match no. 13 pp. 349-356 1982

MOLECULAR INTERACTIONS IN BIOLOGICAL SYSTEMS. VI.[1] AN INFORMATIONAL APPROACH BASED ON THE STERIC DIFFERENCE METHOD.

I.Motoc[a], O.Dragomir-Filimonescu[b] and R.Vâlceanu[a]

a) Chemistry Research Centre
B-dul M.Viteazul 24
1900 Timişoara, Romania.

b) Research Centre for Computing Technique
Str.Gh.Lazăr 9
1900 Timişoara, Romania.

(Received: October 29, 1981)

Abstract. The paper presents an informational approach for mapping the biological receptor. This approach is based on the IRS concept (i.e., Investigated Receptor Space) the SIBIS algorithm and the idea that the information content of a biomolecule may be partitioned into two parts : the fraction due to the interaction with the receptor cavity and the fraction due to the interaction with the receptor walls. The method (denoted by $\widetilde{SD}$) is applied with excellent results (i.e., $r = 0.993$, $F = 277.913$, $EV = 0.984$). The $\widetilde{SD}$ method is useful in QSAR and drug design in order to take into account the degree of uncertainty associated with the IRS.

1. Introduction

Within the Steric Difference (SD^{*}) method[2] one considers the series of bioactive structures S_1, S_2,..., S_n which elicit the biological response via the same mechanism and obtains[1,3] the corresponding Investigated Receptor Space (IRS) superimposing the n structures S_i according to the rule that one superimposes the pharmacophoric group of these structures. The derived space $\langle IRS \rangle$ is obtained[3] from the IRS by partitioning the IRS vertices into cavity vertices (c), wall vertices (w) and sterically irrelevant (i) ones. The $\langle IRS \rangle$ space

is viewed as a connected graph : the vertices p and q belonging to IRS are connected if and only if the edge p-q may represent a chemical bond ; $\langle \text{IRS} \rangle$ is specified as :

$$\langle \text{IRS} \rangle = \left[c(m_1,\ldots) \; ; \; w(n_1,\ldots) \; ; \; i(p_1,\ldots) \right]$$

where $m_1,\ldots,\; n_1,\ldots,\; p_1,\ldots,$ index the c-, w-, and i-type vertices of $\langle \text{IRS} \rangle$.

2. The $\widetilde{\text{SD}}$ Algorithm

Let us associate with the m vertices of the IRS the following finite probability scheme :

$$\begin{pmatrix} j \\ \alpha_j \end{pmatrix} = \begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha_1 & \alpha_2 & \ldots & \alpha_m \end{pmatrix}$$

α_j is the probability that the considered effectors interact with the receptor space centered around the vertex j of IRS ; α_j is computed as :

$$j = \sum_{i=1}^{n} x_{ij} \Big/ \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij} \quad ; \quad 0 \leqslant \alpha_j \leqslant 1$$

where x_{ij} are the entries of the <u>occupancy matrix</u>[1] $\underline{\underline{X}} = \left[x_{ij} \right]$, where $x_{ij} = 1$ if the vertex j is occupied by a non-hydrogen atom belonging to the effector i, and $x_{ij} = 0$ otherwise ; n and m stand for the number of effectors, and the number of IRS vertices, respectively.

The <u>information matrix</u> is $\underline{\underline{\widetilde{X}}} = \left[\widetilde{x}_{ij} \right]$, where $\widetilde{x}_{ij} = x_{ij} \alpha_j \log_2 \alpha_j$. The mean quantity of information $\bar{I}_i$ carried out by the effector i is computed as :

$$\bar{I}_i = \sum_{j=1}^{m} \widetilde{x}_{ij} \quad , \text{ bits / effector}$$

$\bar{I}_i$ is gained due to the effector interaction with the receptor cavity and walls, respectively. Accordingly, one may partition $\bar{I}_i$ as :

$$\bar{I}_i = \tilde{SD}_{c,i} + \tilde{SD}_{w,i}$$

($\tilde{SD}_c$ is the information revealed by the interaction with the receptor cavity, and $\tilde{SD}_w$ is the information revealed by the interaction with the receptor walls).

$\tilde{SD}_c$ and $\tilde{SD}_w$ are easily computed using $\tilde{\underline{I}}$ and $\langle IRS \rangle$:

$$\tilde{SD}_c = \sum_{j=cavity} \tilde{x}_{ij} \quad , \quad \tilde{SD}_w = \sum_{j=wall} \tilde{x}_{ij}$$

The vertices of i-type do not carry information and they are therefore neglected.

