match no. 12 pp. 117-126 1981

# MOLECULAR INTERACTIONS IN BIOLOGICAL SYSTEMS
# I. STERIC INTERACTIONS. THE SIBIS ALGORITHM

Ioan Motoc[a,*] and O. Dragomir-Filimonescu[b]
a) Chemical Engineering Department
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
b) Research Centre for Computing Technique
Str. Gh. Lazar 9
1900 Timisoara, Romania
(Received: February 1981)

Summary. The SIBIS algorithm represents the advanced version of our Steric Difference (SD*) method. The aim of SIBIS is to generate bioactive structures using the computer. The algorithm is based on the method of least squares with subsidiary conditions. SIBIS is applied with very good results on the hapten-antibody interactions. A study of the stability, convergence and auto-ajustability of the algorithm is also presented.

## 1. Introduction

We present a new method to map the biological receptor. The method is termed SIBIS (abbreviation for Steric Interactions in BIological System). The aim of the SIBIS method is to compute the optimal standard used within the Steric Difference method[1,2], i.e., the SD* version[3-6]. The SIBIS algorithm is based on the method of least squares with subsidiary conditions.

SIBIS is useful in QSAR and drug design[7,11].

*) On leave from Chemistry Research Centre,
B-dul M. Viteazul 24, 1900 Timisoara, Romania.

2. The SIBIS Algorithm

We proceed by:

1) superimpose all the $S_i$ structure of the considered series of bioactive molecules $S_1$, $S_2$,...,$S_n$ according to the point 2 of the Steric Difference algorithm[1,3]. The resulted pattern reflects the topology of the receptor space investigated by the effectors which are being considered. This space is termed Investigated Receptor Space, abbreviated by IRS.

2) IRS constitutes[8] a convenient topological frame, namely: the m vertices of the IRS constitute the topological coordinates for the effector atoms. Accordingly, the chemical structure $S_i$ is described by the vector $\underline{X}_i = [x_{ij}]$, j = 1,2,...,m, where $x_{ij} = k_t$ if the vertex j of the IRS is occupied by the atom t of the effector i, and $x_{ij} = 0$ otherwise. $k_t$ is Austel value[12] associated with the atom t. The matrix $\underline{\underline{X}} = [x_{ij}]$, i = k,2,...,n, j = 1,2,...,m describes the structure of the effector series.

3) the derived space <IRS> is obtained from a given IRS by: one partitions the IRS vertices into cavity (c), wall (w) and sterically irrelevant (i) ones. IRS is viewed as a connected graph, i.e., the vertices p and g are connected if and only if the edge p-g may represent a chemical bond. The standard is defined by the c-type subgraph. <IRS> is specified as:

$$<\text{IRS}> = [c(m_1,\ldots) ; w(n_1,\ldots) ; i(p_1,\ldots)]$$

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

The SD* steric parameters are easily computed (the matrix $\underline{\underline{X}}$ and <IRS>):

$$SD^*_c = \sum_{\substack{\text{all j of} \\ \text{c-type}}} x_{ij} \quad \text{and} \quad SD^*_w = \sum_{\substack{\text{all j of} \\ \text{w-type}}} x_{ij}$$

The SIBIS algorithm consists of the following steps:

i) consider the starting <IRS> denoted by $<IRS>_{init}$. Compute the corresponding SD*-equation, i.e.,

$$BR = a + b\,SD^*_c + c\,SD^*_w \quad , b > 0 \ , c < 0$$ (BR: biological response)

and its correlation coefficient.

ii) change the attribute of the vertex j of the IRS (i.e., c↦w or i, w↦c or i, i↦c or w) if and only if the following two conditions hold:

1) the resulting SD*-equation has a better correlation coefficient; and

2) the subgraphs of c- and i- type vertices, respectively, are left connected.

The changes are performed until further improvements are not possible.

iii) the resulting <IRS> is considered as $<IRS>_{init}$ and step 2 is carried out for all vertices j = 1,2,...,m.

iv) continue steps 2 and 3 until no change of the vertex attribute occurs. The resulted <IRS> is optimal, denoted by $<IRS>_{opt}$. The computing procedure is stopped.

Concerning the SIBIS algorithm, we note:

1) the connectivity of the c- type subgraph must be preserved to conform with the definition of the standard (i.e., an organic compound is described by a connected graph[15]).

2) the connectivity of the i- type subgraph is also preserved in order to prevent meaningless improvements of the resulting SD*-equations.

