Neural Nets for the Simulation of Molecular Recognition within MS-Windows Environment

Jarosław Polański

Institute of Chemistry, University of Silesia, Szkolna 9, 40-006 Katowice, Poland

Received September 14, 1995⊗

A neural method for the comparison of the molecule series, using a self-organizing map (SOM) procedure, is presented. The coordinates of individual atoms within the most active molecule mimicking the geometry of the receptor are trained with SOM algorithms to form a template which provides a basis for recognizing the similarity within the compounds series. Therefore the SOM matrices produced for individual molecules, if analyzed further by neural procedures or by the calculation of the matrix determinants, provide reasonable classifications according to the molecule's structure or actual activity observed, provided that this is limited by shape factors. All computational operations are performed within MS-WINDOWS environment, using commercial programs of HYPERCHEM and MATLAB.

INTRODUCTION

A molecule is a primary objective of any chemical study, the basic and the most substantial unit that determines the nature and behavior of any substance found in the real world. In particular, molecular recognition processes determine the response evoked within any biological receptor. Most often, however, a complicity of the interactions does not allow for the understanding of the phenomena occurring, and usually it is similarities or common motifs within drug molecules (pharmacophores) that imply the stimulation of the same receptor site. As a consequence, many approaches offering a comparison of molecules have been inspired. Probably the most interesting ones among them raise a possibility of a direct shape comparison by overlaying an individual test molecule on the reference, while observing and describing the degree of their match. The latter started from simple two-dimensional templates,[1] which resemble methods of the graph theory and gradually developed to a three-dimensional description offered by Molecular Shape Analysis (MSA)[2,3] and further methods, in particular by complex Comparative Molecular Field Analysis (CoMFA) approach,[4] aimed at the comparison of molecular surfaces and surfaces' properties, e.g., the electrostatic potential.[5]

Despite all of that, the comparison still poses problems because no single, general, and obvious numerical model of the molecule can be proposed, and often it is molecules of the very different size and character that must be compared as the ones stimulating the same receptor. Moreover most often the molecular system under study is only one among many conformations possible, and there is no evidence that it is any minimal energy state that makes an active conformer.

SELF-ORGANIZING MAPS

Self-organizing maps (SOM) were designed to yield the nets preserving the topology, while reducing the dimensionality of the input objects.[6] A neural approach, using the Kohonen mapping[6] of the molecules, has been proposed recently to project the electrostatic potential from van der Waals surface into two-dimensional maps.[7,8] It has been found that molecular patterns obtained can be used to visualize interactions of some active compounds with muscarinic and nicotinic receptors.[7,8] That technique can also be used for a comparison of the molecule's shape.[9] This consists in preparing the reference (template or comparative) net for the most active analog among the series, which simulates the geometry of the receptor. The method offers a unique possibility for the direct comparison of the molecular surfaces described by a large number of coordinates, preserving their original shape and electronic features, and such a situation closely imitates real molecular recognition processes. Technically it means thousands of points taken from van der Waals surface to be processed by the net to produce SOM pattern. One should, however, realize that what is being done by self-organizing maps, is associating the signals of similar inputs into common outputs; therefore, substantial averaging of inputs are taking place. As a consequence, intuitively it is atoms (coordinates) that make natural attractors for all points located on theirs van der Waals surfaces. From the practical point of view, a molecule could be represented by a relatively small amount of data.

The current approach is aimed at the use of atom coordinates (and charges) for feeding SOM to prepare two-dimensional plots of the molecules that would be capable of identifying similar molecular objects, therefore simulating molecular recognition processes.

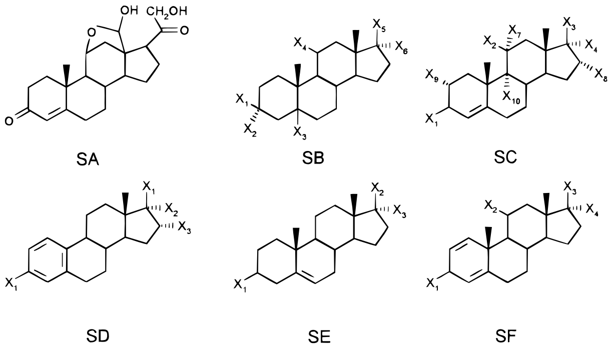
PROCEDURES

Model Building. The structures of all compounds within the series were prepared and used as obtained after modeling by HYPERCHEM 4.0. The compounds under analyses were specified in Table 1 and 2. These form a few series which were extensively studied previously,[4,5,9,10] which enables a comparison of the net performances.

Neural procedures were programmed using MATLAB (MS-WINDOWS) environment. The basic script files were supplemented by some additional ones which give the possibility of generating van der Waals surfaces from the coordinates of the molecules' atoms and the simulation of the toroidal neighborhood for the SOMs. Figure 1 shows

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1996.

Table 1. Structure and CBG/TBG Affinity Data for Steroids Series of the SA−SE Structures[a]



no.	structure	X$_1$	X$_2$	X$_3$	X$_4$	X$_5$	X$_6$	X$_7$	X$_8$	X$_9$	X$_{10}$	CBG log 1/K	TBG log 1/K
1	SA											−6.279	−5.322
2	SB	OH	H	H[b]	H	OH	H					−5.000	−9.114
3	SE	OH	OH	H								−5.000	−9.176
4	SC	=O	H	=O				H	H	H	H	−5.763	−7.462
5	SB	H	OH	H[b]	H	=O						−5.613	−7.146
6	SC	=O	OH	COCH$_2$OH	H			H	H	H	H	−7.881	−6.342
7	SC	=O	OH	COCH$_2$OH	OH			H	H	H	H	−7.881	−6.204
8	SC	=O	=O	COCH$_2$OH	OH			H	H	H	H	−6.892	−6.431
9	SE	OH	=O									−5.000	−7.819
10	SC	=O	H	COCH$_2$OH	H			H	H	H	H	−7.653	−7.380
11	SC	=O	H	COCH$_2$OH	OH			H	H	H	H	−7.881	−7.204
12	SB	=O		H[b]	H	OH	H					−5.919	−9.740
13	SD	OH	H	H								−5.000	−8.833
14	SD	OH	H	OH								−5.000	−6.633
15	SD	OH	=O[g]	H								−5.000	−8.176
16	SB	H	OH	H[c]	H	=O						−5.225	−6.146
17	SE	OH	COMe	H								−5.225	−7.146
18	SE	OH	COMe	OH								−5.000	−6.362
19	SC	=O	H	COMe	H			H	H	H	H	−7.380	−6.944
20	SC	=O	H	COMe	OH			H	H	H	H	−7.740	−6.996
21	SC	=O[d]	H	OH	H			H	H	H	H	−6.724	−9.204
22	SF	=O	OH	COCH$_2$OH	OH			H	H	H		−7.512	?[e]
23	SC	=O	OH	COCH$_2$OCOMe	OH			H	H	H	H	−7.553	?[e]
24	SC	=O	=O	COMe	H			H	H	H	H	−6.779	?[e]
25	SC	=O	H	COCH$_2$OH	H			OH	H	H	H	−7.200	?[e]
26	SC[f]	=O	H	OH	H			H	H	H	H	−6.144	?[e]
27	SC	=O	H	COMe	OH			H	OH	H	H	−6.247	?[e]
28	SC	=O	H	COMe	H			H	Me	H	H	−7.120	?[e]
29	SC[f]	=O	H	COMe	H			H	H	H	H	−6.817	?[e]
30	SC	=O	OH	COCH$_2$OH	OH			H	H	Me	H	−7.688	?[e]
31	SC	=O	OH	COCH$_2$OH	OH			H	H	Me	F	−5.979	?[e]

[a] According to refs 4, 5, and 10.[4,5,10] [b] Of the 5-α steroid series. [c] Of the 5-β steroid series. [d] Assumed to be =O (testosterone) as indicated in the Figure 6,[4] and not as −OH in tables[4] and further publications (compare ref 12). [e] Unknown. [f] H (hydrogen) instead of Me at C$_{10}$ steroid skeleton. [g] No X$_1$ at C$_{17}$.

