match no. 11 pp.169-183 1981

NEW INFORMATIONAL INDICES FOR INTERACTIONS IN BIOLOGICAL SYSTEMS

Ioan Moţoc
Chemistry Research Centre
Mihai Viteazu 24
1900 Timişoara, Romania

(Received November 1980)

Summary. The investigated receptor space (IRS) by a given series of effectors is defined as the graph G whose overlap partition is $G_1,\ldots,G_n$. The graph G_L, $1 \leq L \leq n$, is the hydrogen-suppressus graph corresponding to the effector molecule L. One computs the mean, $\bar{I}(G)$, and total, I(G), quantity of information contained in the investigated receptor space. $\bar{I}(G)$ or I(G) induces an ordering of IRS's and constitues a criterion of the IRS relevance. These indices are of use in the quantitative structure-activity relations. Two concrete series of hapten-antibody interactions and one series of carboxypeptidase inhibitors are studied.

1. Definitions

i) The effector (pharmacon, biomolecule, or bioactive molecule) is a chemical compound which elicits a biological response in an organism.

ii) The effector series is a collection of effectors which elicit a biological response of specifyed type *via* the same mechanism.

iii) The pharmacophore (or, pharmacophoric group) is the set of atoms belonging to the effector which trigger the biological response to occur.

iv) The biological receptor is a structurated entity, located in general on membranes, which mediates the biological response to occur.

2. Notations

$G_L = (X_L, \Gamma_L)$ is the hydrogen-suppressed graph corresponding to the effector L.

m_L = Card X_L is the number of vertices in the graph G_L.

$G = (X, \Gamma)$ is the graph of the investigated receptor space (IRS).

m = Card X is the number of vertices in the graph G.

n is the number of effectors in the considered series.

$X = [x_{ij}]$, where $x_{ij} = 1$ if the vertex $j \in X$ and $j \in X_i$, and $x_{ij} = 0$ if $j \in G$ and $j \notin X_i$. X expresses the topology of the effectors within the topological frame G.

r, s, P, EV and t have the usual statistical connotation.

The informational indices belong to the class of topological indices[2,3,4]. For other basic definitions in graph theory one may consult refs. 5 and 15.

3. The Investigated Receptor Space

In order to compare the topology of the effectors of a given series, one superimposes the considered structures. The rule[6-8] is to superimpose the pharmacophores. The superposition of the n effectors of the series reflects the topology of the

receptor space explored by the considered effectors. This space is termed "the investigated receptor space" (IRS).

In order to get an easily to use method to perform the geometrical congruences we recommand to neglect the differences lower than ± 0.20 Å and $\pm 20^{\circ}$ in the bond lengths and bond angles, respectively.

Concerning IRS, one may easily prove :

Proposition. The following statements are equivalent :

1) $G = \bigcup_{L=1}^{n} G_L$

2) G is an IRS if and only if $G_1, \ldots, G_n$ is an overlap partition of G.

The IRS concept has been used within the MTD method[11] in a slightly modifyed acception.

4. New Informational Indices

Let consider an effector series given by the matrix X :

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & & & & \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & & & & \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}$$

One may associate to the m vertices of the corresponding graph G of the IRS the following finite probability scheme:

$$\begin{pmatrix} 1 & \cdots & j & \cdots & m \\ p_1 & \cdots & p_j & \cdots & p_m \end{pmatrix}$$

where

$$p_j = \sum_{i=1}^{n} x_{ij} \Big/ \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij} = N_j / N$$

Obviously, $0 \leqslant p_j \leqslant 1$, and $\sum_{j=1}^{m} p_j = 1$.

One may define, according to the well known equations[1,10], the mean quantity of information :

$$\overline{I}(G) = - \sum_{j=1}^{m} p_j \log_2 p_j, \text{ bits/IRS} \qquad (1)$$

and the total quantity of information :

$$I(G) = N.\overline{I}(G) = N \log_2 N - \sum_{j=1}^{m} N_j \log_2 N_j, \text{ bits/IRS} \qquad (2)$$

$\overline{I}(G)$ or $I(G)$ induces an ordering or IRS's and furnishes a criterion of the IRS relevance.

