Molecular Shape Descriptors.
3. Steric Mapping of Biological Receptor

I. Motoc [a, b], G. R. Marshall [a], and J. Labanowski [a]

Z. Naturforsch. **40a,** 1121 – 1127 (1985); received May 25, 1985

The paper presents and illustrates and improved version of the SIBIS algorithm. SIBIS is a self-consistent type algorithm which is well suited to delineate the main steric features of biological receptor, identify missing relevant atoms and/or presence of superfluous atoms in the molecules studied, and to provide an approximate estimate for the receptor bulk tolerance. SIBIS calculations for a set of substituted benzoates in the anti-p-(p′-azophenylazo)benzoate system yield an excellent QSAR ($n = 18$, $r = 0.989$, $s = 0.192$, $F = 208.16$) which is further supported by the information correlation coefficients, R (R's express quantitatively the pairwise degree of the molecular shape relatedness); the optimal receptor steric map is in agreement with experimental evidences.

1. Introduction

The three-dimensional molecular shape descriptors (3D-MSD) discussed in the previous papers [1, 2] introduce a topographical basis into quantitative structure activity-activity relationships (QSAR's) and offer a way to approach the analysis of the conformational variable [3] in the conceptual framework of QSAR.

The most active compound of the set under study contains the most information, in a complementary sense, about the receptor and, therefore, represents the logical candidate for reference structure. Note, however, that this does not necessarily imply that it provides the best topographical fit to the receptor active site: one should consider the possibility of missing sterically relevant atoms and/or presence of sterically superfluous (irrelevant) atoms. The problem is to determine, out of the bioactivity of the molecules available, a better topography of the active site.

A classification methodology, the SIBIS algorithm [4], has been developed to map the receptor space explored by the molecules under study, i.e., to identify areas which correspond to the receptor active site, areas occupied by the receptor, and, respectively, areas which, pointing away from active site, offer little opportunity for interaction with the receptor, being, therefore, steric irrelevant; further, an approximate estimate for the receptor bulk tolerance is also provided.

SIBIS is a self-consistent type algorithm based on the least squares method (with, or without subsidiary conditions); the convergence criterium is the best overall agreement between the observed (Y_i) and estimated ($\hat{Y}_i$) bioactivities. The new features which have been incorporated in SIBIS are described in the next section.

It is worth noting that SIBIS approach provides a way to coanalyse structurally diverse compounds and the results may transcend the chemical series on which they were based, namely, they may indicate novel lead structures.

2. The SIBIS Algorithm

Consider the molecules M_i, $1 \leqq i \leqq n$, which exhibit their biological activity *via* a common mode of action, and let M_r be the most active compound of the set. To obtain a framework on which to base a computational effort seeking the optimization of the reference structure, M_r, one must develop an appropriate basis from which to describe numerically the stereochemistry of the molecules considered, and a method to estimate the overlapping and non-overlapping van der Waals volumes of reference *vs.* compared molecules. Towards this end, one proceeds by superimposing the n molecules over M_r, the superposition procedure being the one

---

[a] Department of Physiology and Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA.
[b] Department of Computer Science, Washington University, St. Louis, MO 63130, USA.

Reprint request to Dr. Ioan Motoc, Department of Cell Biology and Physiology, Box 8101, Washington University, School of Medicine, St. Louis, MO 63130, USA.

0340-4811 / 85 / 1100-1121 $ 01.30/0. – Please order a reprint rather than making your own copy.

described in [2], supplemented with an allowance for contracting into a single vertex all nearby atoms, i.e., the atoms $p \in M_i$ and $q \in M_r$ will occupy the same vertex if the Euclidean distance $d(p, q) \leqq \delta$, where δ is given; in order to simplify the problem, the hydrogen atoms are neglected. Geometrical congruences can be performed accurately by using an appropriate molecular graphics package, e.g. [5]. The obtained pattern of vertices mimicks the essential topographical features of the receptor space explored (investigated) by the n molecules; it is called [4, 6] the investigated receptor space, abbreviated by IRS.

