match no. 13 pp. 333-343 1982

MOLECULAR INTERACTIONS IN BIOLOGICAL SYSTEMS. IV[1]. QUANTITATIVE MEASURE OF THE INFORMATION CONTENT OF THE INVESTIGATED RECEPTOR SPACE.

I. Motoc[a] and P.M. Reilly[b]

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

[b] University of Waterloo
Chemical Engineering Department
Waterloo, Ontario N2L 3G1, Canada

(Received: October 29, 1982)

Abstract. Using our IRS concept (i.e., Investigated Receptor Space), the paper shows a convenient method to compute the mean quantity of information ($\overline{I}_i$) of a biomolecule as well as the mean quantity of information ($\overline{I}$) of the IRS associated with a specified effector series. The derived quantity $\Delta\overline{I} = \overline{I}_{max} - \overline{I}$ is a measure of the uncertainty associated with a given IRS. $\overline{I}$ and $\Delta\overline{I}$ (or, $100\,\Delta\overline{I}/\overline{I}_{max}$) furnish general criteria for ordering the IRS's, or the effector series. The method here outlined is applied to hapten - antibody interaction and to the binding to carbonic anhydrase of sulfonamides.

1. Introduction

Within the Steric Difference method[2-4] the structure of the bioactive molecules belonging to a reaction series is described by the hydrogen-suppressed graphs corresponding to the bioactive compounds under consideration. The topology of the molecules is quantitatively compared using the IRS (i.e., Investigated Receptor Space) defined[5] as :

Definition 1. The graph G is an IRS if and only if $G_1, \ldots, G_n$ is an overlop partition of G.

G_l, $l = 1, 2, \ldots, n$, stands for the hydrogen-suppressed graph associated with the effector l.

The Definition 1 is equivalent with :

Definition 2. The graph G is an IRS if and only if $G = \bigcup_{l=1}^{n} G_l$.

Using the IRS concept and notions of information theory[6,7], one develops quantitative measures of the information content of the IRS or effector series and the uncertainty associated with the IRS, respectively. One obtains a convenient basis for ordering the effector series and IRS's

2. The Mean Quantity of Information and the Uncertainty of the IRS.

Consider the IRS occupancy matrix defined as : $X = \left[x_{ij}\right]$ with $x_{ij} = 1$ if the vertex $j \in$ IRS is occupied by a non-hydrogen atom belonging to the effector i, and $x_{ij} = 0$ if it is not occupied.

The finite probability scheme associated with the IRS :

$$\begin{pmatrix} 1 & 2 & \ldots & m \\ p_1 & p_2 & \ldots & p_m \end{pmatrix} = \begin{pmatrix} j \\ p_j \end{pmatrix}$$

describes the probability that the considered n effectors interact with the receptor space centered around the vertex $j \in$ IRS.

p_j , $j = 1, 2, \ldots, m$, are computed according to equation :

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

The mean quantity of information carried out by the effector i is :

$$\bar{I}_i = -\sum_{j=1}^{m} x_{ij}\, p_j \log_2 p_j \quad , \quad \text{bits/effector} \ (2)$$

The mean quantity of information contained in the IRS (and, implicitly, in the corresponding effector series) is :

$$\bar{I} = -\sum_{j=1}^{m} p_j \log_2 p_j \quad , \quad \text{bits/IRS} \qquad (3)$$

Equation (3) establishes an hierarchical order of the IRS's (and, implicitly, of the corresponding effector series) : $(IRS)_a$ dominates $(IRS)_b$ (or, the series a dominates the series b) if $I_a > I_b$, or written more compactly $(IRS)_a \succ (IRS)_b \Leftrightarrow I_a > I_b$.

