match no. 3 pp. 257 - 262 1977

MINIMAL TOPOLOGICAL DIFFERENCES, MAPPING OF SITES OF BIOLOGICAL RECEPTORS AND GRAPHS

(received: September 1977)

Zeno Simon,[a] Ilie I. Bădilescu[b] and Traian Racoviţan[c]

[a] University of Timişoara, Bvd. V. Pârvan 4, Timişoara, Romania,
[b] Chemistry Department, Pedagogical Institute, Piteşti, Romania,
[c] Computing Centre of the Petrochemical Works, Piteşti, Romania.

ABSTRACT. A procedure of determining the best probable shape of effector molecules for a given biological receptor site is described. The procedure is based on minimal topological differences, an approximate measure for steric fit. It uses the experimentally determined biological activities of a series of **N** molecules with known structures. This procedure is an optimisation problem for a tricoloured graph.

In a previous paper,[1] the minimal topological difference,[+)] MTD, was introduced, as an approximate measure of the nonoverlapping volumes when a molecule (in a low energy conformation allowing maximal superposition) is superimposed on a standard S. This standard should be the complementary copy of the combining site of a certain biological receptor (in its "active" conformation). It was assumed that the affinity, i.e. biological activity, of various potential effector molecules M_i towards the receptor, has a general tendency to decrease in a linear fashion versus this MTD. A procedure was also outlined to obtain the "best" standard, S_{opt}, the one which yields the highest correlation between MTD and the biological activity, and thus gives the most probable shape of the effector site.[2-5]

We want to outline here the connection of this MTD - site mapping procedure with graphs, and give some relations which may be useful in connecting this procedure to graph theory.

[+)] In all previous papers,[1-10] the term "minimal steric difference", MSD, was always used. However, "minimal topological difference", MTD, is a more suitable term.

The problem may be stated as follows. There are N molecules M_i, with known experimentally determined biological activities, A_i^{exp}, with respect to a receptor site ; the "best" standard, S_{opt}, should be inferred from these data. First the N molecules are superimposed, atom per atom, to give an atomic network - the hypermolecule $\underline{H}$, according to certain rules,[2-4] (hydrogen atoms are neglected, differences in bond lengths and angles also, but differences between covalent bonds and Van der Waals contact distances are not ignored ; reactive groups or groups giving strong interaction occupy always the same position in $\underline{H}$). Thus $\underline{H}$ is a graph, whose M vertices (indexed j = 1 to M) are connected by covalent bonds existing in at least one of the N molecules. Each molecule M_i is represented by one subgraph in $\underline{H}$, or by several subgraphs if more than one conformation g is considered for M_i. To characterise the M_i's on $\underline{H}$, a tri-indexed variable is used (where index i characterizes the molecule, index g its conformation, and index j the vertex of $\underline{H}$).

$$x_{i,g,j} = \begin{cases} = 1 \text{ , if vertex j of } \underline{H} \text{ is occupied in conformation g of molecule } M_i \text{ ;} \\ = 0 \text{ , if vertex j is empty, i.e. absence of an atom in that vertex in the conformation g of the molecule } M_i \text{ .} \end{cases} \qquad (1)$$

The vertices j may correspond to the interior of the receptor cavity (marked by $\delta_j = 1$, $\gamma_j = 1$),[2] to the walls of the receptor cavity ($\delta_j = 1$, $\gamma_j = 0$), or to regions irrelevant for steric fit (outside the receptor cavity, $\delta_j = 0$, $\gamma_j = 0$). The MTD is obtained by counting the unsuperposable atoms, i.e. the sterically relevant vertices occupied either only in S or in M_i, in the maximally superposable conformation. Thus, let us have a hypothetical standard S, defined by:[2-4]

$$S = \text{Min(MTD)} = \underset{(\text{over } g)}{\text{Min}} \sum_{j=1}^{M} \delta_j \left| x_{i,g,j} - \gamma_j \right| \qquad (2)$$

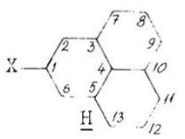
Also, a simpler equivalent formula can be used:

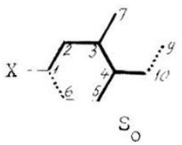
T a b l e 1

Example of a hypermolecule H with M = 13 vertices ; of N = 5 superimposed subgraphs M_i , and of standard S_o ; entries within the main part of the table are $x_{i,g,j}$ values ; X denotes a functional group.

M_i a)	i	g \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	MTD_i
-C	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6
-C-C	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	5
		2	1	0	0	0	0	1	0	0	0	0	0	0	0	7
-C-C-C-C-C	3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	2
		2	1	1	1	0	0	0	1	1	0	0	0	0	0	3
		3	1	1	1	1	0	0	0	0	0	1	0	0	0	2
-C-C-C_6H_5	4	1	1	1	1	1	0	0	1	1	1	1	0	0	0	2
		2	1	0	0	1	1	1	0	0	0	1	1	1	1	4
-C(C-C-C)(C-C-C)	5	1	1	1	1	0	1	1	1	0	0	0	1	0	0	3
S_o		γ_j	1	1	1	1	1	0	1	0	0	1	0	0	0	
		δ_j	1	1	1	1	1	1	1	0	1	1	0	0	0	
		ε_j	-1	-1	-1	-1	-1	+1	-1	0	+1	-1	0	0	0	

a) -C denotes $-CH_3$, -C-C denotes $-C_2H_5$, etc.





