Cross-validated r-squared measuring the predictability of the model (if $q^2 < 0.4$, the model is considered as non-significant).
[b] Optimal number of PLS components chosen for final analysis.
[c] Relative contributions of each field in the QSAR.
[d] Values in parentheses are defined by SYBYL as 'optimal values'.

static and lipophilic fields (Table 5).

It first appears that there is a good correlation between the α_1 affinity and the lipophilic field alone, in contrast to the steric and electrostatic fields. The graphical results (Figs. 8 and 9) show that the information encoded in the MLP is related to a large area near the aromatic ring of clonidine, where an increase in lipophilicity enhances α_1 affinity (cyan contours), and to a large area near the imidazole ring of clonidine, where a decrease in lipophilicity enhances α_1 affinity (yellow contours). Figure 9 illustrates the lipophilic fields of the most and least active compounds, respectively, superimposed on the lipophilic regions of the model. In these figures, the green and red regions represent the MLP field of the compound, i.e., the lipophilic and hydrophilic regions of the ligands as generated by the MLP program.

The left side of Fig. 9 represents the best ligand in the series, with green and cyan regions well superimposed. This indicates that the lipophilic characteristics of the compound fit well with the lipophilic regions of the model. This is also true for the red regions of the compound superimposed on the yellow region of the model. In contrast, the right side of Fig. 9 represents the poorest ligand in the series, with the green regions (lipophilic characteristics of the compound) being superimposed on the yellow regions (hydrophilic regions generated from the model).

We thus conclude that near the aromatic ring, an increase in lipophilicity enhances receptor affinity, while near the imidazole ring of clonidine a decrease in lipophilicity enhances affinity.

CONCLUSIONS

The MLP presented here is able to predict quantitatively the 3D lipophilicity variation as a function of connectivity and conformation. MLP can also add significant information to 3D QSAR and CoMFA, emphasising the composite nature of lipophilicity and assessing the relative contribution of its components to receptor affinity.

Because the lipophilicity field encodes the steric and electrostatic fields, the set of compounds under study dictates which of the contributions, i.e., steric, electrostatic and/or lipophilic, is correlated with activity. However, care must be taken to avoid 'overfitting' by using cross-correlated properties, and not to misinterpret the statistical results.

NOTE ADDED IN PROOF

The necessity to have a more sensitive distance function beyond the van der Waals radius was also recognized by Brickmann and co-workers [29].

ACKNOWLEDGEMENTS

B.T. and P.A.C. are indebted to the Swiss National Science Foundation for support.

REFERENCES

1 Van de Waterbeemd, H. and Testa, B., In Testa, B. (Ed.) Advances in Drug Research, Vol. 16, Academic Press, London, 1987, pp. 87–227.
2 Testa, B. and Seiler, P., Arzneim.-Forsch. Drug Res., 31 (1981) 1053.
3 El Tayar, N., Testa, B. and Carrupt, P.-A., J. Phys. Chem., 96 (1992) 1455.
4 Tsai, R.-S., Testa, B., El Tayar, N. and Carrupt, P.-A., J. Chem. Soc., Perkin Trans. II, (1991) 1797.
5 Audry, E., Dallet, P., Langlois, M.H., Colleter, J.C. and Dubost, J.P., Prog. Clin. Biol. Res., 291 (1989) 63.
6 Fauchère, J.L., Quarendon, P. and Kaetterer, L., J. Mol. Graphics, 6 (1988) 203.
7 Furet, P., Sele, A. and Cohen, N.C., J. Mol. Graphics, 6 (1988) 182.
8 Kellogg, G.E., Semus, S.F. and Abraham, D.J., J. Comput.-Aided Mol. Design, 5 (1991) 545.
9 SYBYL 5.41, 5.55, 6.0, Tripos Associates, St. Louis, MO, 1993.
10 Broto, P., Moreau, G. and Vandycke, C., Eur. J. Med. Chem., 19 (1984) 61.
11 Ghose, A.K. and Crippen, G.M., J. Comput. Chem., 7 (1986) 565.
12 Kantola, A., Villar, H.O. and Loew, G.H., J. Comput. Chem., 12 (1991) 681.
13 Audry, E., Dubost, J.P., Colleter, J.C. and Dallet, P., Eur. J. Med. Chem., 21 (1986) 71.
14 Rekker, R.F. and De Kort, H.M., Eur. J. Med. Chem., 14 (1979) 479.
15 Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979, p. 352.
16 Kim, K.H., Quant. Struct.–Act. Relatsh., 11 (1992) 309.
17 Croizet, F., Dubost, J.P., Langlois, M.H. and Audry, E., Quant. Struct.–Act. Relatsh., 10 (1991) 211.
18 Hirono, S., Liu, Q. and Moriguchi, I., Chem. Pharm. Bull., 39 (1991) 3106.
19 Camilleri, P., Watts, S.A. and Boraston, J.A., J. Chem. Soc., Perkin Trans. II, (1988) 1699.
20 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
21 MEDCHEM Software 3.54, Daylight Chemical Information System, Inc., Irvine, CA, 1989.
22 Fujita, T. and Nishioka, T., Prog. Phys. Org. Chem., 12 (1976) 49.
23 Tsantili-Kakoulidou, A., El Tayar, N., Van de Waterbeemd, H. and Testa, B., J. Chromatogr., 389 (1987) 33.
24 Gaillard, P., Carrupt, P.-A., Testa, B. and Tsai, R.-S., manuscript in preparation.
25 Hudson, B.D., George, A.R., Ford, M.G. and Livingstone, D.J., J. Comput.-Aided Mol. Design, 5 (1992) 191.
26 Singh, P. and Sharma, R.C., Quant. Struct.–Act. Relatsh., 9 (1990) 29.
27 Timmermans, P.B.M.W.M., De Jonge, A., Thoolen, M.J.M.C., Wilffert, B., Batink, H. and Van Zwieten, P.A., J. Med. Chem., 27 (1984) 495.
28 Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
29 Heiden, W., Moeckel, G. and Brickmann, J., J. Comput.-Aided Mol. Design, 7 (1993) 503.