The maximum mean information content of the IRS is :

$$\bar{I}_{max} = \log_2 m \quad , \quad \text{bits} \qquad (4)$$

Accordingly, the derived quantity :

$$\Delta\bar{I} = \bar{I}_{max} - \bar{I} \quad , \quad \text{bits} \qquad (5)$$

is a measure of the degree of uncertainty associated with the IRS. The quantity $100\ \Delta\bar{I}/\bar{I}_{max}$ represents the percent of the obtainable information left unused due to the particular choice of the effector series (i.e., structures and number of points). The order relation induced by equation (5) is :

$$(IRS)_a \succ (IRS)_b \Leftrightarrow \Delta\bar{I}_a < \Delta\bar{I}_b$$

3. Applications

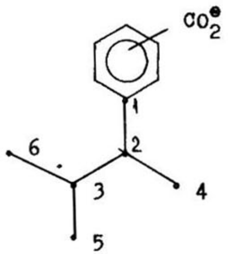
3.1. Hapten - Antibody Equilibria for Substituted Benzoic Acids.

The relative equilibrium constants (K_{rel}) for the combination of substituted benzoic acids in the o - , m - , and and - p - azobenzoate system[8a] (E - series of stereoisomers) are collected in Table 1.

Table 1. Effects of Substituents in the Attachement-Homologous Position on Combination in the E-*o*-, -*m*- , and -*p*- azobenzoate System.

Hapten	K_{rel}, $CO_2^{\ominus}$ in the position :		
	ortho	meta	para
H–C_6H_4–$CO_2^{\ominus}$	1.o	1.o	1.o
H_3C–	1.7	1.9	2.6
Cl–	2.9	3.0	2.8
Br–	2.5	4.8	5.1
I–	4.2	6.0	6.5
H_3CO–	3.3	3.0	3.8
O_2N–	1.1	1.4	2.8
$^{\ominus}O_2C$–	0.6	0.75	2.1
HO–	0.31	1.1	2.3
H_2N–	0.76	1.4	1.9
$H_3CC(O)HN$–	1.2	7.8	7.4

The IRS for the haptens displayed in Table 1 is shown in Figure 1.



<u>Figure 1</u>. IRS for the haptens collected in Table 1. (this IRS is denoted by $[IRS]_a$)

The occupancy matrix corresponding to $\left[IRS\right]_a$ is :

$$X = \left[X_{ij}\right] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{array}\right] \end{array}$$

The probability scheme associated with the $\left[IRS\right]_a$ is :

$$\begin{pmatrix} j \\ p_j \end{pmatrix}_{j=1,2,\ldots,6} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0.3793 & 0.3448 & 0.1379 & 0.0690 & 0.0345 & 0.0345 \end{pmatrix} \qquad (6)$$

The mean quantity of information contained in the $\left[IRS\right]_a$ is given by equation (3) as :

$$\bar{I}_a = -\sum_{j=1}^{6} p_j \log_2 p_j = 2.0557 \text{ , bits/IRS} \qquad (7)$$

The maximum mean information content of the $\left[IRS\right]_a$ is :

$$\bar{I}_{max,a} = \log_2 6 = 2.5849 \text{ , bits/IRS}$$

and the degree of uncertainty associated with the $\left[IRS\right]_a$ is computed as :

$$\Delta \bar{I}_a = \bar{I}_{max,a} - \bar{I}_a = 0.5292 \text{ , bits/IRS , and}$$

$$100\, \Delta\bar{I}_a / \bar{I}_{max,a} = 20.47\ \%$$

3.2. Hapten-Antibody Equilibria for Succinanilates and Succinamates.

The relative equilibrium constants (K_{rel}) for the combination of succinanilates and succinamates in the E - p - azosuccinanilates system[8b] are collected in Table 2.

Table 2. Closeness of Fit Around the Benzene Ding of E - p-azosuccinanilate Antibodies.

Hapten	K_{rel}
Succinanilate	1.00
Succinamate	0.04
N - Methyl succinamate	0.05
p - Nitrosuccinanilate	4.10
p - Bromosuccinanilate	2.20
p - Aminosuccinanilate	0.95
m - Bromosuccinanilate	1.80
o - Bromosuccinanilate	0.40
α - Naphthylsuccinamate	0.32
β - Naphthylsuccinamate	2.20

The IRS for the haptens displayed in Table 2 is shown in Figure 2.

