match no. 13 pp. 357-368 1982

STERIC FIT IN QSAR. THE MTD AND SIBIS-METHODS, CRITIQUE AND PERSPECTIVES

Z.SIMON
Institute of Medicine,Pţa 23 August 2,Timişoara 1900,Romania

(Received: February 26, 1982)

Summary. The MTD and SIBIS methods are compared, with emphasis upon their limitations. Both methods use very similar procedures to describe molecular stereochemistry, but different steric fit parameters and different optimization procedures are used. The main difficulty in extending those methods to molecules with high flexibility, as large as oligopeptides, is the description of molecular stereochemistry - the construction of the hypermolecule (the investigated receptor space).

1. Introduction

Methods to account for steric fit in quantitative structure-activity relations were developed in the last years[1], starting from the minimal steric difference (MSD)-concept for steric misfit[2]. The MTD-method[1,3-6] allows to establish QSAR's for series of usual drug molecules with widely different stereochemistry. A large number of parameters are used in the MTD-method to describe molecular stereochemistry, but the correlation of calculated and experimental biological activities for test series of molecules (not used in obtaining the regressional equation) indicates the MTD results as significant[1]. Thus, a linear correlation coefficient r=0.747 was obtained for a test series of N=25 estrogenic steroids[7], r=0.816 for N=16 cardiotoxic aglycones[6] and r=0.846 for 24 α-aryloxypropionic acids with auxinic activity[8]. The SIBIS-method[9-15] uses a more elaborate optimization procedure and steric fit parameters and can be applied to the same type

of molecules as the MTD-method.

At present, it is not possible to extend these methods to flexible molecules of the size of oligopeptides or larger. Considerable difficulties appeared in MTD-studies[8,16] even for a series of 75 α -aryl- and α-aryloxy-propionic acids with auxinic activity. Therefore we will discuss here the MTD and SIBIS-methods with special emphazis upon their limitations and the approximations of the procedures used in describing molecular stereochemistry and steric fit.

2. Description of molecular stereochemistry

The description of molecular stereochemistry is based, both in the MTD and the SIBIS-methods, upon a connected graph - the hypermolecule ($\hat{H}$) or, as very suggestively called within the SIBIS-method, the investigated receptor space (IRS). This graph is a model for tridimensional molecular stereochemistry.

The $\hat{H}$ or IRS-graph is obtained by the approximative atom per atom superposition procedure (H-atoms neglected) of the first MSD-paper[2]. The vertices of the graph correspond to approximative atomic positions, the edges to possible bonds between atoms. The superposition of the N molecules "i" of the studied series yields M vertices "j" and the occupation or nonoccupation of vertex "j" in molecule "i" is characterised by the x_{ij}-parameter. The stereochemistry of molecule "i" is described by the $x_{i1}, x_{i2}, \ldots, x_{iM}$-vector.

The superposition procedure involves several steps at which difficulties may appear :

i) The relative orientation of the molecules must be indicated before they can be superimposed; some groups, occupying always the same positions within the receptor, must be found. This is easy if a few polar groups or a large part of the molecular core are common to the active molecules. If not, a preliminary pattern recognition approach will be required to find out a common constellation of atoms - the "pharmacophore"[9,11].

ii) Molecules with chains of σ-bonds have several low energy conformations, which are isoenergetic within a few kT-units and the conformation fitting best the receptor will be adopted. The best fitting conformation may be guessed as the one superimposing best upon a highly active rigid molecule. This rule may be not sufficient if flexible molecules prevail; several conformations for each molecule will have to be considered, increasing the number M of vertices in the investigated receptor space[1].

iii) Low (isoenergetic) energy conformations are usually considered as those in which intramolecular π-electronic interactions are not interrupted and in which two vertices occupied by nonbonded atoms are separated by at least one unoccupied vertex. The planar conformation of FIG.1, in which the CH_3-group occupies vertex j=1, the CH-group, j=5, and j=6 is empty, should be allowed according to this rule, but the H...C (of CH_3) distance is 1.45 Å, while the sum of van der Waals distances is 2.7 Å; considerable steric strain must therefore appear.

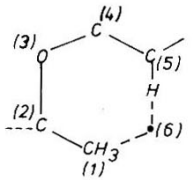

FIGURE 1

iv) If cycles of different sizes (especially condensed cycles) are to be superimposed, for example a cyclopentyl moiety upon a phenyl or cycloheptyl moiety, the superposition of certain atoms (their attribution to one j-vertex or another) becomes ambiguous. The same is true for superposition of atomic triplets with large differences in bond lengths and angles (e.g. the CCC constellations of a cyclopropyl and of a cyclopentyl group).