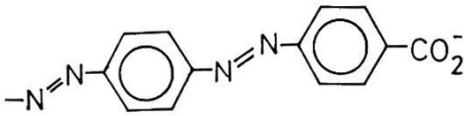
3) we recommend building up $<IRS>_{init}$ as: c- vertices correspond to the non-hydrogen atoms of the most potent drug of the series; the other vertices are of w- type ones; one introduces one virtual i- type vertex, i.e., a vertex $j_0$ such that $x_{ij_0} \equiv 0$ for all effectors i = 1,2,...,n. $j_0$ is connected with those vertices of the IRS for which one wishes to

to check the steric relevance.

3. Application

We apply the SIBIS algorithm to study the interaction of benzoates in the anti-p- (p'-azophenylazo) benzoate system. The biological parameters used in the present calculations are taken from ref. 8. The data are collected in Table 1 ($K_{rel}$ is the relative equilibrium constant).

$\langle IRS \rangle_{init}$ is shown in Figure 1. The standard, S, corresponds to the antigenic part of the immunogen, i.e.,



($CO_2^-$ is the pharmacophoric group).

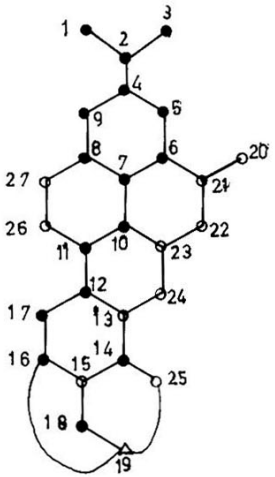



Figure 1. Hapten-antibody interaction, $\langle IRS \rangle_{init}$

(• : c- vertex, o: w- vertex ; △: i- vertex (virtual))

The vector $\underline{X}_i$ is easy to write using IRS shown in Figure 1. For example $\underline{X}_1$ (hapten no. 1 in Table 1) is:

$\underline{X}_1$ = [ 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ] ;

$\underline{X}_{11}$ (hapten no. 11 is Table 1) is:

$\underline{X}_{11}$ = [ 1 1 1 1 1 1 1 1 1 1.7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ] etc.

There resulted the following SD*-equations:

$$\log K_{rel} = -1.683(\pm 0.073) + 0.213(\pm 0.006)SD^*_c - 0.367(\pm 0.049)SD^*_w \quad (1)$$

(n= 23, r = 0.947, s = 0.336, F = 55.576, EV = 0.888)

with $\langle IRS \rangle_{init}$ = [ c(1:18) ; w(20:27) ; i(19) ]

$$\log K_{rel} = -3.357(\pm 0.064) + 0.379(\pm 0.005)SD^*_c - 0.365(\pm 0.040)SD^*_w \quad (2)$$

(n = 23, r = 0.960, s = 0.293, F = 74.584, EV = 0.914)

with $\langle IRS \rangle_{opt}$ = [ c(1:13,16,17) ; w(14,20:24,26,27) ; i(15,18,19,25) ]

Equations (1) and (2) are significant at >99 percentile level (F statistic) and are in excellent agreement with the experimental facts, i.e., antibodies are relative rigid macromolecules[9] and the antibody active site "copies" the shape of the antigenic part of the immunogen[8].

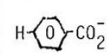
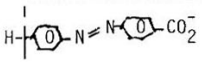
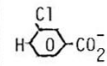
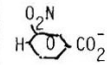
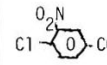
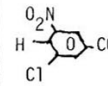
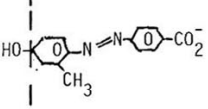
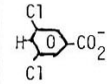
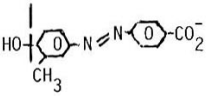
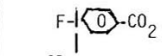
## 4. The SIBIS algorithm: convergence, stability and auto-adjustability

The computations reported below use the biological parameters corresponding to the carboxypeptidase inhibitors[12] collected in Table 2 ($Y_1$ stands for -log K, K being the inhibition constant).

4.1 Convergence. The SIBIS algorithm is considered to converge if $\langle IRS \rangle_{opt}$ does not depend on the considered $\langle IRS \rangle_{init}$.

The data collected in Table 3 argue that SIBIS converges. The standard S corresponds to the compounds no. 3, 7 and 10, respectively (see Table 2), within $\langle IRS \rangle_{init,4}$, $\langle IRS \rangle_{init,3}$ and $\langle IRS \rangle_{init,1}$.