It is of interest to characterize the degree of uncertainty of the vertex j of the graph G corresponding to IRS. For this purpose, one may use the entropy H_j of the vertex :

$$H_j = - \log_2 p_j , \quad 0 \leqslant H_j \leqslant 1 \qquad (3)$$

or, simply, the 100 p_j values.

We may easily adapt our Steric Difference method[6-8] (for an up-to-date review on this topic one may consult ref.9) to include the vertex uncertainty. One proceeds as follows : the SD steric parameters corresponding to the effector i against the IRS specifyed by the graph G :

$$SD_c^* = \sum_p x_{ip} k_p , \quad SD_w^* = \sum_q x_{iq} k_q \qquad (4)$$

should become :

$$\overline{SD}_c^* = \sum_p x_{ip} k_p \cdot 100\, p_p, \quad \overline{SD}_w^* = \sum_q x_{iq} k_q \cdot 100\, p_q \qquad (5)$$

The summations are taken over the vertices p belonging to the receptor cavity, and, respectively, the vertices q belonging to the receptor wall.

k_p characterizes[12] sterically the atom p (i.e., k_H = = 0 ; k = 1 for the first row elements, except F, k_F = 0.8 ; k = 1.2, 1.3, 1.7 for the 2-nd, 3-nd and 4-th row elements, respectively).

x_{ip}, x_{iq} are the entries of the matrix X which describes the structure of the considered effectors within the topological frame G.

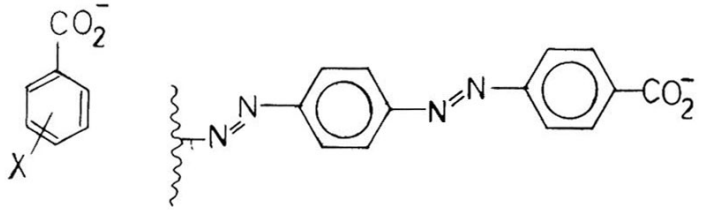
Similarly, one may adapt the MTD method[11] to account for the vertex uncertainty (the resulted parameter is denoted by $\overline{\text{MTD}}$).

5. Applications

The utility of the vertex uncertainty (expressed by 100 p_j) is investigated calculating the SD^*- and $\overline{SD}^*$ - type equations (see relations 4 and 5) for two series of hapten -antibody interactions and the SD^*, $\overline{SD}^*$, MSD and $\overline{\text{MSD}}$ equations for a series of carboxypeptidase inhibitors.

5.1. Interaction of benzoates in the anti-p-(p'--azophenylazo) benzoate system

The relative equilibrium constants, K_{rel}, for the combination of haptens (I) with the antibody against (II) are taken from ref. 13a.



I II

The K_{rel} values and SD* and $\overline{SD}$* parameters are collected in Table 1. The Steric Difference parameters were computed using the IRS shown in Figure 1. The pharmacophore is the $COO^{\bullet}$ group, and the antigenic part of the immunogen (II) is considered complementary to the receptor cavity.

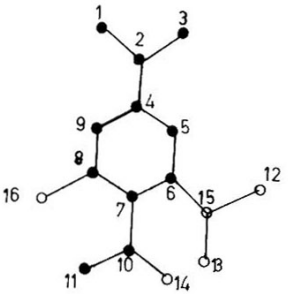


Figure 1. The investigated receptor space (IRS) by the haptens I (• - cavity vertex ; o - wall vertex).

Table 1. Haptens I : K_{rel} values and Steric Difference parameters.