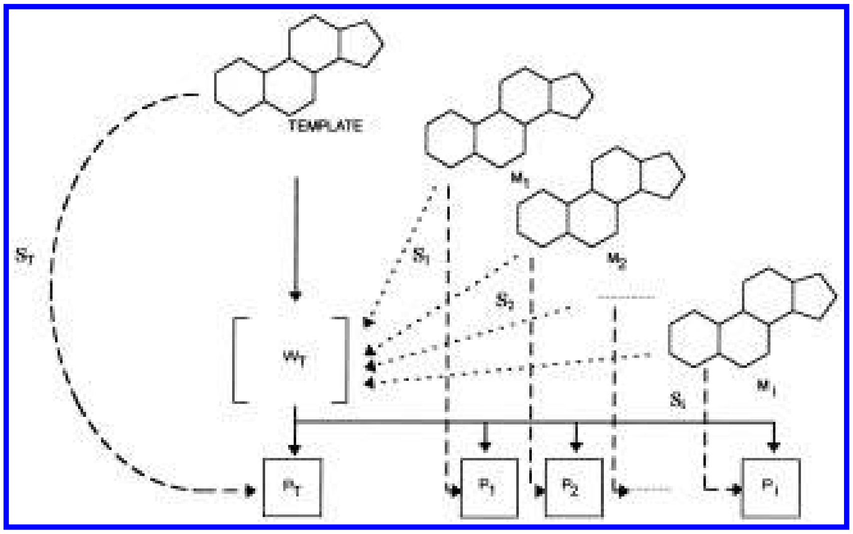
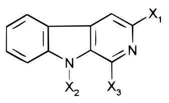
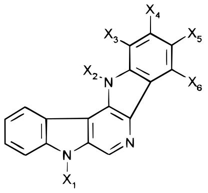


Figure 1. The organization of the net. A template molecule (x, y, and z coordinates of its atoms) is used to train the weight matrix W_T, which is shown by an arrow line. Coordinates of the other analogs are brought into W_T (dotted lines) simulating the patterns P_i. Finally S signals, atom numbers, or atomic charges are transmitted into the patterns linking output neurons with individual input points.

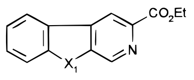
Table 2. Structural and Activity Data for the Analogs of A−G Series According to Ref 10



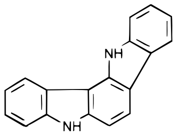
A



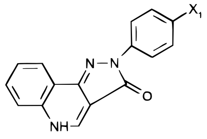
B



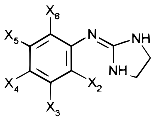
C



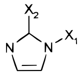
D



E



F



G

no.	structure	X_1	X_2	X_3	X_4	X_5	X_6	activity[a]	no.	structure	X_1	X_2	X_3	X_4	X_5	X_6	activity[a]
1	A	CO_2Me	H	H				−0.699	39	F		Br	H	Br	H	Br	7.46
2	A	CO_2Et	H	H				−0.699	40	F		Cl	H	Cl	H	Cl	7.75
3	B	H	H	H	H	H	H	−0.602	41	F		Me	H	Me	H	Me	10.78
4	A	OEt	H	H				−1.380	42	F		Cl	H	Cl	H	H	8.73
5	A	Oi−Pr	H	H				−2.699	43	F		Me	H	Me	H	H	10.56
6	A	OBu	H	H				−1.991	44	F		Cl	H	H	Cl	H	8.50
7	A	OMe	H	H				−2.093	45	F		Br	H	H	H	Br	7.80
8	A	OPr	H	H				−1.042	46	F		Cl	H	Br	H	Cl	7.72
9	A	COPr	H	H				−0.447	47	F		Cl	H	Me	H	Cl	8.29
10	A	Bu	H	H				−2.389	48	F		Cl	H	NO_2	H	Cl	6.86
11	B	H	H	Me	H	H	H	−1.919	49	F		Cl	H	OMe	H	Cl	8.57
12	B	H	H	H	Me	H	H	−1.000	50	F		Cl	H	H	H	Cl	8.05
13	B	H	H	H	H	Me	H	−2.350	51	F		Et	H	H	H	Et	10.61
14	B	H	H	H	H	H	Me	−3.836	52	F		F	H	H	H	F	8.18
15	B	Me	H	H	H	H	H	−3.066	53	F		Me	H	Br	H	Me	10.21
16	B	H	Me	H	H	H	H	−2.196	54	F		Me	H	Cl	H	Me	10.25
17	B	Me	Me	H	H	H	H	−3.283	55	F		Me	H	H	H	Me	10.53
18	E							−3.295	56	F		Cl	H	Me	H	H	9.41
19	A	H	H	H				−3.210	57	F		Cl	H	H	H	H	9.15
20	A	CO_2t-Bu	H	H				−1.000	58	F		Me	H	Cl	H	H	9.99
21	C	C(=O)						−4.415	59	F		Me	H	H	H	H	10.23
22	C	C(=NOH)						−3.699	60	F		H	H	H	H	H	10.05
23	C	O						−3.964	61	G	Me	Br					3.82
24	C	CH_2						−2.833	62	G	Me	F					2.30
25	D	H						0.398	63	G	Me	H					7.12
26	D	Cl						0.222	64	G	Me	NH_2					8.54
27	E	OMe						1.000	65	G	Me	NO_2					−0.48
28	B	H	H	H	H	H	OMe	−2.398	66	G	H	Br					3.79
29	B	H	H	H	H	H	Cl	−2.854	67	G	H	Cl					3.55
30	A	Cl	H	H				−1.653	68	G	H	Et					7.73
31	A	NO_2	H	H				−2.097	69	G	H	F					2.40
32	A	CO_2	H	H				−2.875	70	G	H	H					6.99
33	A	CO_2Me	H	Et				−3.877	71	G	H	Me					7.86
34	A	H	H	Et				−5.398	72	G	H	NH_2					8.46
35	A	H	H	Me				−4.093	73	G	H	NO_2					−0.81
36	C	C(=O)N(H)						−3.380	74	G	H	C_6H_5					6.48
37	C	S						−3.230	75	G	H	NC_5H_5					5.36
38	F		Cl	Cl	H	H	H	8.55	76	G	H	SMe					5.95

[a] CGS ligands with binding data for the benzoodiazepine receptor inverse agonist site, log ($1/IC_{50}$) for **1**−**37**; affinity data to the cocaine receptor, pK_a, for **38**−**60**; binding to monoamine oxidase, pK_a, for **61**−**76**.

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 2 & 1 & 0 & 2 & 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 0 & 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 0 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 1 & 2 & 0 & 1 & 2 \\ 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 1 \\ 2 & 2 & 2 & 2 & 1 & 1 & 2 & 1 & 0 \end{bmatrix}$$

grid

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 2 & 3 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 1 & 2 & 3 & 2 & 3 \\ 2 & 1 & 0 & 3 & 2 & 1 & 4 & 3 & 2 \\ 1 & 2 & 3 & 0 & 1 & 2 & 1 & 2 & 3 \\ 2 & 1 & 2 & 1 & 0 & 1 & 2 & 1 & 2 \\ 3 & 2 & 1 & 2 & 1 & 0 & 3 & 2 & 1 \\ 2 & 3 & 4 & 1 & 2 & 3 & 0 & 1 & 2 \\ 3 & 2 & 3 & 2 & 1 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 3 & 2 & 1 & 2 & 1 & 0 \end{bmatrix}$$

Manhattan

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

toroidal

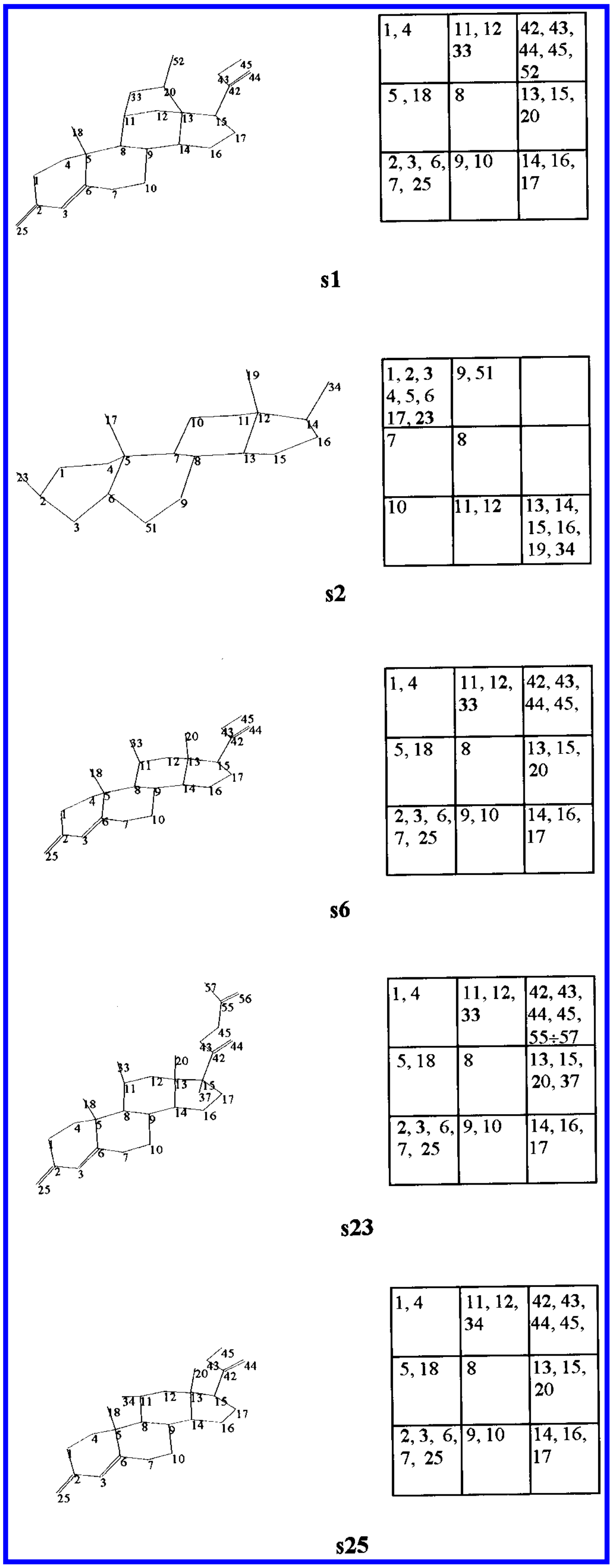
Figure 2. Distance matrices defining a topology of the neighborhoods. The *i*th column of the matrix shows the distances of the *i*th neuron from the *j*th one defined by the individual *j*th row.