The Steric Difference method may be reformulated in terms of $\tilde{SD} = [\tilde{SD}_c ; \tilde{SD}_w]$ parameters removing the steps 3,4 and 5 of the SD^* algorithm[2] by the steps 3', 4' and 5' as follows :

3') Count the number of superposable atoms 1,2,...,p of the effector S_i over the standard S. Compute :

$$\tilde{SD}_{c,i} = \sum_{j=1}^{p} \tilde{x}_{ij}$$

4') Count the unsuperposable atoms 1,2,...,q of S_i over S . Compute :

$$\tilde{SD}_{c,i} = \sum_{j=1}^{q} \tilde{x}_{ij}$$

5') Compute the regression equation :

$$BR = a + b\,\tilde{SD}_c + c\,\tilde{SD}_w$$

by means of the least squares method ; BR stands for the biological response elicited by S_i.

The SIBIS algorithm[3] may be adapted to the $\tilde{SD}$ framework as follows :

1) Consider $\langle \text{IRS} \rangle$ initial. Compute the corresponding $\widetilde{\text{SD}}$-equation BR vs $\widetilde{\text{SD}}_c$, $\widetilde{\text{SD}}_w$ and its correlation coefficient.

2) Change the attribute of the vertex j of $\langle \text{IRS} \rangle$ (i.e., c $\longrightarrow$ w or i, w $\longrightarrow$ c or i, i $\longrightarrow$ w or c) if and only if the following two conditions hold :

i) the resulting $\widetilde{\text{SD}}$ - equation has better correlation coefficient.

ii) the subgraphs of c - and i - type vertices are left connected. The changes are performed until further improvements are not possible.

3) The resulting $\langle \text{IRS} \rangle$ is considered as $\langle \text{IRS} \rangle$ initial and step 2 is carried out for all vertices j = 1,2,...,m.

4) Continue steps 2 and 3 until no change of the vertex assignment occurs. The resulted $\langle \text{IRS} \rangle$ is $\langle \text{IRS} \rangle_{opt}$ and computing procedure is stopped.

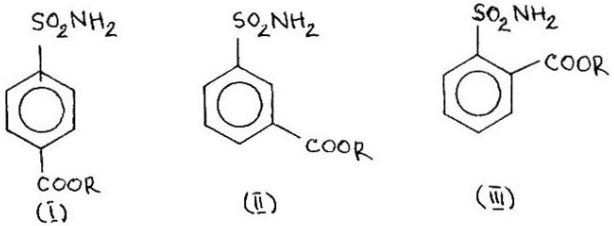
We note that changes c or w $\longrightarrow$ i imply to reevaluate each time the probabilities α_j and, accordingly, the entries of the matrix $\underset{\sim}{\underline{X}}$.

The FORTRAN-IV implementation of the $\widetilde{\text{SD}}$ method is closely related to the implementation of the SD* method[4]: the difference consists in a module which computes the $\widetilde{\text{SD}}_c$ and $\widetilde{\text{SD}}_w$. The structure of the input is as in SIBIS, except the key word which in this case is "SIBIS - IV".

$\widetilde{\text{SD}}$ is useful to evidence the vertices (i.e., the atoms and their topological coordinates) which have an important weight in conditioning the biological response.

3. Application

We apply the $\widetilde{\text{SD}}$ method to study the binding of sulfamyl benzoyl esters I - III to carbonic anhydrase.



The biological parameters, i.e., affinity constant, AC in M^{-1}, were reported in ref. 5 ; log AC values are displayed in Table 1.

Table 1. Carbonic anhydrase : experimental and computed biological parameters.

Compl. no.	R	log AC (exp.)	log AC (eq.1)	log AC (eq.2)
I. 1.	Me	7.98	7.746	8.031
2.	Et	8.5o	8.o9o	8.37o
3.	n-Pr	8.77	8.465	8.741
4.	n-Bu	9.11	8.795	9.o69
5.	n-Pent	9.39	9.o54	9.329
6.	n-Hex	9.39	9.197	9.329
II. 7.	Me	6.16	6.452	6.037
8.	Et	6.21	6.827	6.408
9.	n-Pr	6.44	7.157	6.736
lo.	n-Bu	6.95	7.416	6.995
11.	n-Pent	6.86	7.559	6.995
III.12.	Me	4.41	4.873	4.578
13.	Et	4.80	4.554	4.811
14.	n-Pr	5.28	4.93o	5.182
15.	n-Bu	5.76	5.259	5.510
16.	n-Pent	6.18	5.518	5.770

$\langle IRS \rangle_{init}$ is shown in Figure 1.

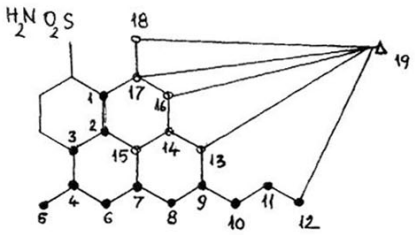


Figure 1. Carbonic anhydrase : $\langle IRS \rangle_{init}$; ● - cavity vertex, o - wall vertex, △- sterically irrelevant virtual vertex. (The vertices which are not numbered in Fig.1 are neglected because they are occupied by all effectors and, accordingly, are informationless).