One can use the IRS as coordinate system and ascribe to each M_i the m-dimensional row vector $\boldsymbol{X}_i = [X_{ij}]$, $j = 1 \div m$; here, m denotes the number of IRS vertices, and $X_{ij} = 1$ if the vertex j is occupied by a non-hydrogen atom of M_i, and $X_{ij} = 0$ if it is empty.

Next, one derives the initial steric map, $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$, of the receptor as follows:

(i) Consider the additional possible connectivities in the original IRS by connecting vertices p and q if the resultant edge may represent a covalent bond.

(ii) Partition the IRS vertices into three classes: vertices assigned to the receptor active site (c), to the receptor backbone (w), and steric-irrelevant areas (i).

The c-type vertices correspond to the non-hydrogen atoms of the reference structure; the other vertices are of w-type. For convenience one introduces an additional vertex of i-type, empty in the molecules considered, and connects it with those IRS vertices for which one wishes to check the steric relevance; obviously, the virtual i-type vertex does not change the information content of the original IRS.

In the framework established above, the overlapping volume of M_r over M_i (denoted by SMD (r, i)) and the non-overlapping volume of M_i over M_r (SMD (i, r)) are approximated [4, 7] by the following steric molecular descriptors (SMD's):

$$\mathrm{SMD}\,(r, i) = \sum_{j \in \{\text{c-type}\}} S_{ij} X_{ij}\,, \qquad (1\,\mathrm{a})$$

$$\mathrm{SMD}\,(i, r) = \sum_{j \in \{\text{w-type}\}} S_{ij} X_{ij}\,, \qquad (1\,\mathrm{b})$$

where S_{ij} is an additive measure of the size of the atom j belonging to the molecule i, e.g., van der Waals fragmental volume [8], k-values calculated in [9], molar refractivity [10], etc., and $X_{ij} \in \boldsymbol{X}_i$.

The SIBIS algorithm consists of the following steps:

1) Consider the steric map $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$ and compute the regression equations (2) or (3) and the corresponding correlation coefficients, r_0:

$$\hat{Y}_i = \begin{cases} \beta_0 + \xi\,\mathrm{SMD}\,(r, i) - \tau\,\mathrm{SMD}\,(i, r) \\ + F(\sigma_1, \sigma_2, \ldots)\,, & (2\,\mathrm{a}) \\ \beta_0 + \xi\,\mathrm{SMD}\,(r, i) - \tau\,\mathrm{SMD}\,(i, r) \\ - \varphi[\mathrm{SMD}\,(r, i)]^2 + F(\sigma_1, \sigma_2, \ldots)\,, & (2\,\mathrm{b}) \end{cases}$$

$$\hat{Y}_i = \begin{cases} \xi\,\mathrm{SMD}\,(r, i) - \tau\,\mathrm{SMD}\,(i, r) \\ + F(\sigma_1, \sigma_2, \ldots)\,, & (3\,\mathrm{a}) \\ \xi\,\mathrm{SMD}\,(r, i) - \tau\,\mathrm{SMD}\,(i, r) \\ - \varphi[\mathrm{SMD}\,(r, i)]^2 + F(s, \sigma, \ldots)\,. & (3\,\mathrm{b}) \end{cases}$$

Here, ξ and τ, or ξ, τ and φ are the corresponding susceptibilities to the molecular shape; F is (usually) a linear function of non-shape factors (expressed quantitatively by $\sigma_1, \sigma_2, \ldots$) which may condition bioactivity. F sorts out the non-shape component of the biological response in order to prevent the contamination of the SMD's due to iterative correction of the SMD (r, i) and SMD (i, r) – see steps 2 and 3 below; this sorting out avoids the respective global coding of beneficial factors by SMD (r, i) and detrimental ones by SMD (i, r).

The quadratic models (2 b) or (3 b) provide an approximate estimate for the receptor bulk tolerance SMD $(r, i)_{\max} = \xi/2\varphi$, i.e., $\partial \hat{Y}/\partial \mathrm{MSD}\,(r, i) = 0$.