the scheme of the net architecture. The most active analog was used to obtain the weight matrix (map) prepared by training the net with coordinates coming from van der Waals surface or simply by the atoms' coordinates. The MATLAB embedded procedure of unsupervised learning with instar rule[11] and neighborhoods simulated by the grid, Manhattan, and toroidal distances were tested. Figure 2 shows the distances for individual weight matrices of the size 3 × 3. Thus the matrix *i*th column shows the distances of the *i*th neuron from the *j*th one defined by the individual *j*th row. Furthermore the trained weight matrix was used as a template to recognize the signals coming from both the template and other analogs. Finally each point coming into any output neuron of the individual pattern P_i are defined by the parameter characterizing directly this point within a molecule (e.g., partial charge of the atom or electrostatic potential for the points coming from van der Waals surface). Input points are located within the resulted SOMs, while the neurons of the outputs can be represented by different parameters. The number of signals coming into a particular output neuron or either the sum or mean value of the signals characterizes the inputted points. The maps (matrices) obtained were used to calculate matrix determinants or to form the inputs to the second layer mapping performed to classify compounds.

RESULTS AND DISCUSSION

Steroid **s1**÷**s31** patterns were prepared according to template technique (as described in PROCEDURES). Compounds were modeled by HYPERCHEM with a steroid skeleton held for each analog within individual stereochemistry subseries. It means that only side chains atoms (i.e., their coordinates) can provide differences within aforementioned subseries of the same skeletons. Figure 3 presents five selected molecular patterns produced while templating by **s6**, one of the most active CBG analog. Values within the maps, prepared for 3 × 3 grid distance, define individual atoms within the molecule; they were transmitted as an S signal, as shown in Figure 1. Although hydrogens were also used to train the template weight matrix and simulate the maps, these were not shown within the patterns to make them more clear. It can be found that similar objects produce similar patterns, i.e., the molecules **s1**, **s23**, and **s25** of the main skeletons matching the template **s6** furnish the maps comparable with the template, while the **s2** definitely differs from the series. Moreover within this series (**s1**, **s23**, and **s25**) atoms of the side chains which cannot be found within the template **s6** tend to be located near these coming from the template which would neighbor them after the superimposition of the molecule on the template. The signal of the atoms of the main skeleton are distributed in the same manner within all patterns of the **s1**, **s6**, **s23**, and **s25**) series.

The maps can be presented as matrices, provided that each output neuron would be characterized by one number. Three main approaches can be taken to the problem; therefore the number of signals or either the sum of signal values or its mean value can yield an output neuron.



s1

1, 4	11, 12 33	42, 43, 44, 45, 52
5 , 18	8	13, 15, 20
2, 3, 6, 7, 25	9, 10	14, 16, 17

s2

1, 2, 3 4, 5, 6 17, 23	9, 51	
7	8	
10	11, 12	13, 14, 15, 16, 19, 34

s6

1, 4	11, 12, 33	42, 43, 44, 45,
5, 18	8	13, 15, 20
2, 3, 6, 7, 25	9, 10	14, 16, 17

s23

1, 4	11, 12, 33	42, 43, 44, 45, 55÷57
5, 18	8	13, 15, 20, 37
2, 3, 6, 7, 25	9, 10	14, 16, 17

s25

1, 4	11, 12, 34	42, 43, 44, 45,
5, 18	8	13, 15, 20
2, 3, 6, 7, 25	9, 10	14, 16, 17

Figure 3. A typical pattern (maps) for five random steroids. Numbers of individual atoms as ascribed by HYPERCHEM and not according to the nomenclature. Hydrogens are omitted within the patterns to make them more clear.

Table 3 identifies the determinants of the matrices **s1**÷**s31**, of different sizes and neighborhoods, prepared for different templates. The partial atomic charge is now the S signal transmitted. Individual determinants D were defined by a number of signals (D_N), sum of signals (D_Σ), and mean signal value (D_M). The size of the smallest map was assumed to provide at least adjacent (distance = 1) and nonadjacent (distance = 2) neurons; i.e., 2 × 2 Manhattan, 3 × 3 grid, and 4 × 4 toroidal neighborhood fulfill the requirement. In Figure 4a−c the aforementioned determinants of the analogs

Table 3. The D_N, D_M, D_Σ Determinants, of the Matrices Formed from Molecular Patterns (Maps) of the Steroids **s1**−**s31**, for Different Neighborhoods and Matrices' Sizes, if Trained with **s6** and **s12**[b]

Template 6 D_N

	map[a]							map[a]					
no.	mn22	mn33	mn44	gr33	gr44	tor44	no.	mn22	mn33	mn44	gr33	gr44	tor44
1	1.0435	0.8220	1.0588	0.1176	1.1425	0.1600	17	2.0483	−4.3429	6.9328	−7.9412	8.9059	−0.5867
2	1.8986	−1.6000	4.7059	−8.3824	−2.6463	0.3600	18	2.1401	−4.5714	6.1008	−9.0000	9.7964	−0.5867
3	1.6812	−3.4286	4.8529	−3.7059	5.3435	−0.5867	19	0.9565	0.6308	1.1176	1.4706	0.9237	0.4600
4	0.7343	0.0615	0.5966	0.0588	0.0076	−0.2333	20	0.8357	0.6308	1.1176	−0.1471	0.4504	0.4600
5	1.6957	−1.3846	3.3613	−6.6176	−2.4427	0.3600	21	0.9130	−0.0835	1.1513	0.8529	0.5827	−0.2333
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	22	1.2609	−0.8462	0.2521	−0.9559	−0.3613	6.6333
7	0.8744	1.1319	0.7647	1.0000	0.4046	1.2000	23	0.8696	2.0132	1.0000	0.7647	1.2748	0.9400
8	0.7101	1.2418	0.5714	−0.1471	0.2672	0	24	0.8696	0.9429	0.9244	1.7059	0.7634	0
9	1.6280	−2.9714	3.4664	−1.5882	3.5623	−0.5867	25	0.9130	1.3846	0.8235	1.2353	0.9618	0.5000
10	0.9130	0.7736	1.0588	1.2353	0.9618	0.4600	26	0.6522	0.4220	0.9832	2.2059	0.4020	−0.9933
11	0.7826	0.1495	1.2269	3.1471	0.3664	−0.5467	27	0.9903	0.7363	1.3039	−0.1176	1.4351	0.2267
12	−0.0725	0.1407	−0.1232	−8.7206	1.1807	1.4933	28	1.0676	0.2637	1.3193	1.5882	1.6031	0.6933
13	0.9179	−0.7846	0.9412	−11.4706	0.3257	0	29	0.6957	1.1363	0.9412	2.8235	0.6870	−0.5000
14	0.7826	−0.8308	1.0756	−12.4265	0.3868	0	30	0.7826	2.4198	1.8067	1.6029	1.0611	−1.1067
15	0.8406	−0.6923	0.6723	−9.5588	0.2036	0	31	0.8406	2.0879	1.3950	2.4118	1.4198	1.9067
16	1.6957	−1.3846	3.3613	−6.6176	−2.4427	0.3600							

