# The Receptor-like Neural Network for Modeling Corticosteroid and Testosterone Binding Globulins

Jarosław Polański

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

Received June 30, 1996[⊗]

A neural-net method for simulation of corticosteroid and testosterone binding globulin (CBG, TBG)-ligand interactions is presented. Molecular modeling provides the geometry and partial atomic charges of 31 steroid molecules. The atomic coordinates within the molecule of the compound of the highest affinity are then used to train a self-organizing map (SOM) that forms a template for the comparison to other molecules. Comparison is done using a series of normalized patterns produced by the SOM. The template SOM, after overlaying on the set of random vectors, mimics the topology of the receptor site and is used to train unsupervisedly a neuron capable of recognizing the degree of similarity between the reference and tested patterns. A good correlation is observed for signals generated by the neuron plotted against the experimental CBG affinities. For TBG affinity modeling a modified procedure is designed which is capable of separating electrostatic and shape effects. The high predictive power of the model is achieved by keeping close analogy to the processes taking place at the real receptor sites.

## INTRODUCTION

A similarity has always focused a strong interest in chemical science, and recently the whole concept of molecular similarity has made substantial gains in importance.[1,2] The studies of molecular recognition within biological receptors and drug design are probably the most important reasons for this. A number of approaches have been developed to measure molecular similarities of the electronic and steric background for the purposes of quantitative structure−activity relationships (QSAR). Many are based on the comparisons between the reference and analyzed molecules. These methods started from a simple two-dimensional reference (template).[3] Three-dimensional description was offered by the Molecular Shape Analysis (MSA)[4,5] and further methods, in particular by the complex Comparative Molecular Field Analysis (CoMFA)[6] approach, aimed at the comparison of molecular surfaces.

Although it is generally accepted that the shape and electrostatic potential of the molecular surface determine interactions between any molecule and receptor site, there are still many problems with the studies of the recognition processes. No general and obvious enough numerical model of the molecule can be proposed which can sufficiently describe molecules of very different sizes and characters which often stimulate the same receptor.

The aim of the current work is to present a neural algorithm capable of modeling various receptor sites. The proposed approach is based on the analysis of molecular similarities like the methods mentioned above. However, by the appropriate design of the assembly of neural nets it was possible to mimic molecular mechanisms of interactions that take place at the real receptor site, and as a result obtain a high predictive power of the procedure, outperforming those reported in the literature.

## METHODS

**Self-Organizing Maps.** Neural self-organizing maps (SOM) were designed to provide the nets preserving the topology, while reducing the dimensionality of the input objects.[7] Gasteiger and Zupan with coworkers inspired the application of the SOMs trained with the Kohonen rule to obtain two-dimensional plots representing three-dimensional molecules.[8,9] The SOM net trained with the coordinates of the points taken from (van der Waals' or Connoly's) molecular surfaces, after projections of the electrostatic potential calculated for these points, gives a two-dimensional molecular map. That technique can be used to visualize the interactions of individual compounds with some biological receptors.[9,10]

**Comparative Self-Organizing Maps.** SOM technique allows for the comparison of the transformed objects. Therefore a reference molecule can be indicated to form a template SOM net that is trained with coordinates coming from van der Waals surface. Such a net can be applied as a base for projections of the surface vector (vectors), e.g., electrostatic potential from another molecule (molecules).[11−13] The resultant feature maps can be seen as a kind of superimposition of the tested and reference molecules. The patterns can be compared, by means of the simple subtraction[12] of the reference and tested maps, or classified by the use of the second neural layer.[11,13]

**Small Self-Organizing Maps.** Molecular surfaces and electrostatic potential can be substituted for the atomic coordinates and partial atomic charges. Such an operation reduces the size of the SOM that should be prepared to a size of few neurons. The maps obtained can be used for the efficient classification of bioactive compounds according to their activity levels, while the results are comparable with those yielded by larger maps.[13]

**Hypermolecule.** Methods called Minimal Topological Difference (MTD) and Minimal Steric Difference (MSD) have been proposed by Simon for comparative shape analysis.[14,15] Comparisons are made by finding topologically overlapping and nonoverlapping molecular elements. The original MTD, MSD, and related methodologies slightly differ from that applied here, but a similar approach is used; an assembly of all atomic positions of the whole molecule