Further, the "no-intercept" regression models (3) offer an alternative to models (2) in situations where the intercept β_0 is not significant i.e., one accepts the null hypothesis $\beta_0 = 0$.

2) Change the classification of the vertex $j \in \langle \mathrm{IRS} \rangle_{\mathrm{init}}$ (i.e., c to w or i; w to c or i) if and only if the following two conditions hold:

(i) the resultant equations (2) or (3) have a better correlation coefficient $r_1 \geqq r_0 + \Delta r$, where Δr is the desired improvement of r_0; and

(ii) the set of c-type vertices are all connected among themselves. These changes of vertex classification are performed until no further improvement is observed.

The conditions (i) and (ii) are designed to identify atoms which are sterically relevant, and respectively, to be consistent with the concept of reference structure: the c-type vertices correspond to the non-hydrogen atoms making up the reference structure and, obviously, M_r is represented by a connected graph. The graph consisting of the i-type vertices may be also preserved and used to define the "entrance" of the active site.

Note that by ignoring condition (ii) above and $F(\sigma_1, \sigma_2, \ldots)$ in (2) or (3), SIBIS can function as a bi-parametric variant of Free-Wilson type [11, 12].

(3) The resultant $\langle \mathrm{IRS} \rangle$ is then considered as $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$ and the step 2 is carried out for all vertices j.

(4) Repeat steps 2 and 3 until self-consistency is achieved, i.e., until the vertex classification no longer changes within the given tolerance Δr on repeated iteration. The resultant $\langle \mathrm{IRS} \rangle$ is optimal, and the computing procedure is terminated.

The set of c-type vertices of $\langle \mathrm{IRS} \rangle_{\mathrm{opt}}$ represents the "best" probable molecular shape complementary to the receptor active site and it may be used to improve the reference structure in calculating the 3D-MSD parameters [1, 2]. Extension of SIBIS approach in combination with 3D-MSD parameters is currently under development.

3. Application: Hapten-Antibody Interactions

We illustrate SIBIS by calculating the optimal steric map of the antibody active site involved in the interaction of substituted benzoates in the anti-p-(p′-azophenylazo)benzoate system. This system is a logical candidate for illustrating SIBIS because there are clear evidences that antibodies are relatively rigid macromolecules [13], and the antibody active site is a structurated cavity which "copies" the shape of the antigenic part of the immunogen [14]; hence, the steric effects would play an important role in this case. Further, this example has little conformational flexibility to cloud the issues.

The investigated receptor space (IRS) shown in Fig. 1A was obtained with p-(p′-azophenylazo)-benzoate moiety as reference structure and the carboxylate group as pharmacophore. By way of illustration, the $\boldsymbol{X}$ vectors ascribed to the haptens no. 1, 12, 13, and 17 in Table 1 are given below:

$$\boldsymbol{X}_1 = [1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$$
$$\boldsymbol{X}_{12} = [1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$$
$$\boldsymbol{X}_{13} = [1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0]$$
$$\boldsymbol{X}_{17} = [1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 0]$$

The initial steric map, $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$, was obtained by considering additional possible connectivities in the original IRS, and introducing the virtual vertex no. 23 connected with the vertices no. 21 and 22 (Fig. 1B) for which checking of steric relevance is legitimate [14].

The relative equilibrium constants, K_{rel}, for the eighteen benzoates considered are taken from [14].

Table 1. Benzoate haptens: observed and calculated equilibrium constants, K_{rel}.