Template 6 D_Σ

	map[a]							map[a]					
no.	mn22	mn33	mn44	gr33	gr44	tor44	no.	mn22	mn33	mn44	gr33	gr44	tor44
1	1.1456	0.7537	−1.9246	27.1933	7.5549	−6.7603	17	1.2856	0.0320	−0.1236	−6.0266	−0.0316	−3.4462
2	−0.1285	0.0183	−0.0272	−1.6165	−0.0141	−1.2657	18	1.4168	0.0369	−0.7146	−7.0739	−0.0332	−3.1981
3	1.0591	0.0327	−0.0324	−4.9392	−0.0259	−2.6620	19	2.0715	0.1040	−0.2091	−0.7740	0.3542	−1.6159
4	1.0567	1.5037	0.2137	−0.0773	0.4102	−1.1060	20	4.7592	0.2039	−0.0842	−17.4511	0.4407	−1.3138
5	−0.2106	0.0281	−0.0244	−3.4698	−0.0240	−1.6413	21	0.1647	0.7040	0.3354	0.0854	0.3595	−3.2794
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	22	4.6816	0.2074	−0.2654	8.4638	0.5168	0.0612
7	1.8933	0.3715	−0.4307	3.4184	0.5151	2.4219	23	0.0663	1.6402	−0.1562	−2.4134	−0.0893	−8.6215
8	2.5011	−0.3752	0.0714	−4.4736	0.2169	−0.0411	24	2.0901	−0.1151	0.0010	−1.6851	−0.2515	−3.4725
9	0.9056	0.0021	−0.0473	−3.9806	−0.0371	−2.3585	25	0.8917	−0.3079	−0.7613	−2.5237	−3.4168	−9.4500
10	0.9044	0.1424	−0.2184	−0.3416	0.6840	−3.0061	26	0.177	1.0361	0.2639	11.0623	−0.5242	−8.6497
11	0.1527	0.2864	−0.0025	2.6552	0.0731	−0.5639	27	2.3061	0.9468	0.4827	−3.3250	0.0747	−9.5897
12	−0.0837	0.1933	0.0644	4.5074	0.3495	−2.3174	28	0.0693	0.0526	0.3144	−12.7113	0.1115	−4.8288
13	−0.1500	0.0021	−0.1237	6.8350	−0.1086	0	29	2.1568	0.2557	−0.1467	−0.7397	1.7857	−6.1797
14	−0.2253	0.0047	0.3827	8.7291	−0.0900	0	30	1.2159	1.4826	−0.1720	−7.2304	0.2786	−6.8998
15	−0.2171	0.0000	−0.0876	7.6591	−0.1194	0	31	−0.6798	0.9568	1.0112	7.9421	4.4306	12.3458
16	−0.0949	−0.0006	0.0180	−3.5689	0.0474	−2.5289							

Template 6 D_M

	map[a]							map[a]					
no.	mn22	mn33	mn44	gr33	gr44	tor44	no.	mn22	mn33	mn44	gr33	gr44	tor44
1	1.3284	0.0314	−2.8998	0.4671	4.1113	1.7407	17	8.1904	−0.0051	−0.0121	−0.0197	0.0076	0.0166
2	3.9185	0.0190	−0.1040	−0.0158	0.0009	0.0074	18	9.4128	−0.0052	−0.0797	−0.0230	0.0068	0.0154
3	5.7564	−0.0058	−0.0045	−0.0162	0.0046	0.0130	19	2.1649	0.1867	−0.1264	0.6040	1.1268	0.9617
4	1.4441	1.2925	2.0553	−0.0010	−1.7895	0.0828	20	3.8677	0.1165	−0.0300	−1.4004	−0.3941	−0.3164
5	4.4536	0.0270	−0.1212	−0.0307	0.0011	0.0096	21	0.1328	1.0269	2.5697	0.0013	−0.9240	0.2188
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	22	4.9574	0.0367	−0.3436	−0.0325	−0.5169	0.0322
7	1.9567	0.8052	−0.0500	−1.0121	−0.6718	−0.0419	23	0.1504	1.0133	−0.2098	−0.0796	−0.2885	−0.3141
8	2.2230	−0.4323	−0.0507	−1.3869	−0.7362	1.2999	24	2.2884	−0.1679	0.1108	0.5263	1.2927	1.3502
9	7.0065	−0.0005	−0.0093	−0.0131	0.0040	0.0113	25	0.8869	−0.3977	−1.3057	0.5563	2.1218	2.8739
10	0.9126	0.2197	−0.1851	0.8106	1.0745	1.3587	26	0.1613	1.1546	1.0259	0.1416	−0.7546	0.6061
11	0.3040	0.9300	−0.0063	0.1541	−0.0174	0.0813	27	2.1284	0.4999	0.5570	0.1081	1.1889	−0.6217
12	−0.0522	−0.1474	0.1358	−0.0235	−0.2272	−0.0340	28	0.3269	0.5571	0.5399	−0.6147	1.1719	0.7694
13	1.8534	0.0003	−0.0894	0.0111	−0.0007	0	29	2.2542	0.3152	−0.2433	0.1000	0.8246	2.1121
14	1.3777	0.0007	0.2420	0.0131	0.0228	0	30	0.8473	1.1885	−0.2166	−0.8017	−0.0458	−2.5447
15	2.2425	0.0000	−0.0886	0.0150	0.0505	0	31	0.1604	0.8881	0.3768	1.3829	−0.7051	−2.7582
16	−1.7627	0.0185	0.0901	−0.0318	−0.0022	0.0138							

Template 12 D_N

	map[a]						map[a]				
no.	mn33	mn44	gr33	gr44	tor44	no.	mn33	mn44	gr33	gr44	tor44
1	238	0	120	−16	−3	8	129	0	330	76	−36
2	124	98	710	0	172	9	0	−250	468	0	132
3	0	−350	624	0	132	10	119	0	330	−8	−9
4	0	0	540	70	−8	11	137	0	486	8	18
5	88	70	568	0	172	12	0	392	1120	936	0
6	128	0	396	45	63	13	0	0	384	0	0
7	137	0	396	216	63	14	0	0	432	0	0

Table 3 (Continued)

Template 12 D_N

	map[a]						map[a]				
no.	mn33	mn44	gr33	gr44	tor44	no.	mn33	mn44	gr33	gr44	tor44
15	0	0	288	0	0	24	0	0	516	−25	−54
16	88	90	648	0	156	25	119	0	330	−20	−6
17	0	−500	858	0	132	26	0	0	420	−8	36
18	0	−550	936	0	132	27	0	0	558	−55	−17
19	0	0	516	−20	−12	28	146	0	504	22	−7
20	0	0	516	71	−9	29	0	0	384	1	63
21	0	0	540	40	−12	30	185	0	558	−45	56
22	120	−6	342	48	42	31	160	0	495	−3	9
23	822	0	672	−24	−56						

Template 12 D_Σ

	map[a]						map[a]				
no.	mn33	mn44	gr33	gr44	tor44	no.	mn33	mn44	gr33	gr44	tor44
1	−0.0600	0	0.0112	−0.0050	−0.0005	17	0	0.0002	0.0000	0	0.0001
2	−0.0133	0.0001	−0.0015	0	−0.0015	18	0	0.0003	−0.0001	0	0.0000
3	0	0.0001	0.0001	0	0.0002	19	0	0	0.0020	−0.0033	−0.0005
4	0	0	0.0026	−0.0018	0.0000	20	0	0	0.0004	−0.0067	0.0003
5	−0.0120	0.0000	−0.0009	0	−0.0015	21	0	0	0.0022	−0.0014	−0.0003
6	−0.0162	0	0.0012	−0.0067	−0.0003	22	−0.0416	0.0024	−0.0192	0.0000	−0.0007
7	−0.0302	0	0.0011	−0.0071	0.0002	23	−0.0425	0	0.0053	−0.0057	−0.0002
8	−0.0281	0	0.0005	−0.0205	−0.0004	24	0	0	0.0019	0.0040	−0.0003
9	0	0.0002	0.0000	0	0.0002	25	−0.0032	0	0.0012	−0.0151	−0.0016
10	−0.0142	0	0.0013	−0.0072	−0.0006	26	0	0	0.0002	0.0001	0.0000
11	−0.0049	0	−0.0002	0.0002	0.0000	27	0	0	0.0039	−0.0017	−0.0001
12	0	0.0004	−0.0013	−0.0001	0	28	−0.0143	0	0.0005	−0.0036	0.0000
13	0	0.0000	0.0002	0	0	29	0	0	0.0019	−0.0002	−0.0001
14	0	−0.0001	0.0006	0	0	30	−0.0305	0	0.0046	−0.0015	−0.0002
15	0	0.0000	0.0002	0	0	31	−0.0165	0	0.0012	0.0082	−0.0002
16	−0.0211	0.0003	−0.0009	0	−0.0082						