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0095-2338(96)00105-9 CCC: $14.00 

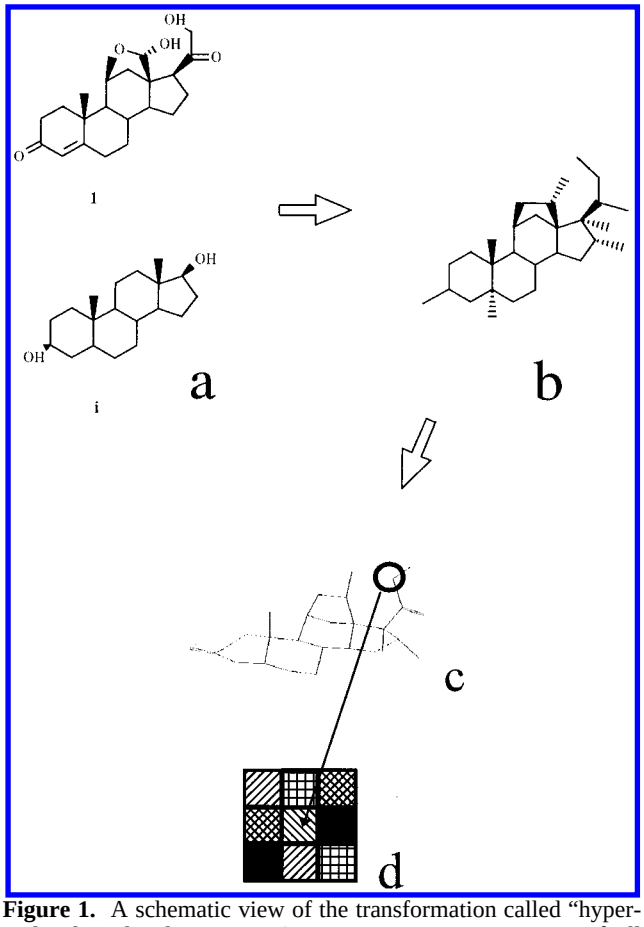



**Figure 1.** A schematic view of the transformation called "hypermolecule", details in text. Atom per atom superposition of all molecules within data shown schematically by two molecules (a) produces a two-dimensional hypothetical structure called hypermolecule **HM** (b), which is capable of describing all analogs **1**−**21**. The signals given by the individual atoms within separate analogs are defined by 1, if the atom can be found in this certain analog, or 0—representing the lack of the atom. For example, the atom encircled gives 1 for the analog **1**, but 0 for the analog **2**. **HM** after defining stereochemistry (that of the analog **12** is taken) is modeled into three-dimensional representation (c). Each analog (a) is modeled using the MM+ force field, and partial atomic charges are projected on the respective atoms defined by three-dimensional HM (c) giving for each molecule a $HM_i$ pattern. Hydrogen-neglected $HM_i$ structures are transformed into SOM maps (d). The mapping procedure is detailed in further illustrations.

series is constructed to form a hypothetical two-dimensional structure capable of defining each molecule of this series. Figure 1 illustrates a procedure including such a transformation which will be called by definition a "*hypermolecule*" (HM). Thus, two-dimensional representations of all molecules (Figure 1a) of the data base are superimposed, atom per atom, which gives a two-dimensional HM pattern (Figure 1b) forming a topological basis for describing each analyzed analog. Formally, a HM pattern can be coded by an *n*-element all ones vector (where *n* is equal to the number of the atomic position within HM), while an *i*th molecule is given by an *n*-element $HM_i$ vector consisting of ones and zeros that code the atoms which exist or do not exist in the *i*th molecule, respectively. After defining a certain stereochemistry a hypermolecule HM can be modeled into a three-dimensional structure (Figure 1c). Each molecule is also modeled into a three-dimensional structure (these structures are not shown in Figure 1), and the respective distribution of the atomic charges is calculated as described in the part Model Building. These partial atomic charges are projected on the respective atomic positions defined by a three-dimensional $HM_i$ patterns (Figure 1c). Such patterns are transformed to the respective SOM maps (Figure 1d). We will show in this work that a transformation described above can be used effectively for explaining some molecular phenomena engaged in molecular recognition.