No.	Hapten	q_4	log K_{rel} obs.	Eq. (5)	Eq. (6)	Eq. (7)
1	Benzoate (B)	−0.0554	0.0000	−0.1082	0.1099	0.2622
2	3-Cl−B	−0.0191	−0.3665	−0.5013	−0.5713	−0.4475
3	3-NO_2−B	−0.0034	−0.9208	−1.0101	−0.8302	−0.7345
4	3-NO_2, 4-Cl−B	−0.1794	−0.3872	−0.3262	−0.3221	−0.5111
5	3-NO_2, 5-Cl−B	0.0158	−1.1549	−1.3675	−1.4832	−1.4442
6	3,5-diCl−B	0.0005	−1.6198	−0.8596	−1.2249	−1.1572
7	4-F−B	−0.2825	0.5563	0.5770	0.6299	0.4111
8	4-Cl−B	−0.2414	0.7324	0.5966	0.6346	0.4856
9	4-Me−B	−0.0570	0.2553	0.1587	0.2938	0.4483
10	4-NH_2−B	−0.1084	0.3222	0.2661	0.3787	0.4483
11	4-OH−B	−0.1488	0.6721	0.3504	0.4454	0.4483
12	4-NO_2−B	−0.2600	0.2553	0.5819	0.2925	0.0431
13	4-(phenylazo)-benzoate (PAB)	−0.1622	1.8261	2.2236	1.7357	1.7552
14	4′-NH_2-PAB	−0.1666	1.8451	1.9683	1.9242	1.9378
15	4′-Me-PAB	−0.1631	1.8129	1.9610	1.9184	1.9378
16	4′-OH-PAB	−0.1645	2.0453	1.9640	1.9207	1.9378
17	2′-Me,4′-OH-PAB	−0.1675	1.9085	1.7056	2.1067	2.1290
18	3′-Me,4′-OH-PAB	−0.1643	2.0960	1.6991	1.9204	1.9378

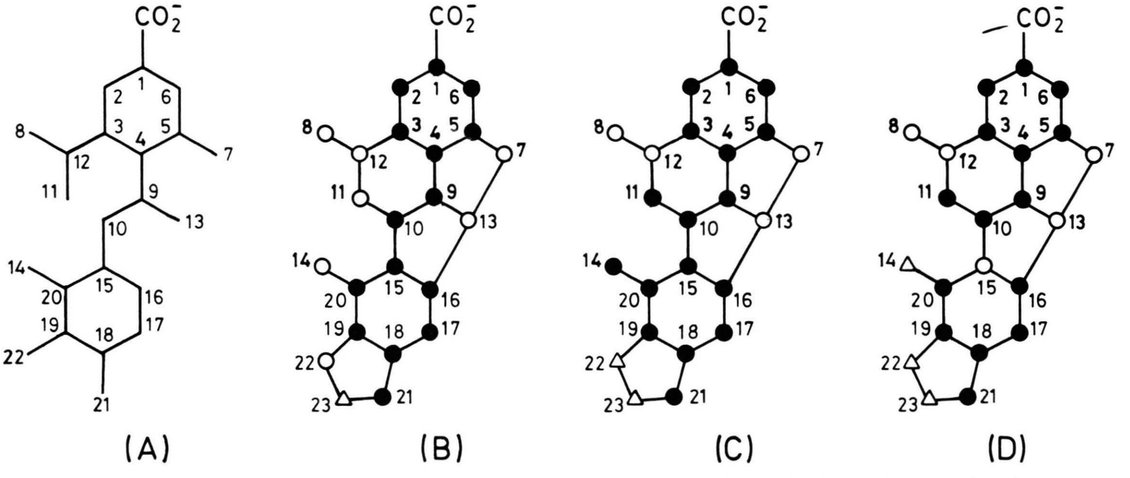

Fig. 1. Hapten-antibody interactions: (A) Investigated receptor space, IRS; (B) Initial steric map, $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$; (C) Optimal steric map, $\langle \mathrm{IRS} \rangle_{\mathrm{opt}}$; (D) Optimal map corresponding to Free-Wilson-type equation (8). (Full circles mark the cavity vertices, the open circles the wall vertices, and the triangle marks the irrelevant vertices.)

The shape descriptors $\mathrm{SMD}(i) = [\mathrm{SMD}(r, i), \mathrm{SMD}(i, r)]$ have been calculated according to (1), using the k_{ij} values [9] as S_{ij}.