These difficulties appear especially if one tries to establish QSAR's for oligopeptides differing by several aminoacid residues. The already existing three dimensional molecular modeling techniques, like the one devised by Gund,[17] could offer a solution for a computerized hypermolecule construction in such cases. Once a "pharmacophoric" constellation has been found, several conformations can be generated

and compared by such techniques. Conformational energy calculations would allow a more rigorous selection of low energy conformations. Algorithms can be devised within such techniques in order to contract nearby atoms in unique vertices. Separate structural information concerning the molecules or the receptor can be considered in such techniques.

3. The optimization procedure

The M vertices of the hypermolecule are the M structural variables which describe the stereochemistry of the receptor; more exactly of the receptor space "investigated" by the studied series of molecules. In all MTD and SIBIS studies published up to now, the number N of molecules in the studied series is less than 3 times the number M of vertices, as required by Topliss and Edwards,[18] but it is difficult to see how the stereochemistry of receptor-effector interactions could be described by a lower number of structural variables. Actually, the success of methods such as MTD or SIBIS is due to the fairly small number of structural variables in the hypermolecules.

The optimization procedures of MTD and SIBIS assign the vertices "j" to the receptor (r)-region, i.e. cavity (c)- or wall (w) regions, or to irrelevant (i)-regions, such as to minimize the sum :

$$Y = \sum_{i=1}^{N} (A_i - \hat{A}_i)^2 \qquad (1)$$

of the quadratic differences between experimental and calculated biological activity.

For the MTD method, the x_{ij}'s which characterize occupation or nonoccupation of vertex j in molecule i are given by (for the sake of simplicity) :

$$x_{ij} = \begin{cases} = 1 & \text{if j is occupied in molecule i} \\ = 0 & \text{if j is empty in molecule i} \end{cases} \qquad (2)$$

and the steric misfit is given by the minimal steric difference, MTD :

$$MTD_i = s + \sum_{j=1}^{M} \varepsilon_j x_{ij} \qquad (3)$$

with $\varepsilon_j = -1$, 0 or +1 for c-, i- and w-type vertices respectively and s is the number of c-vertices. The regressional equation is of the type :

$$\hat{A} = \alpha - \quad MTD \qquad (4)$$

The beneficial and detrimental contributions of atoms in c- and w-vertices, respectively, are considered of equal magnitude, which is certainly a crude approximation. The optimization is started with the "start map" S_o, a collection of ε_1^o, ε_2^o,..., ε_M^o -values, corresponding to a reasonable, probable, shape of the receptor, as indicated by structures of active and inactive molecules.[6] The result is the "optimized receptor map", S^*, the optimal ε_1^*, ε_2^*,..., ε_M^* -assignements of the M vertices to the c-, i- and w-regions. The reliability of the ε_j-assignements is assessed by calculating the increase $\Delta Y(\varepsilon_j)$ of Y produced by each, separate ε_j^* by ε_j substitution in the optimized map, S^*.[6] If several conformations are possible for a given molecule, the one yielding the lowest MTD-value must be considered. Other structural variables can be included into α . The optimization procedure can be performed also by hand if only MTD is considered as structural variable and the series of molecules is not too large.

The prediction of the most active possible molecule is the principal aim of the SIBIS method and k_p-Austel[19] steric parameters are used to characterize atoms or groups of different size in the j-vertex :

$$x_{ij} = \begin{cases} = k_p & \text{if j is occupied by a "P"-atom in i} \\ = 0 & \text{if j is empty in i} \end{cases}$$

The SD_c^* and SD_w^* - (steric difference)-parameters are calculated for the atoms of molecule "i" in c- and in w-regions, respectively :

$$SD^{*}_{ci} = \sum_{\text{all j in c}} x_{ij} \quad ; \quad SD_{wi} = \sum_{\text{all j in w}} x_{ij} \qquad (6)$$

In the informational SIBIS-version[10], the x_{ij}-values are multiplied by uncertainly factors related to an informational content. The regressional equation is of the type :

$$\hat{A} = a + b\ SD^{*}_{c} + c\ SD^{*}_{w} \qquad (7)$$

The optimization procedure starts from a derived space, $\langle IRS \rangle_{init}$, with the vertices of the most active molecule considered as c-vertices, and the other ones as w-vertices. An optimal $\langle IRS \rangle_{opt}$ is obtained by changes in vertex assignments, but these are limited by connectivity conditions[11]: all c-type and i-type vertices must give, each separately, a single connected graph. A redundant i-type vertex is introduced at the beginning of the optimization procedure and is connected with those vertices for which one wishes to check the steric relevance. In $\langle IRS \rangle_{opt}$, the c-graph gives the shape of the most active predicted molecule.