## RESULTS AND DISCUSSION

**Steroids Data.** The 31 steroid molecules specified in Table 1 are complexed by the CBG and TBG proteins. The rigid steroid skeleton makes the series an interesting object for the similarities studies. In fact, many different approaches attempted at modeling both affinities.[16−19] Despite that, only CBG has been defined well enough to specify the shape as the factor limiting binding phenomena. Figure 2 illustrates the schematic view of such a receptor that can be described by a sterically well defined cavity which interacts with a molecule at the certain positions.

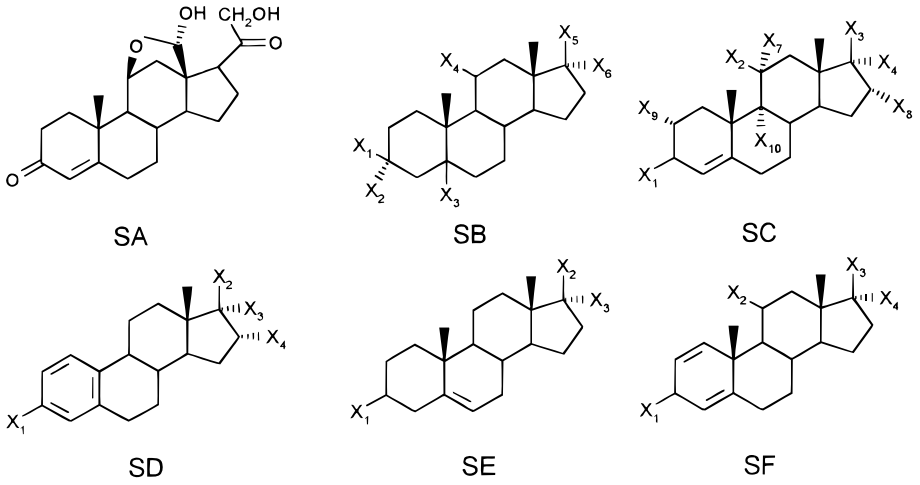
The TBG receptor site cannot be defined so easily, and it is not a coincidence that recent studies focused only on the CBG.[13,18] The similarity studies involving comparative SOM technique ranks the structures according to the CBG affinity but not the TBG one.[11,13]

**Model Building.** The structures of all compounds have been modeled by HYPERCHEM 4.0. Starting geometries were obtained from a 3D model builder. Then, each molecule was optimized using the MM+ force field, and partial atomic charges were calculated with the AM1 method. The HM (Figure 1), assumed to be a hydrogen-neglected skeleton capable of defining each atom of the hydrogen neglected steroid structures **1**−**21**, i.e., the analogs of the reported TBG affinities (compare Table 1), is given the stereochemistry of the steroid skeleton consistent with that of the molecule **12** and modeled into three-dimensional structure as all other molecules. The structures modeled by HYPERCHEM were additionally compared with the data obtained from Gasteiger et al.,[18] which includes some recent revisions of the steroid structures reported in previous investigations. The latter were modeled by a 3D builder of CORINA, while partial atomic charges were calculated by PETRA.[20] As in previous studies each steroid was represented by a single conformation.

**Neural Procedures.** A comparative SOM technique detailed in previous work[13] is used in this work to obtain a series of molecular SOM maps of the steroids **1**−**31**. Figure 3 illustrates the idea of the net which should simulate the operation of the receptor site.

Modeling any ligand-receptor interactions usually poses problems due to the ill-defined topology of the real receptors. Therefore, it seems reasonable to assume that the most active analog simply represents the receptor. It is somewhere near its geometry where optimal interactions can take place. The 3D geometry of the most active molecule, defined by the atomic coordinates, makes a reference SOM net represented in Figure 3 by a weight matrix $W_T$. Such a matrix transforms a molecule, defined by a set of atoms, into its SOM pattern. Therefore a series of comparative SOMs can be produced from the projections of partial atomic charges from each analogs. The projection of partial atomic charges, indicated in Figure 3 with the label "coloring" (for a template molecule) with the respective $S_i$ signals (where *i* means *i*th molecule), can go on with summing or averaging signals