The haptens collected in Table 1 have been previously investigated [10] by CNDO/2 calculations with the goal of identifying electronic factors which may condition the K'_{rel}s. It was found that only the charge on the substituent in position 4 (for the compounds no. 13–18 the diazo group was considered as substituent in position 4), q_4, is significant,

$$\log K_{\mathrm{rel}} = -7.9092\,(\pm 2.5526)\,q_4 - 0.4737\,(\pm 0.3997) \quad (4)$$

with $r = 0.612$, $s = 0.957$, $F = 9.60$. Further, the hapten hydrophobicity (as expressed by π, [15], or f, [16]) has no significant linear or parabolic relationship with $\log K_{\mathrm{rel}}$. These results suggest a linear function of q_4 as the appropriate functional form of $F(\sigma_1, \sigma_2, \ldots)$ in (2) or (3).

With the choice $F(\sigma_1, \sigma_2, \ldots) = a + b q_4$ in (2a), the following SIBIS equations have been derived:

$$\begin{aligned} \log K_{\mathrm{rel}} = {} & 0.2636\,(\pm 0.0226)\,\mathrm{SMD}(r, i) \\ & (11.64) \\ & - 0.2644\,(\pm 0.0646)\,\mathrm{SMD}(i, r) \\ & (-4.10) \\ & - 2.0886\,(\pm 0.9973)\,q_4 \\ & (-2.09) \\ & - 1.8055\,(\pm 0.2334) \\ & (-7.74) \end{aligned} \quad (5)$$

with $r = 0.972$, $s = 0.306$, $F = 79.034$, and the intercorrelation matrix (5a),

	$\log K_{\mathrm{rel}}$	SMD (r, i)	SMD (i, r)	q_4
$\log K_{\mathrm{rel}}$	1.000			
SMD (r, i)	0.894	1.000		
SMD (i, r)	−0.430	−0.082	1.000	
q_4	−0.612	−0.397	0.449	1.000

and

$$\begin{aligned} \log K_{\mathrm{rel}} = {} & 0.1812\,(\pm 0.0132)\,\mathrm{SMD}(r, i) \\ & (13.75) \\ & - 0.5177\,(\pm 0.0568)\,\mathrm{SMD}(i, r) \\ & (-9.12) \\ & - 1.6516\,(\pm 0.6302)\,q_4 \\ & (-2.62) \\ & - 1.0686\,(\pm 0.1705) \\ & (-6.27) \end{aligned} \quad (6)$$

with $r = 0.989$, $s = 0.192$, $F = 208.16$, and the intercorrelation matrix (6a),

	$\log K_{\mathrm{rel}}$	SMD (r, i)	SMD (i, r)	q_4
$\log K_{\mathrm{rel}}$	1.000			
SMD (r, i)	0.862	1.000		
SMD (i, r)	−0.805	−0.440	1.000	
q_4	−0.612	−0.359	0.573	1.000

Equation (5) corresponds to $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$ and (6), calculated with $\Delta r = 0.001$, to the optimal steric

Table 2. Benzoate haptens: The information intercorrelation matrix.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.000					C						B						A
2	0.898	1.000																
3	0.816	0.880	1.000															
4	0.758	0.817	0.928	1.000														
5	0.758	0.817	0.928	0.862	1.000													
6	0.822	0.915	0.805	0.748	0.897	1.000												
7	0.951	0.854	0.776	0.836	0.721	0.781	1.000											
8	0.898	0.806	0.733	0.844	0.681	0.738	0.990	1.000										
9	0.926	0.831	0.756	0.842	0.702	0.761	0.997	0.998	1.000									
10	0.926	0.831	0.756	0.842	0.702	0.761	0.997	0.998	1.000	1.000								
11	0.926	0.831	0.756	0.842	0.702	0.761	0.997	0.998	1.000	1.000	1.000							
12	0.816	0.733	0.667	0.743	0.619	0.761	0.880	0.880	0.882	0.882	0.882	1.000						
13	0.655	0.588	0.535	0.596	0.496	0.538	0.705	0.705	0.707	0.707	0.707	0.713	1.000					
14	0.632	0.568	0.516	0.575	0.479	0.520	0.681	0.682	0.683	0.683	0.683	0.689	0.966	1.000				
15	0.632	0.568	0.516	0.575	0.479	0.520	0.681	0.682	0.683	0.683	0.683	0.689	0.966	1.000	1.000			
16	0.632	0.568	0.516	0.575	0.479	0.520	0.681	0.682	0.683	0.683	0.683	0.689	0.966	1.000	1.000	1.000		
17	0.612	0.550	0.500	0.557	0.464	0.503	0.660	0.660	0.661	0.661	0.661	0.667	0.935	0.968	0.968	0.968	1.000	
18	0.612	0.550	0.500	0.557	0.464	0.503	0.660	0.660	0.661	0.661	0.661	0.667	0.935	0.968	0.968	0.968	0.938	1.000