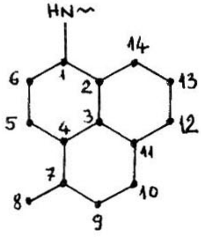


Figure 2. IRS for the haptens collected in Table 2 (this IRS is denoted by $[IRS]_b$).

The occupancy matrix corresponding to $[\mathrm{IRS}]_b$ is :

$\underline{\underline{X}} =$

	1	2	3	4	5	6	7	8	9	1o	11	12	13	14
1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	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
5	1	1	1	1	1	1	1	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	0	0	0	1	0	0	0
8	1	1	1	1	1	1	1	0	0	0	0	0	1	0
9	1	1	1	1	1	1	0	0	0	0	1	1	1	1
10	1	1	1	1	1	1	1	0	1	1	1	0	0	0

The probability scheme associated with $[\mathrm{IRS}]_b$ is :

$$\begin{pmatrix} j \\ p_j \end{pmatrix}_{j=1,2,\ldots,14} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.1364 & 0.1212 & 0.1212 & 0.1212 & 0.1212 \\ 6 & 7 & 8 & 9 & 10 \\ 0.1212 & 0.09091 & 0.0152 & 0.0303 & 0.152 \\ 11 & 12 & 13 & 14 & \\ 0.0454 & 0.0152 & 0.0152 & 0.0303 & \end{pmatrix}$$

The mean quantity of information contained in the $[\mathrm{IRS}]_b$ is given by equation :

$$\bar{I}_b = - \sum_{j=1}^{14} p_j \log_2 p_j = 3.4268 \quad , \quad \text{bits / IRS} \qquad (8)$$

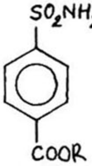
The maximum mean information content of the $[\mathrm{IRS}]_b$ is :

$$\bar{I}_{max.,b} = \log_2 14 = 3.8073 \quad , \quad \text{bits / IRS} ,$$

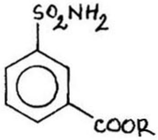
and the degree uncertainty associated with the $[\mathrm{IRS}]_b$ is computed as : $\Delta\bar{I}_b = \bar{I}_{max\ ,b} - \bar{I}_b = 0.3805$, bits / IRS , and $100. \Delta\bar{I}/\bar{I}_{max\ ,b} = 10\ \%$.

3.3. Affinity constants for sulfonamides

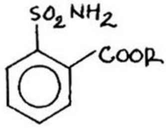
The affinity constants[9] of the sulfonamides I,II and III for carbonic anhydrase (AC, in mole^{-1}) are collected in Table 3



I



II



III

The corresponding IRS is shown in Figure 3 :

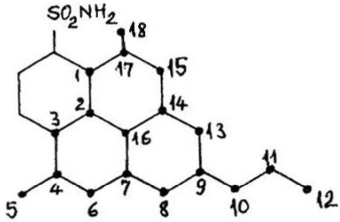


Figure 3. IRS for the sulfonamides collected in Table 3 (this IRS is denoted by $[IRS]_c$).

The matrix (9) is the occupancy matrix corresponding to IRS_c . The probability scheme associated with the IRS_c is :

$$\begin{pmatrix} j \\ p_j \end{pmatrix}_{j=1,\ldots,19} = \left(\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & & & \\ 0.1088 & 0.1088 & 0.1088 & 0.0408 & 0.0408 & 0.0408 & & & \\ 7 & 8 & 9 & 1o & 11 & 12 & 13 & 14 & 15 \\ 0.0748 & 0.0680 & 0.0748 & 0.0544 & 0.0340 & 0.01360 & 0.0272 & 0.0680 & 0.0340 \\ 16 & 17 & 18 & & & & & & \\ 0.0340 & 0.0340 & 0.0340 & & & & & & \end{array} \right)$$

The mean quantity of information, $\bar{I}$, contained in the $[IRS]_c$ is:

$\underline{\underline{X}}$ =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	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
4	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
7	1	1	1	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0
8	1	1	1	0	0	0	1	1	1	0	0	0	0	1	1	0	0	0
9	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	0	0	0
10	1	1	1	0	0	0	1	1	1	1	1	0	0	1	1	0	0	0
11	1	1	1	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0
12	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1
13	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1
14	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	1	1	1
15	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	1	1	1
16	1	1	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1

(9)

$$\bar{I}_c = - \sum_{j=1}^{18} p_j \log_2 p_j = 3.9802 \text{ , bits / IRS}$$

The maximum mean information content of the $[IRS]_c$ is :

$$\bar{I}_{max,c} = \log_2 18 = 4.1699 \text{ bits / IRS}$$

and the degree of uncertainty associated with the $[IRS]_c$ is computed according to equation :

$$\Delta\bar{I}_c = \bar{I}_{max,c} - \bar{I}_c = 0.1897 \text{ , bits / IRS , and } 100\,\Delta\bar{I}_c/\bar{I}_{max,c} = $$

$$= 4.55 \%$$

4. Conclusions

i) The paper has introduced and applied an informational method for ordering the effector series and/or IRS's : the series p dominates the series q, or/and the $[IRS]_p$ dominates the $[IRS]_q$ if $\bar{I}_p > \bar{I}_q$, or $\Delta\bar{I}_p < \Delta\bar{I}_q$, or $100\,\Delta\bar{I}_p/\bar{I}_{max.p} <$ $< 100\,\Delta\bar{I}_q/\bar{I}_{max,q}$.

The three IRS's here studied are ordered according to the above criteria as : $[IRS]_c \succ [IRS]_b \succ [IRS]_a$ (2.0557 < 3.4268 < 3.9802 ; 0.5292 > 0.38o5 > 0.1897 ; 20.47 > 10 > 4.55).

ii) The present approach of the of the information content of the investigated receptor space allows to develop an informational version of our Steric Difference method. This version, denoted by $\widetilde{SD}$, will be discussed in forthcoming papers of the series.

iii) The above ideas may be easily reformulated using Onicescu's information energy[10,11] , $E = \sum_{j=1}^{m} p_j^2$.

Table 3. Binding parameters of sulfonamides to carbonic anhydrase.

	No	R	log AC
I.	1.	Me	7.98
	2.	Et	8.50
	3.	n - Pr	8.77
	4.	n - Bu	9.11
	5.	n - Pent	9.39
	6.	n - Hex	9.39
II.	7.	Me	6.16
	8.	Et	6.21
	9.	n - Pr	6.44
	10.	n - Bu	6.95
	11.	n - Pent	6.86
III.	12.	Me	4.41
	13.	Et	4.8o
	14.	n - Pr	5.28
	15.	n - Bu	5.76
	16.	n - Pent	6.18

References

1. Previous part : O. Dragomir-Filimonescu, I.Motoc and I.Muscutariu, Math.Chem., in press.
2. I.Motoc, I.Muscutariu and D.Ciubotariu, Kexue Tongbao, 26, 361 (1981).
3. I.Motoc, Arzneim.Forsch./Drug.Res., 31, 290 (1981).
4. I.Motoc and O.Dragomir-Filimonescu, Math.Chem., in press.
5. I.Motoc, Math.Chem., 9 (1981), in press).
6. G.Shannon and W.Weaver, "Mathematical Theory of Communication", Univ. of Illinois Press,Urbana, 1949.
7. A.Mowshowitz, Bull.Math.Biophys., 3o, 175 (1968).
8. a)D.Pressman and A.L.Grossberg, "The Structural Basis of Antibody Specificity", Benjamin,New York,1968,p.54 ; b)idem , p.119 and 121.
9. R.W.King and A.S.V.Burgen,Proc.R.Soc.Lond.B,193,107(1976).
lo. O.Onicescu, C.R.Acad.Sci.Paris A, 263, 841 (1966).
11. O.Onicescu, St.cerc.mat., 18, 1419 (1966).