The optimized map, S^{*}, of the MTD-method depends rather strongly upon the start map, S^{o}; local, nonabsolute Y - minima may result in the optimisation procedure[1,5]. Even for strong "perturbations" (interchanges) of the experimental A_i-values, the correlation coefficient r decreases only moderately[20]. A reliability of less than 95% results for most vertex assignments[6,8,21]. Nevertheless, MTD-papers on different series of compounds (estrogenic steroids and stilbestrol derivatives[7,21,22]) yield identical assignments for most common vertices. Different assignments result for vertices occupied by hydrophobic groups in one series and by hydrophylic groups in the other series.

The optimized map, $\langle IRS \rangle_{opt}$ of the SIBIS-method is almost independent of the start-vertex attributions (the $\langle IRS \rangle_{init}$). In a SIBIS-study on 13 carboxypeptidase inhibitors (carboxylic acids)[11] only 2 our of the 19 vertices of $\langle IRS \rangle_{opt}$ change assignments for 4 different $\langle IRS \rangle_{init}$.

The stability to perturbations of A_i-values is also improved (unperturbed r=0.814, strong perturbation r=0.554, in the cited study[11]). Reliability tests in SIBIS-studies seen to have been performed only upon r and the a,b and c-coefficients of the regressional equations (7). Use of the informational correction for the x_{ij}'s, yields - for the three cited examples, almost no increase of r, a strong increased t-Student test value for the b-coefficient, but almost no t-increase for the c-coefficient.[10]

The connectivity conditions of the SIBIS-optimization procedure are responsible for the improved stability of this method.[11] Such conditions could be used to account for independent structural information concerning the receptor - if available. The connectivity conditions formulated in ref.[11] can force the $\langle IRS \rangle_{opt}$ - to describe the receptor as a single cavity with a single entrance. This seems rather restrictive for the high steric complexity possible for biological receptors. As example, the receptor responsible for cardiotoxicity of aglycones, as indicated by the optimised map of a MTD-study[6], should be a cleft with a wide lateral opening, while a SIBIS-study gives a quite different $\langle IRS \rangle_{opt}$ (I.Moţoc, personal communication). The check upon a test series of 16 cardiotoxic aglycones, not used in the optimization procedures, is satisfactory only for the results of the MTD study.

4. Nature of steric fit parameters

Binding of effector molecules to biological receptors implies both attractive forces, largely due to hydrophobic bonding, and repulsive forces, due to deformation (conformational changes) of the receptor cavity by steric misfit. Matching or mismatching of electrical charges or hydrogen bonding groups on the receptor and effector may also intervene, as well as other types of intermolecular forces. Steric misfit will be irrelevant for cavity regions with high flexibility. It is thus difficult to separate a pure steric

fit (misfit) - effect.

The separate consideration of cavity and wall vertices in the SIBIS-method is certainly a significant improvement over the MTD-method, but assignments of vertices to the c-, w-, or i-regions may be due, accidentally, to matching or mismatching of different intermolecular force types also in the SIBIS-method.

A variant of SIBIS, the HIBIS-method accounts specially for hydrophobic bonding.[12] The atoms situated at different vertices are characterised by Rekker's[23] fragmental constants (let us term them f_{ij}, in analogy with x_{ij}). The vertices are now assigned to only two regions: receptor-(r), including both c- and w-regions, and irrelevant ones -(i).

The regressional equation of HIBIS is :

$$\hat{A} = a + bf + cf^2 \qquad (8)$$

$$f_i = \sum_{\text{all } j \text{ in } r} f_{ij} \qquad (9)$$

The optimization procedure is the same as for SIBIS and an optimized derived space, $\langle IRS \rangle_H$ (to discern from $\langle IRS \rangle_S$ of SIBIS) results.

An attempt to combine the SIBIS and HIBIS method, for a series of 20 dihydropholate reductase inhibiting triazine derivatives was described,[13] in which SD_c^*, SD_w^* and f are used as structural parameters, but the $\langle IRS \rangle_H$ and $\langle IRS \rangle_S$, obtained by separate optimization procedures, are not identical.

It would be interesting to characterise attractive efector-receptor forces by f, repulsive forces by SD_w according to a regressional equation :

$$A = a + bf + cf^2 + dSD_w^* \qquad (10)$$

with f defined according to eq.(9), SD_w^*-according to (5) and (6). A unique assignment of vertices to the c-, w- and i-regions (r=c+w) should pe performed by a SIBIS-type optimization procedure.