map shown in Figure 1 C. Here, r denotes multiple correlation coefficient, s is standard deviation, F is the test for the significance of the correlation, and the quantities in parentheses below the coefficients in (5) and (6) are Student t-values.

The estimate for the receptor bulk tolerance, obtained using SIBIS with (2 b), is $\mathrm{SMD}(r, i)_{\max} \cong 0.5015/0.0292 \cong 17.2$; although in good agreement with $\langle \mathrm{IRS} \rangle_{\mathrm{opt}}$, the estimate is rather unreliable due to significant collinearity between SMD (r, i) and $[\mathrm{SMD}(r, i)]^2$.

Comparison of $\langle \mathrm{IRS} \rangle_{\mathrm{init}}$ and $\langle \mathrm{IRS} \rangle_{\mathrm{opt}}$ indicates that the antibody active site studied here consists of two regions characterized by distinct steric features: the deeper part of the cavity, corresponding to the benzene moiety bearing the carboxylate group is rigid and fits tightly the benzene ring. The outer part of the cavity is flexible and can easily adjust its size and shape.

Equation (6) represents an excellent correlation, which accounts for 97.8% of the variability in the K_{rel} values. Further, molecular shape, as expressed by SMD's, plays the dominant role in determining the equilibrium constants, as argued by recomputing (6) without q_4,

$$\begin{aligned}\log K_{\mathrm{rel}} = {} & 0.1862\,(\pm 0.0154)\ \mathrm{SMD}\,(r, i) \\ & (12.10) \\ & - 0.5914\,(\pm 0.0582)\ \mathrm{SMD}\,(i, r) \\ & (-10.16) \\ & - 0.8549\,(\pm 0.1766)\,, \\ & (-4.84)\end{aligned} \tag{7}$$

with $r = 0.984$, $s = 0.226$, $F = 221.99$.

The q_4 term in (6) gives a low (i.e., the additional variability accounted for by q_4 is only 1%) but still significant contribution (see (4)); it models a monopol-monopol interaction and its negative regression coefficient indicates the presence of a net positive charge on the receptor wall corresponding approximately to the vertex 9 of IRS.

Use of SIBIS as a bi-parametric variant of Free-Wilson type (see discussion following point 2 of SIBIS algorithm) yields the following equation

$$\begin{aligned}\log K_{\mathrm{rel}} = {} & 0.3014\,(\pm 0.0140)\ \mathrm{SMD}\,(r, i) \\ & (21.55) \\ & - 0.6021\,(\pm 0.0509)\ \mathrm{SMD}\,(i, r) \\ & (-11.83) \\ & - 1.6386\,(\pm 0.1452)\,, \\ & (-11.28)\end{aligned} \tag{8}$$

with $r = 0.988$, $s = 0.196$, $F = 297.38$, and the intercorrelation coefficient r (SMD(r, i), SMD(i, r)) $= 0.019$. The optimal map corresponding to (8) (Fig. 1D) resemble closely the optimal steric map shown in Fig. 1C: the significance of the different classification of the vertices no. 14 and 15 is marginal as they bring about an improvement in r value amounting only 0.004. Comparison of (6) and (8) indicates, as expected, that (8) attempts to account for the missing regressor q_4 (which has a beneficial effect for all haptens but no. 5 and 6, [10]) by increasing the weight of SMD(r, i) relatively to SMD(i, r), i.e., the ratio $\xi:\tau$ increases from 0.18:0.52 in (5), to 0.30:0.60 in (8).