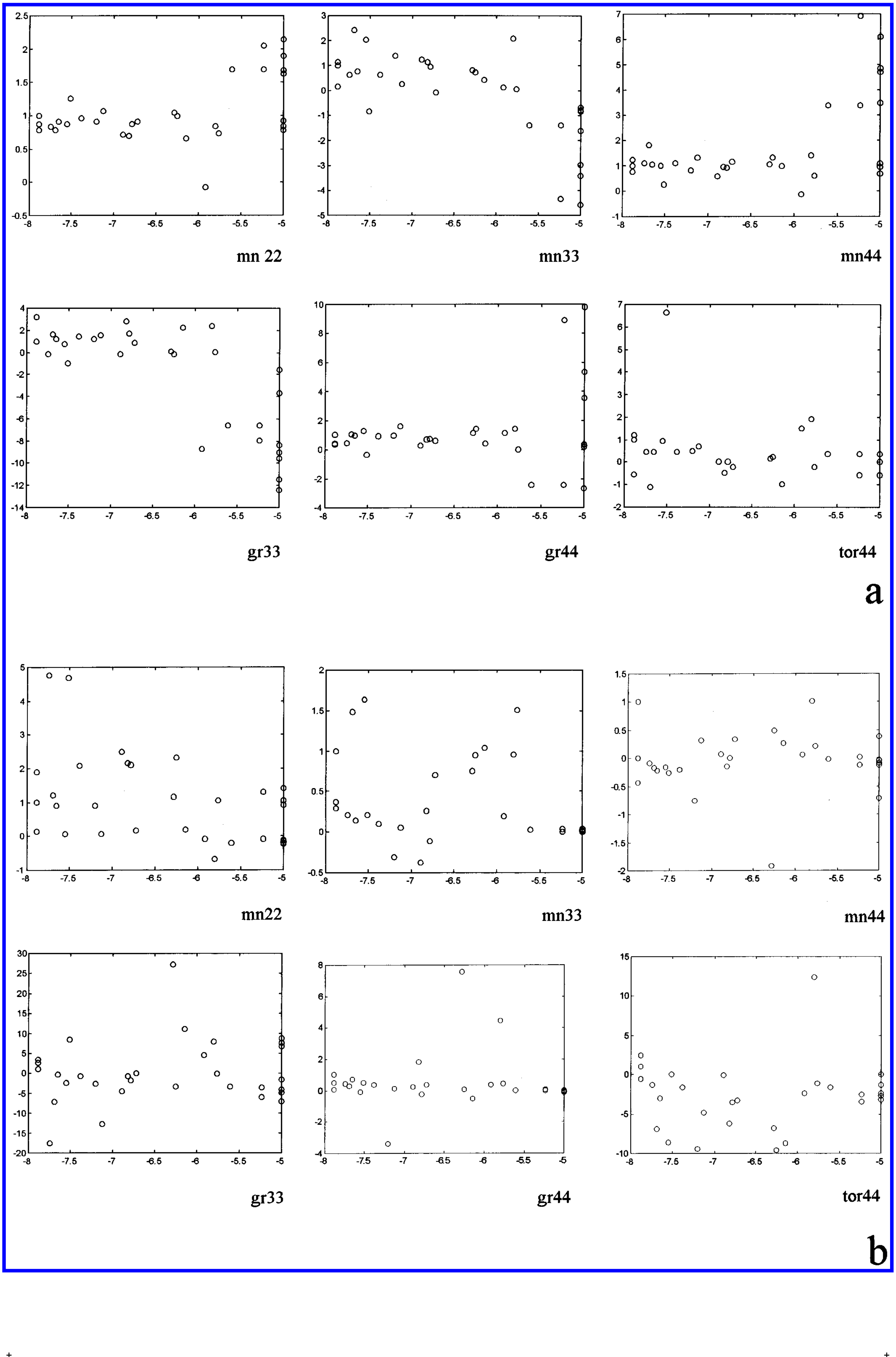
Template 12 D_M

	map[a]							map[a]					
no.	mn33	mn44	gr33	gr44	tor44	×10^3	no.	mn33	mn44	gr33	gr44	tor44	×10^3
1	0.0016	0	0.0311	−0.0771	−0.0034		17	0	0.0003	0.0000	0	−0.0811	
2	−0.0001	0.0003	−0.0145	0	−0.0198		18	0	0.0003	−0.0001	0	−0.0921	
3	0	0.0001	0.0001	0	−0.0711		19	0	0	0.0197	−0.0544	−0.0046	
4		0	0	0.0048	−0.0211	0.0002	20	0	0	0.0088	−0.0976	−0.0015	
5	−0.0001	0.0003	−0.0105	0	−0.0200		21	0	0	0.0041	−0.0241	−0.0014	
6	0.0014	0	0.0048	−0.0308	−0.0049		22	−0.0003	0.0253	−0.0606	0.0009	−0.0268	
7	0.0013	0	0.0033	−0.0400	−0.0019		23	0.0004	0	0.0132	−0.0209	−0.0009	
8	0.0018	0	0.0019	−0.1686	−0.0054		24	0	0	0.0207	0.1081	−0.0053	
9	0	0.0004	0.0000	0	−0.0794		25	0.0004	0	0.0044	−0.0336	−0.0115	
10	0.0015	0	0.0049	−0.0372	−0.0043		26	0	0	0.0004	−0.0101	−0.0006	
11	−0.0001	0	−0.0026	0.0074	0.0000		27	0	0	0.0243	−0.0011	−0.0047	
12	0	0.0028	−0.0012	−0.0036	0		28	0.0013	0	0.0028	−0.0687	−0.0009	
13	0	0.0001	0.0005	0	0		29	0	0	0.0165	−0.0457	−0.0035	
14	0	−0.0007	0.0015	0	0		30	0.0014	0	0.0080	0.0671	−0.0034	
15	0	0.0001	0.0007	0	0		31	0.0014	0	0.0033	0.0084	−0.0020	
16	−0.0002	0.0009	−0.0130	0	−0.1537								

[a] Neighborhood mn, Manhattan; gr, grid; tor, toroidal; size: 22, 2 × 2; 33, 3 × 3; 44, 4 × 4. [b] For the **s6** template the individual determinants values are divided by D_N(**s6**), D_Σ(**s6**), and D_M(**s6**), respectively.

templated by **s6** are plotted vs compounds CBG activity (specified in Table 1). Although it is not easy to establish simple and general rules, some regularities can be obviously observed. Therefore, within D_N plots (Figure 4a) of the 3 × 3 size, compounds are distributed among two ranges, distinguishing analogs of the higher and lower activities. It is necessary to analyze the meaning of the D_N descriptor to understand the regularity. First of all it is a distribution of the signal within the map that is defined by D_N, which means it is a "topology" of the map, consequently a *topology* of the molecule, and its shape that is described by the parameter. No other information (charge) is transmitted to D_N descriptors. To make the situation clear, it should be mentioned here that the shape limits CBG activity.[5] Thus D_N indexes can be used to characterize molecule's shape.

A different situation can be observed for the D_M plots (Figure 4c). Now, mostly, all but the lowest activity analogs are fairly distributed within the plot. Even the analogs of the highest CBG values (**s6**, **s7**, **s11**) are defined by quite different D_M values. Analogs of the lowest CBG provide the D_M in the relatively narrow range. Moreover, the D_M of the latter resembles, in many plots, that of the s11 one; individual values should be extracted from Table 3, to enable the comparisons. As the D_M parameter obviously brings information about the distribution of charge within the molecule, it can be speculated that low CBGs differing from


mn 22
mn33
mn44
gr33
gr44
tor44
a
mn22
mn33
mn44
gr33
gr44
tor44
b
-8
-7.5
-7
-6.5
-6
-5.5
-5