5. Conclusions

The empirical procedure[1,2] to describe molecular stereochemistry, used in the MTD and SIBIS studies published till now, can be used to molecules up to the size of steroidal hormones, di- or tri-peptides. The simplicity of the MTD-method recommends it especially for preliminary trial-and-error studies, to search a crude receptor model and to discern pertinent intermolecular force-features. The receptor map (the derived optimized $\langle$ IRS $\rangle$) and the regressional equation can be obtained by the SIBIS-method, or a combined SIBIS-HIBIS method (in the sense of eq.(10) of the previous paragraph), but one must avoid excessively restrictive vertex-connectivity conditions. The ability to account for more than one conformation for each molecule should be included into the optimization procedure.

For oligopeptides or macromolecules as effectors, computer algorithms for constructing the hypermolecule (the IRS) are to be created. The number M of vertices, the stereochemical structure variables, will dramatically surpass, hereby, the number N of studied molecules. Several receptor models able to "explain" the biological activities will result, and separate structural information concerning the receptor and effectors would be required to discern between them.

Acknowledgements are due to dr.I.Moţoc from the Chemical Research Centre of Timişoara for communication of results concerning a SIBIS-study of cardiotoxic aglycones.

Literature

1. A.T.Balaban, A.Chiriac, I.Moţoc, Z.Simon: "Steric Fit in Quantitative Structure Activity Relations", Lecture Notes in Chemistry, vol.15, Springer, Berlin 1980, chapters 4,5
2. Z.Simon, Z.Szabadai: Studia Biophys. (Berlin) 39, 123(1973)
3. Z.Simon, A.Chiriac, I.Moţoc, S.Holban, D.Ciubotariu, Z.Szabadai, Studia Biophys. (Berlin) 55, 217(1976)
4. Z.Simon, I.I.Bădilescu, T.Racoviţan: MATCH 3, 257(1977)
5. Z.Simon, I.I.Bădilescu, T.Racoviţan: J.Theor.Biol. 66, 485 (1977)
6. Z.Simon.N.Dragomir, M.G.Plauchithiu, S.Holban, H.Glatt, F.Kerek: Eur.J.Med.Chem. 15, 521(1980)
7. V.Popoviciu, S.Holban, I.I.Bădilescu, Z.Simon: Studia Biophys. (Berlin) 69, 75(1978)
8. A.Chiriac, V.Chiriac, I.Moţoc, S.Holban, Z.Simon: Rev.Roum. Chim. (in press)
9. I.Moţoc: Preprint Univ.Timişoara, Serie Chimie 5/1980
10. I.Moţoc: MATCH 11, 169(1981)
11. I.Moţoc, O.Dragomir-Filimonescu: MATCH 12, 117(1981)
12. I.Moţoc, O.Dragomir-Filimonescu: MATCH 12, 127(1981)
13. O.Dragomir-Filimonescu, I.Moţoc, I.Muscutariu: MATCH 12, 135(1981)
14. I.Moţoc. I.Muscutariu, D.Ciubotariu: Kexue Tongbac (Beijing) 26, 361(1981)
15. I.Moţoc: Arzneim.Forsch, 31, 290(1981)
16. A.Chiriac, V.Chiriac, Z.Simon: Preprint Univ.Timişoara, Serie Chimie 4/1981
17. P.Gund, J.Andose, J.B.Rhodes, G.M.Smith: Science 208, 1425(1980)
18. J.G.Topliss, R.P.Edwards: J.Med.Chem. 22, 1238(1979)
19. V.Austel, E.Kulter, W.Kalbfleisch: Arzneim.Forsch. 29, 585(1979)
20. I.Moţoc: MATCH 5, 275(1979)
21. Z.Simon, E.Caspi: submitted to J.Theor.Biol.
22. I.Niculescu-Duvăz, I.Elian, T.Crăescu: Oncologia (Bucharest) 19, 211(1980)
23. R.F.Rekker: "The Hydrophobic Fragmental Constant", Elsevier, Amsterdam, 1977

Appendix. Test of MTD vs. SIBIS results for cardiotoxic aglycones.