**Table 1.** Structure and CBG/TBG Affinity Data for the Steroids Series **1**−**31** 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 lg 1/*K* | TBG lg 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 | $COCH_2OH$ | H | | | H | H | H | H | −7.881 | −6.342 |
| 7 | SC | =O | OH | $COCH_2OH$ | OH | | | H | H | H | H | −7.881 | −6.204 |
| 8 | SC | =O | =O | $COCH_2OH$ | OH | | | | H | H | H | −6.892 | −6.431 |
| 9 | SE | OH | =O | | | | | | | | | −5.000 | −7.819 |
| 10 | SC | =O | H | $COCH_2OH$ | H | | | H | H | H | H | −7.653 | −7.380 |
| 11 | SC | =O | H | $COCH_2OH$ | OH | | | H | H | H | H | −7.881 | −7.204 |
| 12 | SB | =O | | H[b] | H | OH | H | | | | | −5.919 | −9.740 |
| 13 | SD | OH | OH | H | H | | | | | | | −5.000 | −8.833 |
| 14 | SD | OH | OH | H | OH | | | | | | | −5.000 | −6.633 |
| 15 | SD | OH | =O | | 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 | $COCH_2OH$ | OH | | | | | | | −7.512 | ?[e] |
| 23 | SC | =O | OH | $COCH_2OCOMe$ | OH | | | H | H | H | H | −7.553 | ?[e] |
| 24 | SC | =O | =O | COMe | H | | | | H | H | H | −6.779 | ?[e] |
| 25 | SC | =O | H | $COCH_2OH$ | 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 | $COCH_2OH$ | OH | | | H | H | Me | H | −7.688 | ?[e] |
| 31 | SC | =O | OH | $COCH_2OH$ | OH | | | H | H | Me | F | −5.797 | ?[e] |

[a] Structures and data according to refs 17 and 18. [b] Of the 5-α steroid series. [c] Of the 5-$\beta$ steroid series. [d] Assumed to be =O (testosterone) as indicated in ref 18 and in Figure 6[6] and not as −OH in tables[6] and further publications (compare ref 18 for mistakes in previous publications). [e] Unknown. [f] H (hydrogen) instead of Me at $C_{10}$ steroid skeleton.

**Figure 2.** A schematic view of the CBG receptor defined by a sterically defined cavity and an interacting molecule (· − · − ·).

coming into the individual SOM neuron. These manipulations transform the SOM patterns into the corresponding summed or averaged $P_i$ matrices. The normalization achieved from the procedure is of key importance; it makes possible the comparison of the molecules that can vary in size. Consequently, the resulted $P_i$ patterns can be analyzed further. Thus, the map of the template molecule $\mathbf{P}_T$ models the receptor geometry and forms the training set for the single unsupervisedly trained neuron. Practically, $\mathbf{P}_T$ of the size 3 by 3 represented by a $3 \times 3 = 9$ element vector is replicated by the addition to a set of a noisy 5000 random nine element vectors, which yields a training set (TRM) for a single neuron labeled as N. A noise added amounts on average to 10% of the original vector. In fact, the original $\mathbf{P}_T$ vector has never been presented to a neural network. Instead, its noisy analogs are used to train an associative memory neuron with unsupervised protocol of the instar rule.[21] Such a neuron