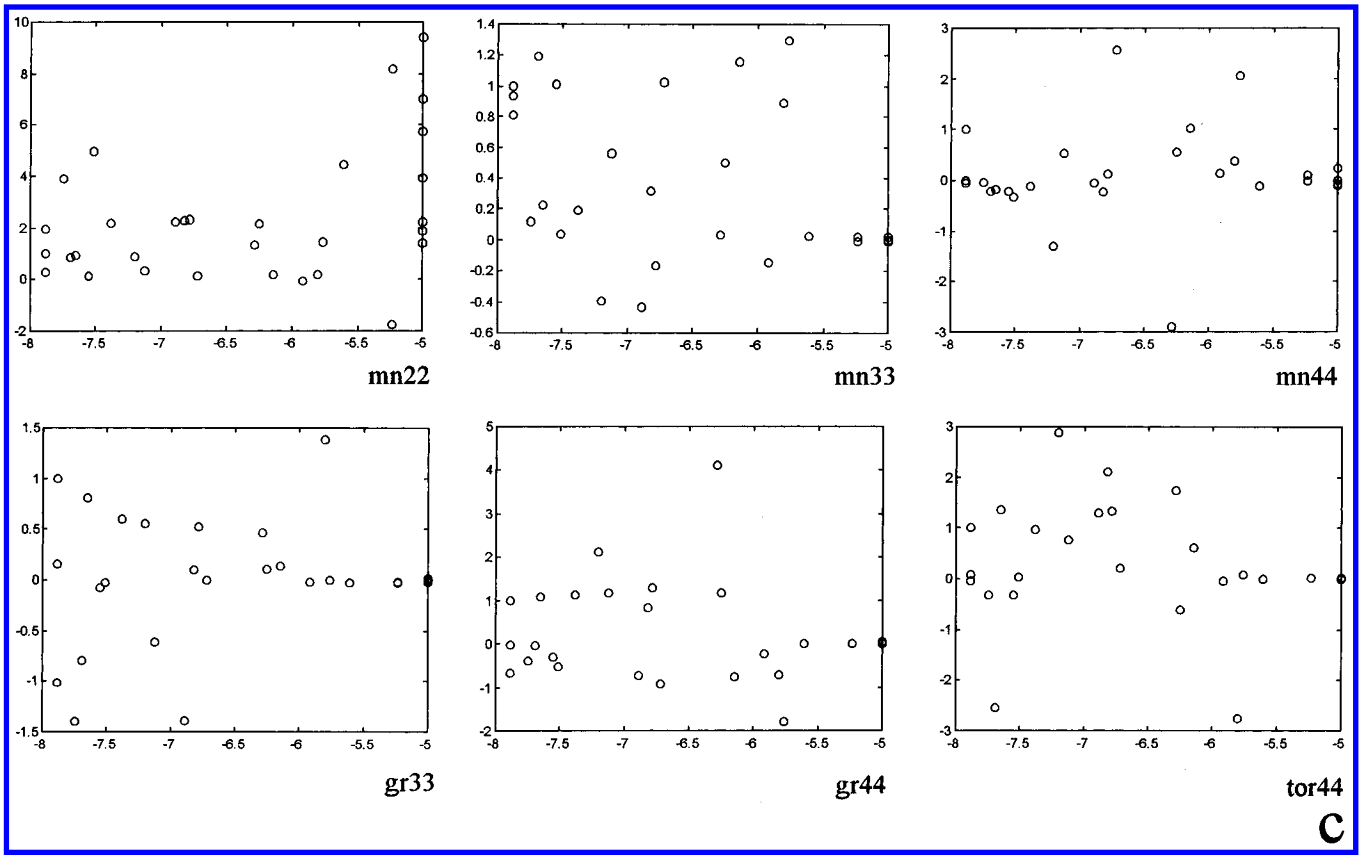


Figure 4. Plots of the matrices' determinants of a, D_N; b, D_Σ; and c, D_M obtained for the patterns produced with the **s6** template vs CBG affinities. If possible (D(**s6**) ≠ 0) D values are scaled by dividing by D(**s6**).

the template by shape provide a better electrostatic match, not necessarily with **s6** but with **s11**. As can be compared with molecules structures (Table 1), **s11** (among **s6**, **s7**, **s11**) actually provides the best fit with the Lows, e.g., there is no C-11 hydroxyl group (X_2-group) in the whole series. Although D_M transmits information on both the shape and partial charges it seems that the descriptor is very sensitive to the latter parameter.

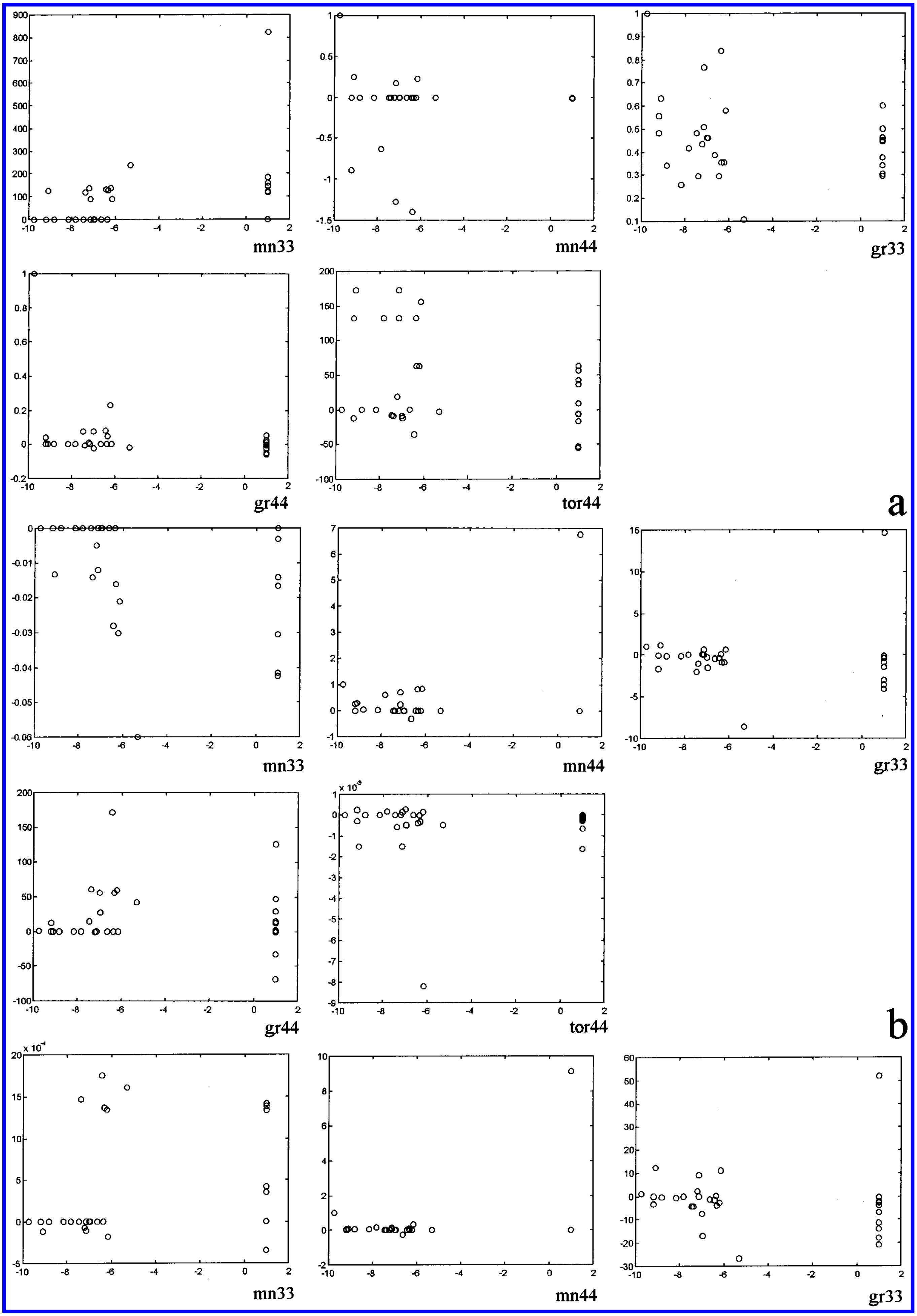
The D_Σ descriptor makes something between D_N and D_M, but in this particular case the relationships obtained cannot be interpreted unambiguously. The distribution of the compounds within the plots depends upon the size and individual neighborhood design.

By analogy Figure 5 provides plots for the **s12** taken as the most active TBG template (the TBG values of 1 were assigned to all compounds, **s22**÷**s31**, of the unknown activity, to make them clearly different within the plot). As indicated before,[5,9] the activity does not depend upon the shape factors. Actually the plots cannot be interpreted easily, which suggests that shape factors still predominate within all descriptors.

Although the example analyzed indicates that matrix determinants carry an information characterizing similarities between the template and individual molecules, the parameters must obviously reduce the data available, bringing the matrix down to a single number. On the other hand, the operation makes easier any analysis of the problems. The approach closely resembles this of topological indexes. The template provides normalization, which makes the series of matrices and their D descriptors invariant of the training procedures; therefore, the patterns obtained can be understood as matrices of similarity, while their determinants D (e.g., D_N, D_M, D_Σ) form indexes of so defined similarity matrices. A lot of parameters different from determinants can be used to characterize the matrices, or even some defined parts of them, and form quantitative structure activity relationships. Another possibility is to bring the whole matrices into the analyses by introducing them as the inputs to neural nets, e.g., SOM procedure, which is capable of classifying objects.

Typical classifications yielded by SOMs fed with the matrices constructed according to the aforementioned rules are shown in Figure 6. Even relatively small SOMs of the first layer provide reasonable classifications for the series where the shape determines compounds activity (i.e., steroid series **s1**−**s31** in relation to their CBG activity). This conclusion corresponds with the one coming from the analyses of the D determinants performed before. The best classifications are provided by the first layer maps describing the quantity of signals coming into the appropriate output neuron (Figure 6a); therefore, the topology of the information within the map still seems to be the most important. A more careful analysis of the plot in Figure 6a reveals that H, M, and L superscripts indicating activity levels (as ascribed in ref 5 according to the values actually observed) also tend to cluster; i.e., L's group at the right side of the plot, while H's group at the left one.