The information correlation coefficient, $R(\alpha, \beta)$, which quantitatively compare the degree of relatedness of the shape of bioactive molecules α and β, [6], offers itself for testing (6).

Note that one may associate with each M_i, *via* the $\boldsymbol{X}_i$ vectors, the finite probability scheme $\boldsymbol{P}_i$:

$$\boldsymbol{P}_i = (P_{ij}), \quad j = 1, 2, \ldots, m;$$

$$P_{ij} = S_{ij} X_{ij} / \sum_{j=1}^{m} S_{ij} X_{ij},$$

$$\sum_{j=1}^{m} P_{ij} = 1, \quad \text{and} \quad 0 \leqq P_{ij} \leqq 1. \tag{9}$$

P_{ij} is the probability that M_i occupies the receptor space centered around the vertex j.

The quantity $E(i)$ is called the information energy content of $\boldsymbol{P}_i$,

$$E(i) = \sum_{j=1}^{m} P_{ij}^2; \quad 1/m \leqq E(i) \leqq 1 \tag{10}$$

and it is a measure of the uniformity of the system described by P_i [17].

Because P_i is related to the shape of the molecule M_i, the degree of relatedness of the probability schemes P_α and P_β will characterize the degree of shape relatedness of the molecules M_α and M_β, provided that the vectors $\boldsymbol{X}_\alpha$ and $\boldsymbol{X}_\beta$ were defined using the same IRS as coordinate system. The information correlation coefficient $R(\alpha, \beta)$, given by

$$R(\alpha, \beta) = \sum_{j=1}^{m} P_{\alpha j} P_{\beta j} / [E(\alpha) \cdot E(\beta)]^{1/2}; \tag{11}$$

$$P_{\alpha j} \in \boldsymbol{P}_\alpha, \quad P_{\beta j} \in \boldsymbol{P}_\beta,$$

expresses quantitatively the relationship between $\boldsymbol{P}_\alpha$ and $\boldsymbol{P}_\beta$, [17], and, accordingly, the relationship between the shapes of the molecules M_α and M_β.

The information correlation coefficient is $R(\alpha, \beta) = 1$ if $\boldsymbol{P}_\alpha$ and $\boldsymbol{P}_\beta$ are identical repartized (i.e., M_α and M_β have the same shape), and $R(\alpha, \beta) = 0$ if $\boldsymbol{P}_\alpha$ and $\boldsymbol{P}_\beta$ are indifferent (i.e., the shape of M_α and M_β are not related); the intermediate values $0 < R(\alpha, \beta) < 1$ are judged using the criteria for the significance of the correlation coefficient r, [18, 19].

The mutual information correlation coefficients among the compounds in Table 1 are collected in Table 2. Because the intercorrelation matrix is symmetrical, only the lower triangular part is displayed.

The information correlation coefficients in Table 2 corroborate the SIBIS results: (i) Compounds with similar shape exhibit similar relative equilibrium constants, provided the variation of q_4 is negligible, e.g., block A in Table 2. (ii) The closely bunched R values for compounds no. 7–11 (block B in Table 2) indicate that the variation of molecular shape within this set of molecules is insignificant and the observed variation of $K'_{\text{rel}} S$ is brought about by q_4 (see also the last two columns in Table 1), in agreement with the SIBIS calculations. (iii) Compounds no. 2–6 (block C in Table 2) show a rather poor correlation between $\log K_{\text{rel}}$ and R values calculated for the whole molecule; this indicates that the shape of different parts of molecule may have different effects (i.e., beneficial *vs* detrimental, beneficial *vs* irrelevant, etc.) on hapten-antibody association. In order to test this assumption, we have calculated the information correlation coefficients between the receptor active site (RAC) and, respectively, receptor backbone (RB) – as they are depicted by the $\langle \text{IRS} \rangle_{\text{opt}}$ shown in Fig. 1C – and the haptens of Table 1.