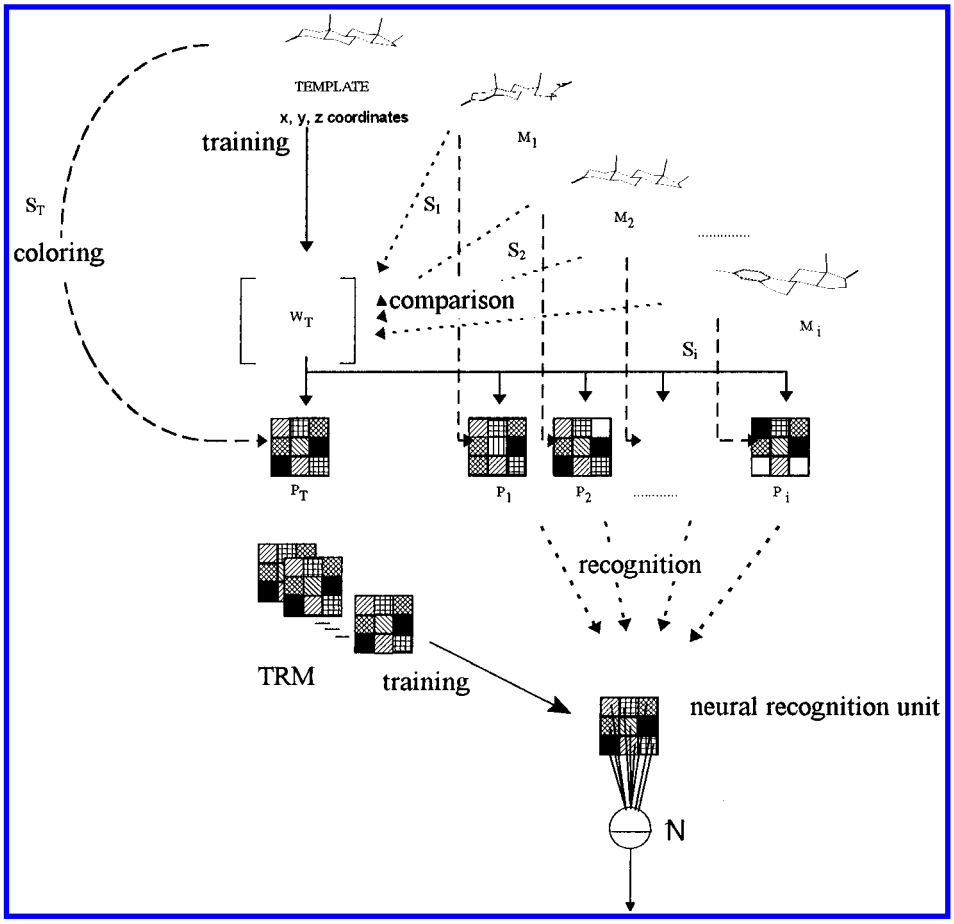



**Figure 3.** A scheme of the assembly of a neural net organized to mimic a receptor site affinity. The procedure is detailed in text.

**Table 2.** Correlation Coefficients of the CBG Models Obtained for Different Templates

| no. | template molecule | correlation coeff $R$ | std devtn $s$ |
|---|---|---|---|
| 1a | **6** | 0.89 | 0.49 |
| 1b | **6** | 0.89 | 0.50 |
| 2a | **6**, **7**, **11** | 0.92 | 0.43 |

[a] Compounds modeled with (Hyperchem). [b] Compounds modeled with 3D model builder CORINA (PETRA).[20]

shapes a unit capable of estimation of similarity between the tested and the template objects, signaling the value close to 1 when recognizing a template-like vector, while 0 corresponds to completely uncorrelated ones. Technically neural procedures were programmed using MATLAB (MS-WINDOWS) environment. The MATLAB NN-Toolbox procedures of unsupervised learning with the instar rule (only slightly different from the Kohonen protocol)[21] were used. Neighborhoods were simulated by the grid distance.[13,22] The "m-files" (scripts) programming the above mentioned procedures are available as Supporting Information.

**Shape Effects.** The three compounds of the highest CBG affinities are analogs **6**, **7**, and **11**. Table 2 specifies the correlations between the observed CBG affinities and the signals generated by the neuron trained with the template shaped after molecule **6** which gives the best model among analogs **6**, **7,** and **11** (models 1a and 1b). It is worth noticing that the structures modeled by CORINA (model 1b) and HYPERCHEM (1a) give almost exactly the same correlation coefficient, $R = 0.89$, and standard deviations $s = 0.49$ (1a) or $s = 0.50$ (1b). Averaging the signals from a few repeated neural procedures, with templates **6**, **7**, and **11** applied