$$MTD = M(S) + \underset{(\text{over } g)}{\text{Min}} \sum_{j=1}^{M} \varepsilon_j \cdot x_{i,g,j} \qquad (3)$$

with $\varepsilon_j = -1$ for the receptor cavity vertices ($\delta_j = 1$, $\gamma_j = 1$), $\varepsilon_j = 1$ for receptor wall vertices ($\delta_j = 1$, $\gamma_j = 0$) and $\varepsilon_j = 0$ for sterically irrelevant vertices ($\delta_j = 0$, $\gamma_j = 0$) ; M(S) is the number of sterically relevant vertices occupied by the hypothetical standard S. In the accompanying example (table 1) the hypermolecule $\underline{H}$ has 13 vertices, while the standard S_o has 7 sterically relevant vertices, which are connected by heavy lines ; the dotted lines in S_o connects the former vertices to receptor wall vertices.

The calculated activities, A_i^{calc}, are given by

$$A_i^{calc} = \alpha - \beta MTD + \ldots . \qquad (4)$$

with α standing for a possible linear correlational equation with respect also to other (continous, usual) structural parameters.

The regressional coefficients, (α , β , etc.), are obtained by minimizing with respect to them the sum Y of square differences:

$$Y = \sum_{i=1}^{N} (A_i^{exp} - A_i^{calc})^2 \qquad (5)$$

If the standard S_o used to calculate the minimal topological differences MTD_i^o is optimal, Y should be minimal also with respect to the values ε_j^o which define S_o . If at a certain vertex t, ε_t^o is substituted by ε_t' , all 2M differences

$$Y(\varepsilon_t' , \varepsilon_t^o) \equiv Y(\varepsilon_t') - Y(\varepsilon_t^o) = \qquad (6)$$

$$= 2\beta \left[\sum_{i=1}^{N} (A_i^{exp} - A_i^{calc}) \cdot \Delta MTD_{i,t} + \frac{\beta}{2} \sum_{i=1}^{N} (\Delta MTD_{i,t})^2 \right]$$

should be nonnegative.

Here:

$$\Delta MTD_{i,t} = MTD'_i - MTD^o_i - \frac{1}{N}\sum_{i=1}^{N} (MTD'_i - MTD^o_i) \qquad (7)$$

and for each M_i a single conformation is possible (all g_i=1):

$$\Delta MTD_{i,t} = (\varepsilon'_t - \varepsilon^o_t)\cdot x_{i,t} - \frac{1}{N}\sum_{i=1}^{N} (\varepsilon'_t - \varepsilon^o_t)\cdot x_{i,t} \qquad (8)$$

The MTD^o_i 's are calculated with respect to S_o, the MTD'_i values with respect to the standard resulted by substituting ε'_t for ε^o_t .

Starting from an initial standard S_o, the most active M_i molecule for example, one performs all 2M differences $\Delta Y(\varepsilon'_t , \varepsilon^o_t)$. A new standard is obtained by performing the substitution corresponding to the most negative difference, and then new MTD'_i values and a new correlational equation (4) are found. The cycle is repeated until only nonnegative differences ΔY result.

Preliminary data [2] suggest that statistically significant correlations between biological activity and the best molecular shape can be obtained for any biological receptor. The selection of the initial standard seems quite critical - the procedure may easily yield local, nonabsolute minima for Y. Also, the statistical reliabilities for the assignment of ε_j for the vertices in S_{opt} remain to be determined ; qualitatively they should the higher, the higher the two values $\Delta Y(\varepsilon_j , \varepsilon^{opt}_j)$, and also the more often vertex j is encountered (either occupied or unoccupied) in all the N molecules M_i .

To conclude, our receptor mapping procedure seems to be an optimization problem for a tricoloured graph.

Acknowledgements. We are very indebted to Prof. Dr. A. T. Balaban for drawing our attention to graph theoretical implications of our procedure.

References

1 Z. Simon and Z. Szabadai,
Studia Biophys.(Berlin), 29, 123 (1973).

2 Z. Simon,
Preprint, Univ. Timişoara, Fac. Stiinţe Natur.,
Ser. Chim., 1976/1.

3 Z. Simon, I. I. Bădilescu, A. Chiriac, D. Ciubotariu,
S. Holban, F. Kerek, I. Moţoc, M. Mracek and Z. Szabadai,
Preprint, Univ. Timişoara, Fac. Stiinţe Natur.,
Ser. Chim., 1976/2.

4 Z. Simon, A. Chiriac, I. Moţoc, S. Holban, D. Ciubotariu
and Z. Szabadai,
submitted to Studia Biophys.(Berlin).

5 Z. Simon, I. I. Bădilescu and T. Kacoviţan,
submitted to J. Theoret. Biol.

6 A. Chiriac, D. Ciubotariu, Z. Szabadai, R. Vîlceanu and
Z. Simon,
Rev. Roumaine de Biochimie, 12, 143 (1975).

7 A. Chiriac, Z. Simon and R. Vîlceanu,
Studia Biophys.(Berlin), 51, 183 (1975).

8 I. I. Bădilescu and Z. Simon,
submitted to Rev. Roumaine Biochimie.

9 C. Drăgulescu, A. Maurer, S. Policec, M. Mracek and
Z. Simon,
unpublished results.

10 Z. Simon and L. Draskovits-Schuch,
in press at Rev. Roumaine Chim.