Table 1. Hapten-antibody interactions: $K_{rel}$ values

| No. | Hapten | $K_{rel}$ | No. | Hapten | $K_{rel}$ |
|---|---|---|---|---|---|
| 1. |  | 1.00 | 18. |  | 67 |
| 2. |  | 0.43 | 19. |  | 70 |
| 3. |  | 0.12 | 20. |  | 65 |
| 4. |  | 0.41 | 21. |  | 111 |
| 5. |  | 0.07 | 22. |  | 81 |
| 6. |  | 0.024 | 23. |  | 125 |
| 7. |  | 3.6 | | | |
| 8. |  | 5.4 | | | |
| 9. |  | 1.8 | | | |
| 10. |  | 5.4 | | | |
| 11. |  | 9.8 | | | |
| 12. |  | 2.1 | | | |
| 13. |  | 4.7 | | | |
| 14. |  | 1.8 | | | |
| 15. |  | 6.8 | | | |
| 16. |  | 5.3 | | | |
| 17. |  | 1.6 | | | |

$\langle IRS \rangle_{init,2}$ is shown in Figure 2.

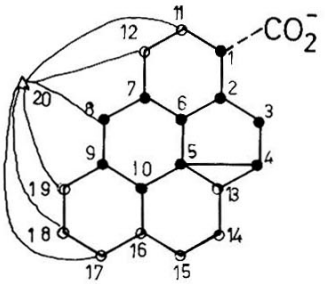



Figure 2. Carboxypeptidase inhibitors: $\langle IRS \rangle_{init,2}$

4.2. Stability. The SIBIS algorithm is considered[13] stable if small perturbations of the input data do not change significantly the resulted SD*-equation and $\langle IRS \rangle_{opt}$.

Table 2. Carboxypeptidase Inhibitors

| No. | Carboxylic acid (inhibitor) | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
|---|---|---|---|---|---|
| 1. | $C_2H_5COO^-$ | 1.00 | 0.996 | 1.01 | 2.34 |
| 2. | $n\text{-}C_3H_7COO^-$ | 2.30 | 2.298 | 2.32 | 0.83 |
| 3. | $n\text{-}C_4H_9COO^-$ | 2.57 | 2.573 | 2.55 | 2.57 |
| 4. | $n\text{-}C_5H_{11}COO^-$ | 2.20 | 2.196 | 2.10 | 2.20 |
| 5. | $(CH_3)_2CH\ CH_2COO^-$ | 2.57 | 2.574 | 2.70 | 2.57 |
| 6. | $C_6H_5COO^-$ | 0.83 | 0.834 | 0.80 | 2.30 |
| 7. | $C_6H_5CH_2COO^-$ | 2.34 | 2.336 | 2.40 | 1.00 |
| 8. | $C_6H_5(CH_2)_2COO^-$ | 2.92 | 2.924 | 2.80 | 1.66 |
| 9. | $C_6H_5(CH_2)_3COO^-$ | 1.70 | 1.697 | 1.75 | 1.48 |
| 10. | β-indolyl $CH_2COO^-$ | 3.11 | 1.109 | 3.11 | 2.26 |
| 11. | β-indolyl $(CH_2)_2COO^-$ | 2.26 | 2.256 | 2.16 | 3.11 |
| 12. | β-indolyl $(CH_2)_3COO^-$ | 1.48 | 1.477 | 1.70 | 1.70 |
| 13. | β-naphtyl $CH_2COO^-$ | 1.66 | 1.664 | 1.50 | 2.92 |

Y's collected in Table 2 would be exactly the same as the published set, i.e., Y's, if rounded to the published number of digits. There resulted the following SD*-equations:

$$Y_1 = 1.159(\pm 0.123) + 0.220(\pm 0.022)SD^*_c - 0.158(\pm 0.047)SD^*_w \qquad (3)$$

($r = 0.814$, $s = 0.404$, $F = 5.911$)

$$Y_2 = 1.154(\pm 0.124) + 0.224(\pm 0.022)SD^*_c - 0.158(\pm 0.049)SD^*_w \qquad (4)$$

($r = 0.812$, $s = 0.408$, $F = 5.902$)

with $\langle IRS \rangle_{init}$ = [ c(1:10) ; w(11:19) ; i(20) ]