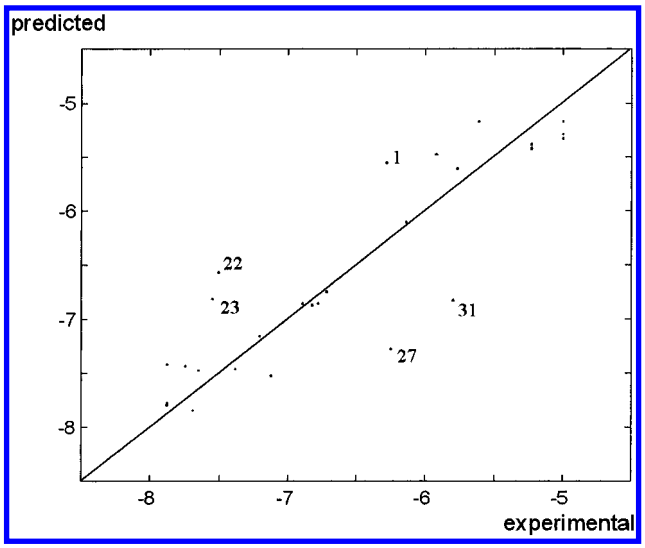



**Figure 4.** Experimental CBG affinities (pK) plotted against the predicted ones, according to model 2 (Table 2). The comparative $P_i$ maps were calculated from the average signals coming into the output neurons. Values from the range of $0 \div 1$ that were originally signaled by neural net were linearly scaled into the CBG range. Only the compounds of the least fit are labeled with numbers.

subsequently, only slightly improves the correlation (model 2 vs 1). Figure 5 illustrates the relationship given by the model 2. The coefficient $R = 0.92$ with the standard deviation $s = 0.43$ describes the correlation. The obtained correlation parameters indicate the predictive power of the model because no information on the actual compounds activity, but only the reference target topology is given to the network during the training step. The models obtained

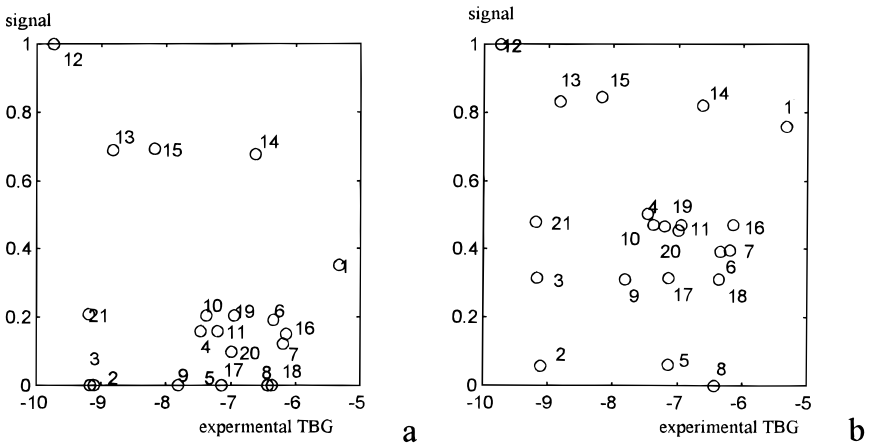



**Figure 5.** Signals generated by the neuron N, if plotted against the experimental TBG affinities for molecular matrices $P_i$ resulted from the procedure using the template **12** calculated from (a) sum of signals and (b) an average signal (details in text). Steroid molecules **1**−**21** are shown only, as the affinity of other anologs (**22**−**31**) is unknown.