No.	Hapten	K_{rel}	SD^*		$\overline{SD}^*$	
			SD_c^*	SD_w^*	$\overline{SD}_c^*$	$\overline{SD}_w^*$
1.	m - Cl	0.43	9	1.2	77.76	7.40
2.	p - Cl	5.30	10.2	0	82.20	0.00
3.	m - NO_2	0.12	9	3	77.76	7.40
4.	p - NO_2	1.80	11	1	82.69	1.23
5.	3-NO_2 , 4-Cl	0.41	10.2	3	82.20	13.57
6.	3-NO_2 , 5-Cl	0.07	9	4.2	77.76	16.52
7.	3,5 - di - Cl	0.024	9	2.4	77.76	10.35

The matrix $X = [x_{ij}]$ describing the structure of the effectors collected in Table 1 is shown below :

i \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0
2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0
4	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0
5	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0
6	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1
7	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1
N_j	7	7	7	7	7	7	7	7	7	3	1	3	3	1	5	2

(N = 81)

The following log K_{rel} *vs* SD_c^* , SD_w^* and $\overline{SD}_c^*$, $\overline{SD}_w^*$ equations were computed (the symbols have the usual connotation):

$$\log K_{rel} = -4.554(\pm 0.172) + 0.486(\pm 0.018)SD_c^* - 0.296(\pm 0.069) SD_w^* \quad (6)$$

$(r=0.890, s=0.371, F=3.802, EV=0.687, r(SD_c^*)=0.767, r(SD_w^*)=-0.782, r(SD_c^*, SD_w^*)=-0.515)$

$$\log K_{rel} = -14.668(\pm 0.165) + 0.184(\pm 0.002)\overline{SD}_c^* - 0.064(\pm 0.017) \overline{SD}_w^* \quad (7)$$

$(r=0.900, s=0.357, F=4.164, EV=0.710, r(\overline{SD}_c^*)=0.800, r(\overline{SD}_w^*)=-0.757, r(\overline{SD}_c^*, \overline{SD}_w^*)=-0.504)$

The equation (7) is slightly better than the equation (6) (see s, F and EV values), but the most important distinction among these equations consists in the t-Student values corresponding to the predictor variables :

	SD_c^*	SD_w^*	$\overline{SD}_c^*$	$\overline{SD}_w^*$
t =	27.38	-4.31	88.79	-3.78

$(t_{0.05\ ;\ 6} = 2.447\ ,\ t_{0.01\ ;\ 6} = 3.707)$

5.2. Interaction of para-substituted benzoates in the anti-p-(p'-azophenylazo)benzoates system.

The K_{rel} for the eighteen haptens collected in the Table 2 are taken from the reference 13 b. The Steric Difference parameters were computed against the IRS shown in Figure 2.

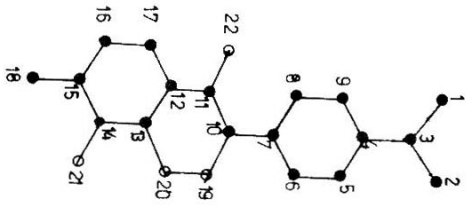
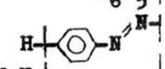
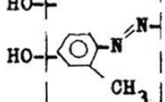
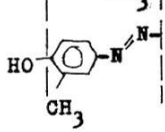


Figure 2. The investigated receptor space by the para-substituted benzoates (● - cavity vertex; O - wall vertex). The matrix X associated to the effector series displayed in the Table 2 is :

i \ j	1	2	3	4	5	6	7	8	9	1o	11	12	13	14	15	16	17	18	19	2o	21	22
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X=9	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0
1o	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	0
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
N_j	18	18	18	18	18	18	18	18	17	1o	8	7	6	6	6	6	5	3	2	1	1	1

(N = 240)

Table 2. Para-substituted benzoate haptens : K_{rel} values and Steric Difference parameters.

No.	Hapten	K_{rel}	SD^*		$\overline{SD}^*$	
			SD_o^*	SD_w^*	$\overline{SD}_o^*$	$\overline{SD}_w^*$
1.	H-	1.0	9	0	64.56	0.00
2.	F-	3.6	9.8	0	71.02	0.00
3.	Cl-	5.3	10.2	0	74.24	0.00
4.	H_2C-	1.8	10	0	72.63	0.00
5.	Br-	5.4	10.3	0	75.05	0.00
6.	I-	9.0	10.7	0	78.28	0.00
7.	H_2N-	2.1	10	0	72.63	0.00
8.	HO-	4.7	10	0	72.63	0.00
9.	O_2N-	1.8	11	1	77.11	1.35
10.	^-O_2C-	5.3	11	1	77.11	1.35
11.	$H_3CC(O)HN$-	1.6	12	1	80.70	0.45
12.	C_6H_5-	6.8	13	2	83.84	2.25
13.		67	17	0	94.60	0.00
14.	H_2N- 	70	18	0	96.84	0.00
15.	H_3C- 	65	18	0	96.84	0.00
16.	HO- 	111	18	0	96.84	0.00
17.	HO- 	81	18	1	96.84	0.90
18.	HO 	125	18	1	96.84	0.45