M_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$R(\text{RAC}, M_i)$	0.612	0.550	0.583	0.634	0.542	0.503	0.660	0.660	0.661	0.661	0.661	0.667	0.935	0.904	0.904	0.904	0.938	0.875
$R(\text{RB}, M_i)$	0.000	0.220	0.333	0.309	0.495	0.403	0.000	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000

(12)

Obviously, if conditions (13) hold,

$$R(\mathrm{RAC}, M_i) \leqq R(\mathrm{RAC}, M_j) \quad \text{and} \qquad R(\mathrm{RB}, M_i) \geqq R(\mathrm{RB}, M_j)\,, \tag{13}$$

the molecules M_i and M_j are comparable, and $\log K_{\mathrm{rel}}(i) \leqq \log K_{\mathrm{rel}}\ (j)$. Inspection of (12) shows that conditions (13), which, although different, are qualitatively related to the SMD, satisfactorily reproduce the experimentally ordering of $\log K'_{\mathrm{rel}}s$, e.g., $R(\mathrm{RAC}, M_1) > R(\mathrm{RAC}, M_i)$, $R(\mathrm{RB}, M_1) < R(\mathrm{RB}, M_i)$, and $\log K_{\mathrm{rel}}(1) > \log K_{\mathrm{rel}}(i)$, $i = 2 \div 6$; $R(\mathrm{RAC}, M_4) > R(\mathrm{RAC}, M_j)$, $R(\mathrm{RB}, M_4) < R(\mathrm{RB}, M_j)$, and $\log K_{\mathrm{rel}}(4) > \log K_{\mathrm{rel}}(j), j = 3, 5, 6$; etc.

It is concluded from the results presented above that a SIBIS type approach does provide a reliable receptor steric map which is useful in rationalizing bioactivity and predicting new bioactive compounds.

Acknowledgement

This work was supported by NIH grant GM24483.

[1] I. Motoc, G. R. Marshall, R. A. Dammkoehler, and J. Labanowski, Z. Naturforsch. **40a,** 1108 (1985).
[2] I. Motoc and G. R. Marshall, Z. Naturforsch. **40a,** 1114 (1985).
[3] G. R. Marshall and I. Motoc, Approaches to the Conformation of the Drug Bound to the Receptors, Top. Mol. Pharm. **3** (1985), in press.
[4] I. Motoc, Quant. Struct. Act. Relat. **3,** 43 (1984).
[5] SYBYL Software Tripos Assoc., Inc., St. Louis, MO, USA.
[6] I. Motoc, Z. Naturforsch. **38a,** 1342 (1983).
[7] I. Motoc, Arzneim.-Forsch./Drug Res. **31,** 290 (1981).
[8] I. Motoc and G. R. Marshall, Chem. Phys. Lett. **116,** 415 (1985).
[9] V. Austel, E. Kutter, and W. Kalbfleisch, Arzneim.-Forsch./Drug Res. **29,** 585 (1979).
[10] I. Motoc, Quant. Struct.-Act. Relat. **3,** 47 (1984).
[11] W. Purcell, G. E. Bass, and J. M. Clayton, Strategy of Drug Design: A Guide to Biological Activity, Wiley, New York 1973.
[12] C. Hansch, C. Silipo, and E. E. Steller, J. Pharm. Sci. **64,** 1186 (1975).
[13] R. J. P. Williams, Angew. Chem. Int. Ed. Engl. **16,** 766 (1977).
[14] D. Pressman and A. L. Grossberg, Structural Basis of Antibody Specificity, Benjamin, New York 1968.
[15] C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York 1979.
[16] R. F. Rekker, The Hydrophobic Fragmental Constant, Elsevier, Amsterdam 1977.
[17] O. Onicescu, C. R. Acad. Sci. Paris **A263,** 841 (1966); St. Cerc. Mat. **18,** 1419 (1966).
[18] P. R. Wells, Linear Free Energy Relationships, Academic Press, London 1968.
[19] J. G. Topliss and R. J. Costello, J. Med. Chem. **15,** 1066 (1972).