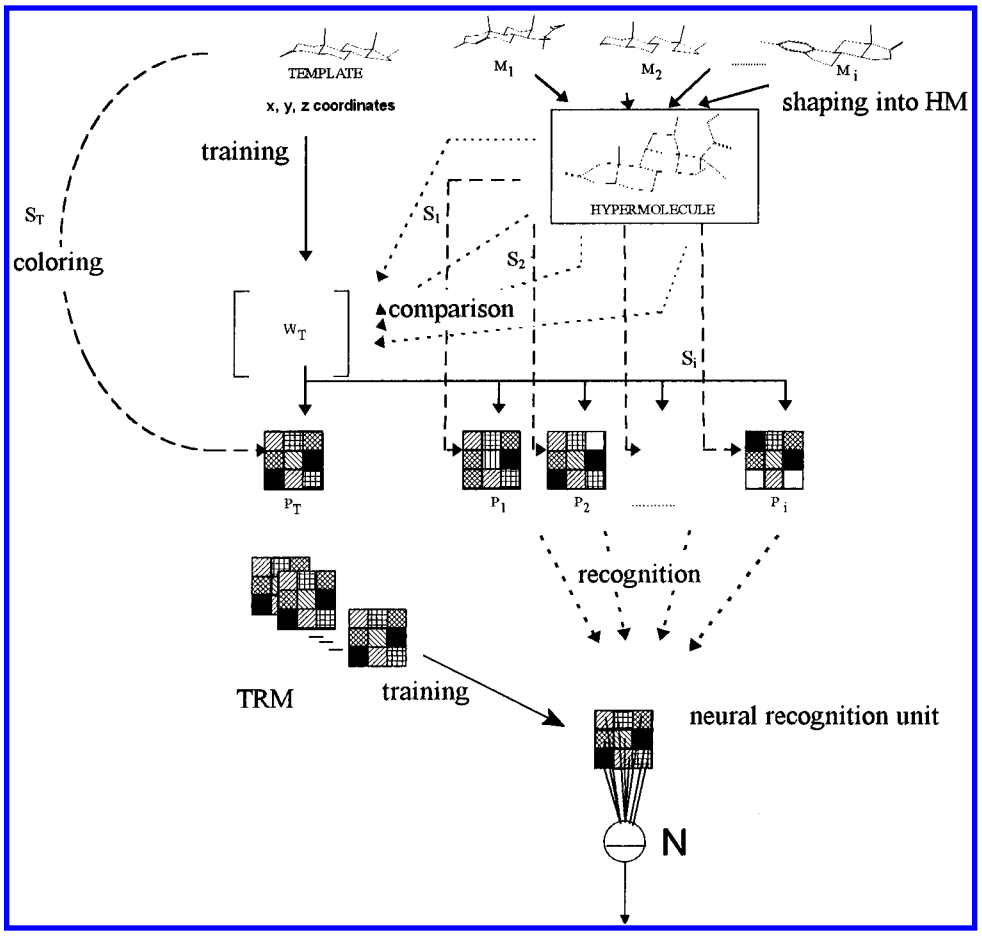



**Figure 6.** A scheme of the procedure modified to study the effects resulting from the charge distribution within the molecule. $W_T$ is trained with a target molecule **12** of the highest affinity, but all analogs are now transformed into the stereochemistry consistent with that of a target molecule by a hypermolecule **HM**, as detailed in text.

are comparable to those based on the neural backpropagation algorithm with cross-validation, involving autocorrelation function ($R$ = 0.92, standard deviation, $\sigma$ = 0.44, after rejection of one outlier).[18] Now, however, only the activity of one compound (three compounds in model 2) is needed to predict the affinities for all other analogs, while the previous method always needs all but one analog to find the affinity of that one.

The procedure designed naturally mimics the operation of the real receptor. A geometry of the most active analog is used to define the receptor-like shape for training neural network. The degree of similarity of other analogs displayed by such a network complies with the actual CBG affinity, which clearly indicates that it is the shape of steroids (modified by the distribution of the charge within molecule) that limits the affinity of binding to the CBG receptor site.

**Electrostatic Effects.** Figure 5 shows the results obtained from the procedure discussed above for the TBG shaped neural unit. Compound **12** is now the analog of the highest affinity. The approach, however, completely fails to predict the actual TBG affinities, which indicates that shape effects do not limit the process. Figure 6 shows the procedure

**Figure 7.** A schematic view of the interactions of the flexible receptor unit capable of accommodating the shape of three molecules (• − • − •). The arrows indicates the possibility of fitting any of the analyzed molecules.

designed to study the effects coming from the charge distribution within the molecules. This procedure mimics a receptor capable of accomodating the shape of the individual stimulating molecule, e.g., the allosteric receptor, as shown schematically in Figure 7. The core of the method is still the same, but now a hypermolecule HM is constructed and modeled to a 3D structure of the stereochemistry consistent with that of the most active TBG analog: **12**. Real analogs **1**−**21** are modeled, and partial atomic charges are calculated for each molecule and projected on the atoms of HM, respectively. The latter operation can be represented by ascribing of the respective charges to all nonzero elements of the individual $HM_i$ vectors describing a real *i*th molecule. The weight matrix $W_T$ is trained with the atomic *x*,*y*,*z* coordinates coming from the molecule **12**. Next, the individual $HM_i$ molecules are processed by the template $W_T$ matrix to produce a series of comparative $P_i$ SOMs. That forms an input for the second neural layer, as described in the previous chapter. In Figure 8 the resulted signals generated for the individual $P_i$ by the N neuron are plotted against the actual TBG affinity. A linear relationship of a very high correlation can be observed for the compounds of the highest activities (**12**, **2**, **3**, **9**, **13**, **15**, **21**). Also compound **1** of the lowest affinity fits the model (Figure 8a).