It seems that the maps of the first layer should compress information, i.e., better classifications are obtained for the maps of the relatively higher number of signals coming into each neuron. Once more as for determinants the 3 × 3 size seems to be the optimal, while the neighborhood construction (grid, Manhattan, toroidal) seems not to be so important; therefore, it can be concluded that there is probably some optimal signal/output neurons ratio which limits the process.


mn33
mn44
gr33
gr44
tor44
a
mn33
mn44
gr33
x 10^-3
gr44
tor44
b
x 10^-4
mn33
mn44
gr33
-10
-8
-6
-4
-2
0
2

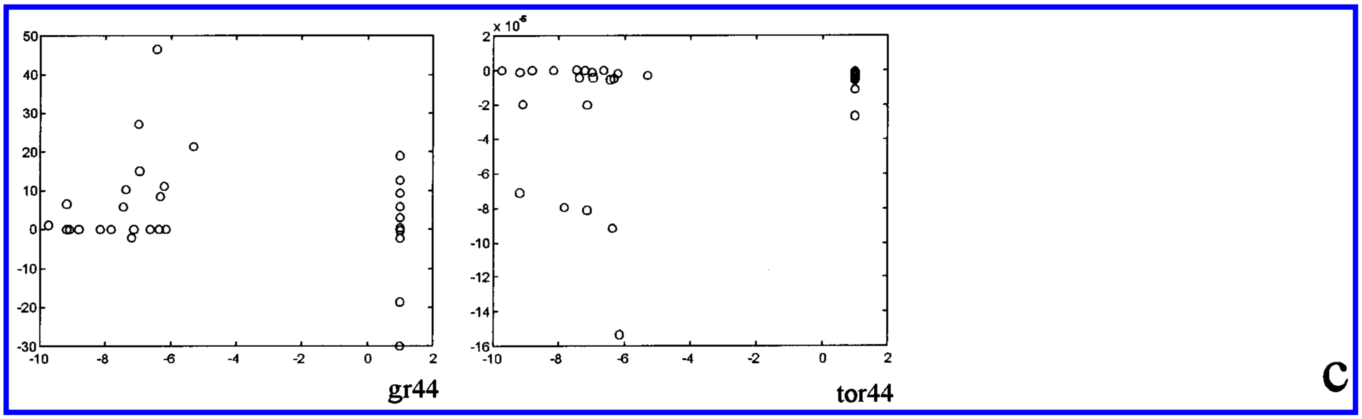


Figure 5. Plots of the matrices' determinants of a, D_N; b, D_Σ; and c, D_M obtained for the patterns produced with the **s12** template vs TBG affinities. If possible (D(s12) ≠ 0) D values are scaled by dividing by D(s12). TBG value of 1 are ascribed to the compounds of the unknown TBG affinity (compare Table 2).

1^M, 23^H	25^H,6^H, 7^H,10^H, 28^H			30^H, 31^M		13^L, 14^L, 15^L
						17^L, 18^L
8^M, 20^H	19^H	11^H		12^M		3^L
	27^M		29^M	24^M		9^L
22^H			21^M, 26^M	4^L		2^L, 5^L, 16^L

a

6^H,10^H, 25^H, 31^M		19^H				22^H
	24^M		27^M, 29^M			7^H, 8^M
1^M				28^H	30^H	20^H
21^M						
	26^M		12^M		2^L	5^L, 16^L
4^L, 23^H						
11^H			13^L, 14^L, 15^L		3^L, 17^L, 18^L	9^L

b

Figure 6. Classifications resulting from the SOM (7 × 7 of the grid neighborhood) fed with molecular maps of the 3 × 3 grid distance, trained with **s6** (atomic coordinates). Superscripts indicate the affinity levels as ascribed earlier[5] (H, high; M, medium; L, low). The **s** is omitted in the compounds' labels. (a) molecular maps prepared from the number of input signals found in individual output neuron and (b) molecular maps prepared from the mean value of input signals defined by the atomic charge.

1			4^L, 26^M, 21^M			
						2^L, 5^L, 16^L
	12^M		22^H			
3^L, 9^L		17^L, 18^L			13^L, 14^L, 15^L	

1/ Highs: 6, 7, 10, 11, 19, 20, 23, 25, 28, 30
Mediums: 1, 8, 24, 27, 29, 31

Figure 7. Classifications resulting from the SOM (7 × 7 of the grid neighborhood) fed with molecular maps of the 7 × 7 toroidal distance, trained with **s6** (complete van der Waals coordinates), while assuming the number of incoming signals as output neurons values. Superscripts indicate the affinity levels as ascribed earlier[5] (H, high; M, medium; L, low). The **s** is omitted in the compounds' labels.

In this particular case the ratios amount to ca. 14, 6, and 3 signals per neuron for 4 (2 × 2), 9 (3 × 3), and 16 (4 × 4) output neurons, respectively. Probably it is the reason why the neighborhood is assumed less important; maybe 3 × 3 toroidally shaped matrices would better mimic reality,[7,8] but the smallest one must cover 4 × 4 neurons. Otherwise only one type of output neurons (distance = 1) would be obtained. Generally, classifications gained can be compared with the ones resulting from the analysis of the full van der Waals surfaces.

Furthermore it should be realized that no information about the compounds activity is transmitted into the net; therefore, reasonable clustering provides a possibility of good predictions. To make them easier, the similarities to, and dissimilarities from, the template can be overexaggerated. This is first of all the location of the empty neurons within the patterns that can be used now; therefore let us ascribe 0 to the empty output neuron and 1 to the occupied one. Classification obtained for such molecular patterns composed of zeros and ones are shown in Figures 7 and 8a,b. The first one resulted from the classification of 7 × 7 toroidally shaped patterns coming from the complete van der Waals surfaces, while the latter are for 3 × 3 grid and 4 × 4 toroidal patterns, respectively. In all three plots almost all Highs

1			4^{L},			12^{M}
			2/3/5/9/ 16/17/ $18all^{L}$			
						13^{L}, 14^{L}, 15^{L}
26^{M}, 21^{M}						

1/ Highs: 6, 7, 10, 11, 19, 20, 22, 23, 25, 28, 30
Mediums: 1, 8, 24, 27, 29, 31

a

1						13^{L}, 14^{L}, 15^{L}
			4^{L}, 26^{M}, 21^{M}			
43			7^{H}	12^{M}		2/3/5/9/ 16/17/ 18 all^{L}

1/ Highs 22, 23, 30
Mediums 1, 27
43/ Highs: 6, 10, 19, 20,25, 28
Mediums:, 8, 24, 27, 29, 31

b

Figure 8. Classifications resulting from the SOM (7 × 7 of the grid neighborhood) fed with molecular maps trained with **s6** (atomic coordinates), while assuming the number of incoming signals as output neurons values, of the (a) 3 × 3 grid distance and (b) 4 × 4 toroidal distance. Superscripts indicate the affinity levels as ascribed earlier[5] and (H, high; M, medium; L, low). The **s** is omitted in the compounds' labels.

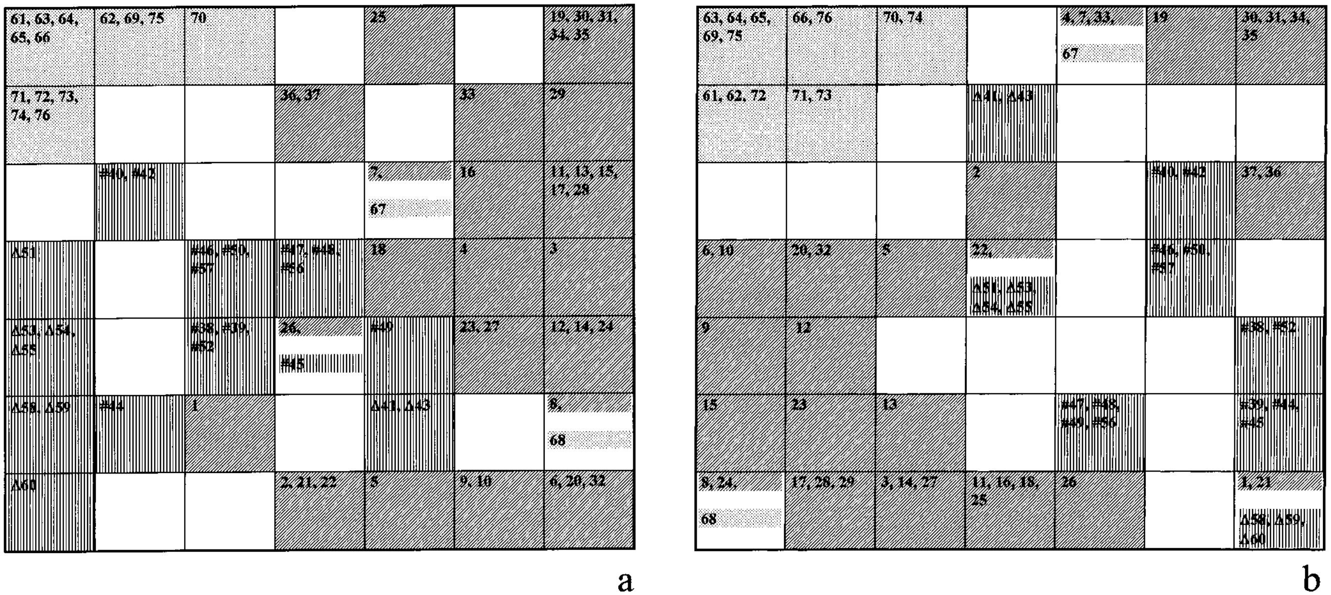
23^{H}
30^{H}, 31^{M}
1^{M}, 6^{H}, 7^{H}, 28^{H}
10^{H}, 11^{H}, 25^{H}, 27^{M}
8^{M}, 19^{M}, 20^{H}, 24^{M}, 29^{M}
22^{H}
21^{M}, 26^{M}
4^{L}
13^{L}, 14^{L}, 15^{L}
12^{M}
2^{L}, 5^{L}, 16^{L}
9^{L}
3^{L}, 17^{L}, 18^{L}