The following QSAR's were derived with the data collected in the Table 2 :

$$\log K_{rel} = -1.245(\pm 0.058) + 0.191(\pm 0.005)\ SD_o^* - 0.161(\pm 0.161(\pm 0.082)\ SD_w^* \qquad (8)$$

$(r=0.95, s=0.23 F=41.38, EV=0.89, r(SD_c^*)=0.93, r(SD_w^*)=-0.01,$
$r(SD_c^*, SD_w^*)=0.133)$

$$\log K_{rel} = -3.999(\pm 0.057)+0.061(\pm 0.0007)\overline{SD}_c^* - 0.175(\pm 0.077)\overline{SD}_w^* \quad (9)$$

$(r=0.95, s=0.23, F=42.62, EV=0.89, r(\overline{SD}_c^*) = 0.93, r(\overline{SD}_w^*)=-0.01,$
$r(\overline{SD}_c^*, \overline{SD}_w^*)=0.07).$

and

	SD_c^*	SD_w^*	$\overline{SD}_c^*$	$\overline{SD}_w^*$
t =	41.29	-1.97	89.17	-2.27

$(t_{0.05;17} = 2.101 , t_{0.01;17} = 2.898)$

We note that on passing from the SD^*- to the $\overline{SD}^*$-type equation, the relevance of the SD_c^* term impoves.

5.3. Carboxypeptidase Inhibitors

The IRS explored by the series of carboxypeptidase inhibitors collected in Table 3 in shown in Figure 3.

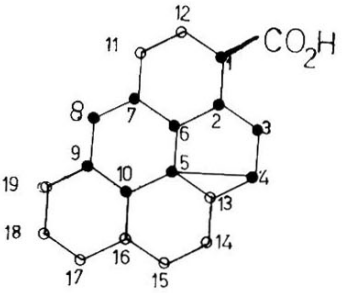


Figure 3. The investigated receptor space by the caboxypeptidase inhibitors (• - cavity vertex ; O - wall vertex).

The X matrix associated to the studied carboxypeptidase inhibitors is given below.

The following equations were derived with the data displayed in Table 3 :

$$Y = 1.144(\pm 0.123) + 0.222(\pm 0.021)SD_c^* - 0.187(\pm 0.047)SD_w^* \quad (10)$$

$(r=0.814, s=0.405, F=5.885, EV=0.595, r(SD_c^*)=0.584, r(SD_w^*)=-0.456, r(SD_c^*, SD_w^*)=0.176)$

$$Y = 3.317(\pm 0.118) - 0.205(\pm 0.018)MSD \quad (11)$$

$(r=0.809, s=0.409, F=9.515, EV=0.624)$

$$Y = 0.041(\pm 0.121) + 0.042(\pm 0.003)\overline{SD}_c^* - 0.052(\pm 0.012)\overline{SD}_w^* \quad (12)$$

$(r=0.820, s=0.398, F=6.166, EV=0.607, r(\overline{SD}_c^*)=0.600, r(\overline{SD}_w^*)=-0.324, r(\overline{SD}_c^*, \overline{SD}_w^*)=0.337)$

and

$$Y = 3.215 - 0.044\ \overline{MSD} \quad (13)$$

$(r=0.814, s=0.404, F=9.828, EV=0.632)$

The SD^*- type equations are slighly better than MSD ones. The main defferences among SD^* - $\overline{SD}^*$, and MSD - $\overline{MSD}$ equations lie in the t-Student values, namely :

	SD_c^*	SD_w^*	MSD	$\overline{SD}_c^*$	$\overline{SD}_w^*$	$\overline{MSD}$
t =	10.69	-3.99	-11.46	20.45	-4.13	-10.81

$(t_{0.05;12} = 2.179,\ t_{0.01;12} = 3.055)$

References

1. D.Bonchev, Math.Chem., 7, 65(1979).

2. D.H.Rouvray, Am.Sci., 61, 729(1973).

Table 3. Carboxypeptidase inhibitors : biological activities and parametrization.