The MTD-method was applied for a series of thirty cardiotoxic aglycones in a previous paper.[6] The hypermolecule resulted by superposition of these thirty molecules with the corresponding vertex numerotation is depicted in FIG.2 and the corresponding optimised receptor map is characterised by following vertex-assignments :

$$S^{*}\begin{cases} c\ ;\ j(\xi_j=-1) : 1,12,14 \\ i\ ;\ j(\xi_j=\ 0) : 2,4\text{-}11,13,15\text{-}20,22\text{-}24 \\ w\ ;\ j(\xi_j=+1) : 3,21 \end{cases}$$

Vertices j=25-45, occupied in all the 30 molecules, can have any assignment in the MTD-method. The corresponding regressional equation is :

$$A = 0.437-1.71MTD-0.135HF \ ;\ r=0.916,\ s=0.34 \qquad (11)$$

HF is the hydrophobicity relatively to dihydroxylated cardenolides, calculated according to Rekker.[23]

In a SIBIS-study on the same series, dr.I.Moţoc (J.Org. Chem., submitted) obtained the following optimal IRS and regressional equation :

$$\langle IRS \rangle_{opt}\begin{cases} c\ ;\ j(\xi_j=-1) : 1,12,15,25\text{-}45 \\ i\ ;\ j(\xi_j=\ 0) : 2,4,8\text{-}11 \\ w\ ;\ j(\xi_j=+1) : 3,5\text{-}7,13,14,16\text{-}24 \end{cases}$$

$$A=18.795+0.862SD_c-0.426SD_w-0.428HF \ ;\ r=0.82,\ s=0.48 \qquad (12)$$

In the optimization procedure of this SIBIS study, only vertices j=2,3,8-10,16 were connected to the redundant, i type vertex, j=46. The Austel steric parameters are $k_p=1$ for all atoms and groups involved.

The test series, used also in the previous paper (ref.[6]), is listed in TAB.1. The occupied vertices j ($x_{ij}=1$),HF,MTD, SD_c^* and SD_w^* - values for the test - molecules are also listed, as well as the experimental activities,A,and those calculated by the MTD-method, $\hat{A}(11)$ and by the SIBIS-method, $\hat{A}(12)$. For the experimental vs. calculated activity correlation, r=0.81 is obtained with the MTD-method, but only r = = 0.10 with the SIBIS method.

TABLE 1 Test series

Name	$j(x_{ij}=1)$	HF	MTD	SD_w	SD_c	A	A(11)	A(12)
Hellebrigenin	1,2,4-7,17	-2.24	0	5	22	+0.73	+0.71	+0.37
12- -OH--Bufalin	1,2,4-6,12,14,20	-0.58	0	4	23	+0.68	+0.51	+0.95
Gamabufotalin	1,2,4-6,12,14.19	-1,44	0	4	23	+0.60	+0.63	+1,32
Scillarenin	1,2,4,5,12,14	+0.05	0	2	23	+0.45	+0.43	+1.54
Strophantidin	2,4-7,12,14,17	-2.47	1	5	22	+0.11	-0.94	+0.48
Sarmentogenin	2,4-7,12,14,17,19	-4.14	1	6	22	-0.07	-0.71	+0.76
Digitoxigenin	2,4-6,12,14	0.00	1	3	22	-0.06	-1.27	+0.27
Bufotalinin	1,2,4-6,12,17	-1,41	1	3	24	-0.17	-1.08	+2.70
Glaucorigenin	2,4-6,12,14,17,22	-3.51	1	5	22	-0.23	-0.80	+0.88
Strophantidol	2,4-7,12,14,17	-2.19	1	5	22	-0.24	-0.98	+0.36
3-epi--Bufalin	1,3-6,12,14,20	-0.87	1	5	23	-0.40	-1.16	+0.65
Uzarigenin	7-10,12,14	0.00	1	3	22	-0.49	-1.27	+0.27
Pachygenol	7-10,12,14,17	-1.85	1	4	22	-2.00	-1.02	+0.64
3-epi-Digitoxigenin	3-6,12,14	0.00	2	4	22	-2.00	-2.98	-0.16
Resibufogenin	1,2,4-6,12,15	-0.03	1	2	24	-2.00	-1.27	+2.43
3-epi-Periplogenin	3-7,12,14	-1.67	2	5	22	-2.00	-2.76	+0.13

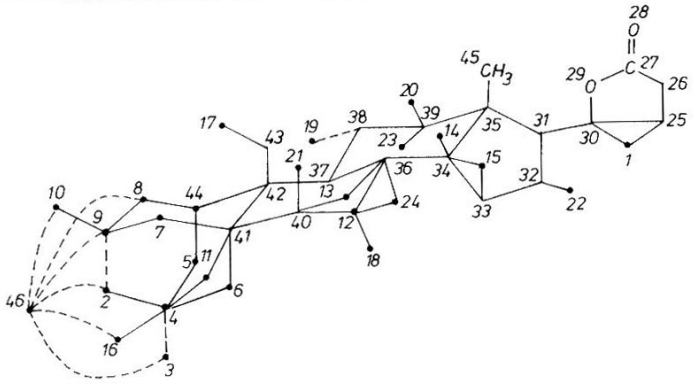

FIGURE 2