Table 3. Convergence: <IRS>'s

| Vertex no. | $\langle IRS \rangle_1$ init. | $\langle IRS \rangle_1$ opt. | $\langle IRS \rangle_2$ init. | $\langle IRS \rangle_2$ opt. | $\langle IRS \rangle_3$ init. | $\langle IRS \rangle_3$ opt. | $\langle IRS \rangle_4$ init. | $\langle IRS \rangle_4$ opt. |
|---|---|---|---|---|---|---|---|---|
| 1 | c | c | c | c | c | c | c | c |
| 2 | c | c | c | c | c | c | c | c |
| 3 | c | c | c | w | c | w | w | c |
| 4 | c | w | c | c | c | c | w | w |
| 5 | c | c | c | c | c | c | c | c |
| 6 | c | c | c | c | c | c | c | c |
| 7 | c | c | w | c | w | c | w | c |
| 8 | c | c | w | c | w | c | w | c |
| 9 | c | c | w | c | w | c | w | c |
| 10 | c | c | c | c | w | c | w | c |
| 11 | w | w | w | w | w | w | w | w |
| 12 | w | w | w | w | w | w | w | w |
| 13 | w | c | c | c | c | c | w | c |
| 14 | w | c | w | c | w | c | w | c |
| 15 | w | w | w | w | w | w | w | w |
| 16 | w | w | w | w | w | w | w | w |
| 17 | w | i | w | i | w | i | w | i |
| 18 | w | w | w | w | w | w | w | w |
| 19 | w | w | w | w | w | w | w | w |
| 20 | i | i | i | i | i | i | i | i |

The equations (3) and (4) and the data collected in Table 3 argue the statement that SIBIS is a stable algorithm.

4.3. Auto-adjustability

The $Y_3$ and $Y_4$ values (Table 2) represent medium and strongly perturbed $Y_1$ values. Using SIBIS, the following equations resulted:

$$Y_3 = 0.756(\pm 0.096) + 0.403(\pm 0.017)SD^*_c - 0.655(\pm 0.058)SD^*_w \quad (5)$$

($r = 0.882$, $s = 0.324$, $F = 10.544$), with

$\langle IRS \rangle_{opt.}$ = [ c(1÷3, 5÷9, 13, 14) ; w(4, 10÷12, 15, 18, 19) ; i(16, 17, 20) ].

$$Y_4 = 0.804(\pm 0.176) + 0.304(\pm 0.033)SD^*_c - 0.156(\pm 0.065)SD^*_w \quad (6)$$

($r = 0.554$, $s = 0.580$, $F = 1.328$), with

$\langle IRS \rangle_{opt.}$ = [ c(1, 2, 5÷7, 10, 12, 15, 16) ; w(3, 4, 8, 9, 11, 13, 14, 17÷19); i(20) ]

The statement that the auto-adjustability of the SIBIS algorithm is quite low is clearly supported by the equation (6), i.e., equations with $r^2 < 0.40$ are chance correlations[14].

References

1. I. Motoc, Can. J. Pharm. Sci., in press.
2. I. Motoc and R. Valceanu, Math. Chem., 8, 333 (1980).
3. I. Motoc, Arzneimittel Forsch., in press.
4. I. Motoc, I. Muscutariu and D. Ciubotariu, Kexue Tongbao, in press.
5. I. Motoc and I. Muscutarin, Math. Chem., this issue.
6. I. Motoc, Preprint, Timisoara Univ., Serie Chim., No. 5/1980.
7. A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, "Steric Fit in QSAR", Springer, Berlin, 1980.

8. D. Pressman and A.L. Grossberg, "Structural Basis of Antibody Specificity", Benjamin, New York, 1968.

9. R.J.P. Williams, Angew. Chem. Int. Ed., 16, 766 (1977).

10. I. Motoc and O. Dragomir-Filimonescu, Eur. J. Med. Chem., in press.

11. V. Austel, E. Kutter and W. Kalbfleisch, Arzneimittel Forsch., 29, 585 (1979).

12. L. Lumry, E.L. Smith and J. Polglase, J. Phys. Chem., 55, 125 (1951).

13. A.E. Beaton, D.B. Rubin and J.L. Barone, J. Am. Stat. Assoc., 71, 158 (1976).

14. J.G. Topliss and R.J. Costello, J. Med. Chem., 15, 1066 (1972).

15. A.T. Balaban, ed., "Chemical Applications of Graph Theory", Academic, London, 1976.