a

61, 62, 63, 64, 65, 66, 69, 71, 72, 73, 74, 75, 76
25, 70
3, 11, 12, 13, 14, 15, 16, 17, 27, 28, 29,
7, 10, 19, 20, 23, 24, 30, 31, 32, 33, 34, 35, 36, 37, 67
1, 2, 4, 6, 8, 9, 18, 21, 22, 68
5, Δ(T)41, Δ43, Δ51, Δ53, Δ54, Δ55, Δ58, Δ59
#38, #39, #40, #42, #44, #45, #46, #47, #48, #49, #50, #52, #56, #57, Δ60

b

Figure 9. Classifications resulting from the SOM organized as a one-dimensional map. The 3 × 3 patterns of the grid distance are inputted, while assuming the number of incoming signals as output neurons values. (a) For steroids series with **s6** template Superscripts indicate the affinity levels as ascribed earlier[5] (H, high; M, medium, L, low). The **s** is omitted in the compounds' labels. (b) For non-congeneric structures of Table 2, with 41 as the template. △ (above 9.99) and # (below 9.99) are used to mark the activity levels of imidazolines series.



61, 63, 64, 65, 66	62, 69, 75	70		25		19, 30, 31, 34, 35
71, 72, 73, 74, 76			36, 37		33	29
	#40, #42			7, 67	16	11, 13, 15, 17, 28
Δ51		#46, #50, #57	#47, #48, #56	18	4	3
Δ53, Δ54, Δ55		#38, #39, #52	26, #45	#49	23, 27	12, 14, 24
Δ58, Δ59	#44	1		Δ41, Δ43		8, 68
Δ60			2, 21, 22	5	9, 10	6, 20, 32

a

63, 64, 65, 69, 75	66, 76	70, 74		4, 7, 33, 67	19	30, 31, 34, 35
61, 62, 72	71, 73		Δ41, Δ43			
			2		#40, #42	37, 36
6, 10	20, 32	5	22, Δ51, Δ53, Δ54, Δ55		#46, #50, #57	
9	12					#38, #52
15	23	13		#47, #48, #49, #56		#39, #44, #45
8, 24, 68	17, 28, 29	3, 14, 27	11, 16, 18, 25	26		1, 21 Δ58, Δ59, Δ60

b

Figure 10. Classifications resulting from the SOM (7 × 7 of the grid neighborhood) fed with molecular maps of the 3 × 3 grid distance trained with atomic coordinates, while assuming number of signals as output neurons values for (a) **14** as the template and (b) **34** as the template. △ (above 9.99) and # (below 9.99) are used to mark the activity levels of imidazolines series.

together with Mediums are put into the neuron occupied by the template **s6**, which makes the operation very interesting for prediction procedures.

Two-dimensional plots can give, however, only ambiguous classifications in relation to one-dimensional parameters. Figure 9a shows the second layer organized as a one-dimensional axis (1 × 15) to be compared with the parameter of the same dimensionality. It can be seen that the distribution observed can be interpreted in terms of biological activities of the analogs input. It is worth noticing that predictions performed with the method do not include any artificial training step which would make use of the compounds activity. Compounds are not forced to be grouped according to their activities. Therefore prediction in the method should be seen rather as an extension of the simulated recognition processes, which makes a close analogy with the reality.

Thus far, analyses performed covered a series of steroid congeners. Let us try, however, to shape, according to the method, a group of noncongeneric molecules **1**−**76**. These were modeled by HYPERCHEM and noted by coordinates without any restriction to preserve similarities. A molecular coordinate system (mcs) is used to record atom which are directly fed into the neural procedure. Molecules inertial axes form the *x*, *y*, and *z* axes of so defined mcs. Figures 10a,b shows classifications obtained for the series **1**−**76**, using templates 14 and 41, respectively. With minor exceptions, compounds were obviously classified into three main series. Moreover imidazoline subseries **38**−**60** were divided into the ones of the higher p*K* amounting to at least 9.99 and Mediums + Lows. The one axis plots also provide reasonable classification, as shown in Figure 9b. Among noncongeners **1**−**76** imidazolines (**38**−**60**) are put into two neighboring neurons, and as within two-dimensional plots they are separated into two groups of the lower and higher activities. Only one exception of compound **60** can be observed.

The approach presented is simple, and the analogy with natural systems is outstanding. Thus a template formed from the most actie molecule should be seen as the receptor model, while the comparison of each molecule performed by the template net models an operation performed by natural receptor unit.

CONCLUSIONS

It has been demonstrated that simple small self-organizing maps trained only with the coordinates of the atoms within molecule are capable of recognizing similar molecular objects. Therefore the template molecule which can furnish a receptor model forms a basis to normalize and prepare molecular patterns (matrices). Depending upon signals transferred these patterns can bear information including either a bare shape or also a charge component. Similarities within patterns can be analyzed simply by means of the calculation of matrix determinants or by inputting them to neural classification. The determinants presented bring the matrices and, consequently, the molecule to a single number which can be regarded as an index characterizing similarity; therefore, resembling topological indexes' approach. The resulted classifications indicate the utility of the procedure for the simulation of the molecular recognition processes; therefore, making possible both finding the similarities within molecules analyzed, distinguishing the effects operating within the series of compounds stimulating one receptor, and predicting the activity for the new analogs.

ACKNOWLEDGMENT

Thanks are due to Dr. Andrzej Polański from the Institute of Automatic Control of the Silesian Technical University Gliwice for valuable comments as well as for providing the possibility of performing some computational operations. The financial support of the Rector of the University of Silesia in Katowice and that of Stefan Batory Foundation from Warsaw which made it possible to present this paper at the XIth International Conference on Computers in Chemical Research and Education in Paris is also gratefully acknowledged.

REFERENCES AND NOTES

(1) Kier, L. B. Indexes of Molecular Shape from Chemical Graphs. *Med. Res. Rev.* **1987**, *7*, 417−440.

(2) Hopfinger, A. J. A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196−7206.

(3) Hopfinger, A. J. Theory and Application of Molecular Potential Energy Fields in Molecular Shape Analysis: A Quantitative Structure−Activity Relationship Study of 2,4-diamino-5-benzylpyrimidines as Dihyrofolate Reductase Inhibitors. *J. Med. Chem.* **1983**, *26*, 990−996.

(4) Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(5) Good, A. C.; So, S. S.; Richards, W. G. Structure−Activity Relationships from Molecular Similarity Matrices. *J. Med. Chem.* **1993**, *36*, 433−438.

(6) Kohonen, T. Self-Organization and Associative Memory; Springer: Berlin, 1988.

(7) Gasteiger, J.; Li, X.; Rudolph, Ch.; Sadowski, J.; Zupan, J. Representation of Molecular Electrostatic Potential by Topological Feature Maps. *J. Am. Chem. Soc.* **1994**, *116*, 4608−4620, and references cited.

(8) Zupan, J.; Gasteiger, J. Neural Networks of Chemists; VCH: Weinheim, 1993; pp 285−291.

(9) Polański, J.; Gasteiger, J. The Comparison of Molecular Surfaces by an Assembly of Self-Organizing Neural Network. Computers in Chemistry ′94″; Technical University of Wroclaw, National Institute of Standards and Technology: Gaithersburg, MD, U.S.A., Wroclaw, 1994; p 88, full publication under preparation.

(10) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR from Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation methods. *J. Med. Chem.* **1993**, *36*, 2929−2937.

(11) Demuth, H.; Beale, M. Neural Network Toolbox. For Use with MATLAB; The MathWorks, Inc.: Natick, MA, 1994; pp 8.1−8.22.

(12) Wagener, M.; Sadowski, J.; Gasteiger, J.; Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic AH Receptor Activity by Neural Networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.

CI9501251