Using the $\widetilde{SD}$-method, one obtains the following results:

$$\log AC = 5.621(\pm 0.134) + 1.005(\pm 0.055)\widetilde{SD}_c - 1.364(\pm 0.141)\widetilde{SD}_w \quad (1)$$

(42.105) (18.283) (-9.66)

(r=0.953, s=0.497, F=39.937, EV=0.895)

with $\langle IRS \rangle = \langle IRS \rangle_{init} = [c(1 \div 12); w(13 \div 18); i(19)]$ and

$$\log AC = 5.938(\pm 0.053) + 0.959(\pm 0.021)\widetilde{SD}_c - 2.290(\pm 0.073)\widetilde{SD}_w \quad (2)$$

(112.662) (46.014) (-31.201)

(r=0.993, s=0.196, F=277.913, EV=0.984)

with $\langle IRS \rangle = \langle IRS \rangle_{opt} = [c(1 \div 11,13); w(14,15,17,18); i(12, 16,19)]$

The numbers in parantheses below the regression coefficients represent the Student t-values, and r,s,F,EV stand for correlation coefficient, standard deviation, Fisher F - value and explained variance, respectively.

Table 2 collects the values of the $\widetilde{SD}$ parameters and the the effector mean quantity of information $\bar{I}_i$.

Table 2. Carbonic anhydrase :results.

Cmpd. no.	$\langle IRS \rangle_{init}$			$\langle IRS \rangle_{opt}$		
	$\widetilde{SD}_c$	$\widetilde{SD}_w$	$\bar{I}_i$	$\widetilde{SD}_c$	$\widetilde{SD}_w$	$\bar{I}_i$
1.	2.114	0.000	2.114	2.183	0.000	2.183
2.	2.455	0.000	2.455	2.537	0.000	2.537
3.	2.829	0.000	2.829	2.925	0.000	2.925
4.	3.156	0.000	3.156	3.267	0.000	3.267
5.	3.414	0.000	3.414	3.537	0.000	3.537
6.	3.556	0.000	3.556	3.537	0.000	3.537
7.	1.775	0.700	2.475	1.830	0.723	2.553
8.	2.149	0.700	2.848	2.218	0.723	2.941
9.	2.477	0.700	3.176	2.560	0.723	3.283
10.	2.734	0.700	3.434	2.830	0.723	3.553
11.	2.876	0.700	3.576	2.830	0.723	3.553
12.	1.088	1.351	2.439	1.118	1.062	2.180
13.	1.088	1.585	2.673	1.361	1.062	2.423
14.	1.462	1.585	3.047	1.748	1.062	2.810
15.	1.790	1.585	3.375	2.091	1.062	3.853
16.	2.047	1.585	3.632	2.361	1.062	3.421

The SD^* - equations corresponding to $\widetilde{SD}$ - equations(1) and (2) are as follows :

$$\log AC = 5.118(\pm 0.128) + 0.363(\pm 0.017)SD^*_c - 0.324(\pm 0.044)SD^*_w \quad (3)$$

(40.136) (21.462) (-7.408)

(r=0.957, s=0.475, F=44.168, EV=0.904)

with $\langle IRS \rangle = \langle IRS \rangle_{init} = [c(1 \div 12); w(13\ 18); i(19)]$ and

$$\log AC = 5.647(\pm 0.041) + 0.341(\pm 0.005SD^*_c - 0.708(\pm 0.020)SD^*_w \quad (4)$$

(136.849) (62.357) (-34.606)

(r=0.996, s=0.154, F=455.943, EV=0.990)

with $\langle IRS \rangle = \langle IRS \rangle_{opt} = [c(1 \div 11,13); w(14,15,17,18); i(12,16,19)]$

The agreement between SD^{*} - and $\overset{\sim}{SD}$ - equations is excellent. It is not surprising because this series of inhibitors is very well designed : the degree of uncertainty associated[6] with the $\langle IRS \rangle$ shown in Figure 1 is. only 4.55 % :

$$100\,\Delta\bar{I} / \bar{I}_{max} = 100(\bar{I}_{max} - \bar{I}) / \bar{I}_{max} = 100(-\log_2 18 + \sum_{j=1}^{18} \alpha_j \log_2 \alpha_j) / \log_2 18 = 100(4.1699 - 3.9802)/4.1699 = 4.55\ \%$$

(for further details on the information content of IRS see ref. 6)

References

1. Preceding part of this series, I.Motoc and R.Vâlceanu, Math.Chem., in press.
2. I.Motoc, Arzneim.Forsch., 31, 290 (1981).
3. I.Motoc, O.Dragomir.Filimonescu, Math.Chem., in press.
4. O.Dragomir-Filimonescu,I.Motoc and I.Muscutariu, Math. Chem., in press.
5. R.W.King and A.S.V. Burgen, Proc.Roy.Soc.London B, 193, 107 (1976).
6. I.Motoc and P.M.Reilly, Math.Chem., in press.