No.	Inhibitor	Y	SD^*		MSD	$\overline{SD}^*$		$\overline{MSD}$
			SD_c^*	SD_w^*		$\overline{SD}_c^*$	SD_w^*	
1.	C_2H_5COOH	1.00	2	0	8	27.96	0.00	49.48
2.	$n\text{-}C_3H_7COOH$	2.30	3	0	7	40.86	0.00	36.58
3.	$n\text{-}C_4H_9COOH$	2.57	4	0	6	51.61	0.00	25.83
4.	$n\text{-}C_5H_{11}COOH$	2.20	5	0	5	59.53	0.00	18.30
5.	$(CH_3)_2CHCH_2CH_2COOH$	2.57	5	0	5	56.99	0.00	20.45
6.	C_6H_5COOH	0.83	4	2	8	46.24	2.16	33.36
7.	$C_6H_5CH_2COOH$	2.34	6	1	5	58.07	5.38	24.75
8.	$C_6H_5CH_2CH_2COOH$	2.92	8	0	2	70.98	0.00	6.46
9.	$C_6H_5(CH_2)_3COOH$	1.70	5	4	9	59.53	17.21	35.51
10.	$\beta\text{-}C_8H_6NCH_2COOH$	3.11	10	0	0	77.44	0.00	0.00
11.	$\beta\text{-}C_8H_6N(CH_2)_2COOH$	2.26	7	4	7	67.75	17.21	26.90
12.	$\beta\text{-}C_8H_6N(CH_2)_3COOH$	1.48	6	6	10	62.37	17.22	32.29
13.	$\beta\text{-}C_{10}H_7CH_2COOH$	1.66	7	4	7	65.60	17.21	29.05

(Y's refer to - log K, K is the inhibition constant, mole/l. (ref. 14)

i \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
5	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0
X=7	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
8	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	1	1	0	0	0	1	0	0	1	1	1	1	0	0	0
10	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
11	1	1	0	0	1	1	1	1	0	1	0	0	1	1	1	1	0	0	0
12	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	1	1
13	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	0	0	0
N_j	13	13	3	3	10	12	5	3	3	7	1	1	5	4	3	4	1	1	1

(N = 93)

3.A.T.Balaban,Theor.Chim.Acta(Berl.),53,355(1979).

4.A.T.Balaban,I.Chiriac,I.Moţoc and Z.Simon,"Steric Fit in QSAR",Lecture Notes in Chemistry, vol.15,Springer,Berlin, 1980.

5. J.W.Essam and M.E.Fischer,Rev.Mod.Phys.,42,272(1970).

6. I.Moţoc,J.Biol.Chem.(in press).

7. I.Moţoc and R.Vâlceanu,Math.Chem.(in press).

8. I.Moţoc,Arzneimittel Forsch.(in press).

9. I.Moţoc,"Quantitative Treatment of the Shape of Biomolecules : The Steric Difference Method",Preprint,Timişoara University,Serie Chimie No.5/1980.

10. C.Shannon and W.Weaver,"Mathematical Theory of Communication",Univ. of Illinois,Urbana,1949.

11. Z.Simon,A.Chiriac,I.Moţoc,S.Holban,D.Ciubotariu and Z.Szabadai,Studia Biophys.(Berl.),55,217(1976).

12. V.Austel,E.Kulter and W.Kalbfleisch,Arzneimittel Forsch., 29,585(1979).

13. a)D.Pressman and A.L.Grossberg,"Structural Baisi of Antibody Specificity",Benjamin,New York,1968,p.60;b)ibid.,p.58.

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

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