It should be realized that a transformation defined by "hypermoleculing", depriving all analogs of their original stereochemistry, makes the charge distribution within a molecule the only observable parameter. The manipulation clearly includes the shape only as a weak side effect., i.e., the larger the molecule the more signals will be transmitted into the net. A more careful analysis of the plot (Figure 8) reveals its V-like shape. An almost perfect straight-line plot will result if one reflects the signals provided by the neuron for the medium activity level across the *y* (signal) = 0.96 line. The overrecognition of the bare charge patterns signaled by the neural net within medium range probably indicates that, in fact, the electrostaticity compensates the shape misfit. Thus, it can be speculated that while the bare electrostaticity can explain the TBG affinity within the highest and lowest (analog **1**) activity range, the shape is probably responsible for the decrease in binding TBG within the medium area.

**Electrostatic and Shape Effects.** Unfortunately, the relationship illustrated in Figure 8 does not provide any predictive value. We need to include shape effects to model the real TBG receptor site. Below we present a slightly modified procedure including the HM transformation, which properly models the TBG affinity. Let us consider once more Figure 7 to explain the basis of the procedure. The core of the method is unchanged. Now, however, the template molecule is selected individually for each stereochemical subseries, to comply with the stereochemistry of this subseries. Real analogs of this subseries, after transformation through the HM, form comparative $P_i$ maps. The map $P_{12}$ which resulted from the procedure is always selected to provide the target pattern for training the second neural layer that gives the information about an optimal shape. E.g., template **6** gives $W_T$ for comparison of the analogs **1**, **4**, **6**, **7**, **8**, **10**, **11**, **19**, **20**, **21**, and the resulted comparative $P_{12}$ trains the N neuron. The procedure is repeated to cover all possible subseries, while the template molecules are selected randomly from the individual subseries.

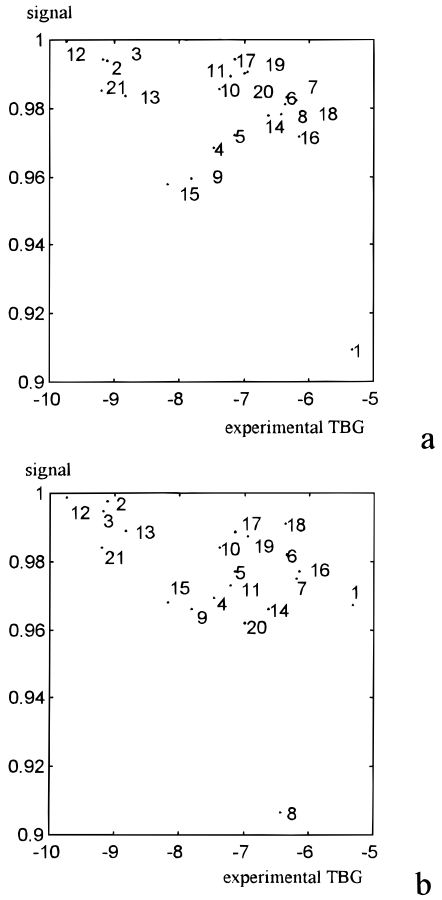



**Figure 8.** Signals generated by the N neuron and plotted against the experimental TBG affinities for molecular maps (matrices) given by the comparison of the individual $HM_i$ with the template **12**. The maps are colored with (a) sum of signals and (b) an average signal. Steroid molecules **1**−**21** are shown only, as the affinity of the anologs (**22**−**31**) is unknown.

Figure 9 plots the generated signals against experimental TBG affinities for all analogs when compared with templates **3**, **6**, and **13**, respectively. Signals of the analogs of the stereochemistry consistent with that of each template (**3**, **6**, and **13**) are now selected from individual plots shown in Figure 9, linearly scaled to the TBG units (by means of linear regression), and plotted in Figure 10, against the actual TBG affinities. Compound **16** which is of the unique stereochemistry is excluded from the procedure. The correlation coefficient ($R = 0.84$, $s = 0.67$) also indicates a predictive power of the model, since no information on the compounds activity but only the target topology and charge distributions is given to the network. The high predictive power of the model makes it comparable to the best ones from the literature,[17] which justifies the rationality of the approach.

The obtained models indicate, moreover, that it is a bare electrostaticity that limits the high and low TBG affinity.

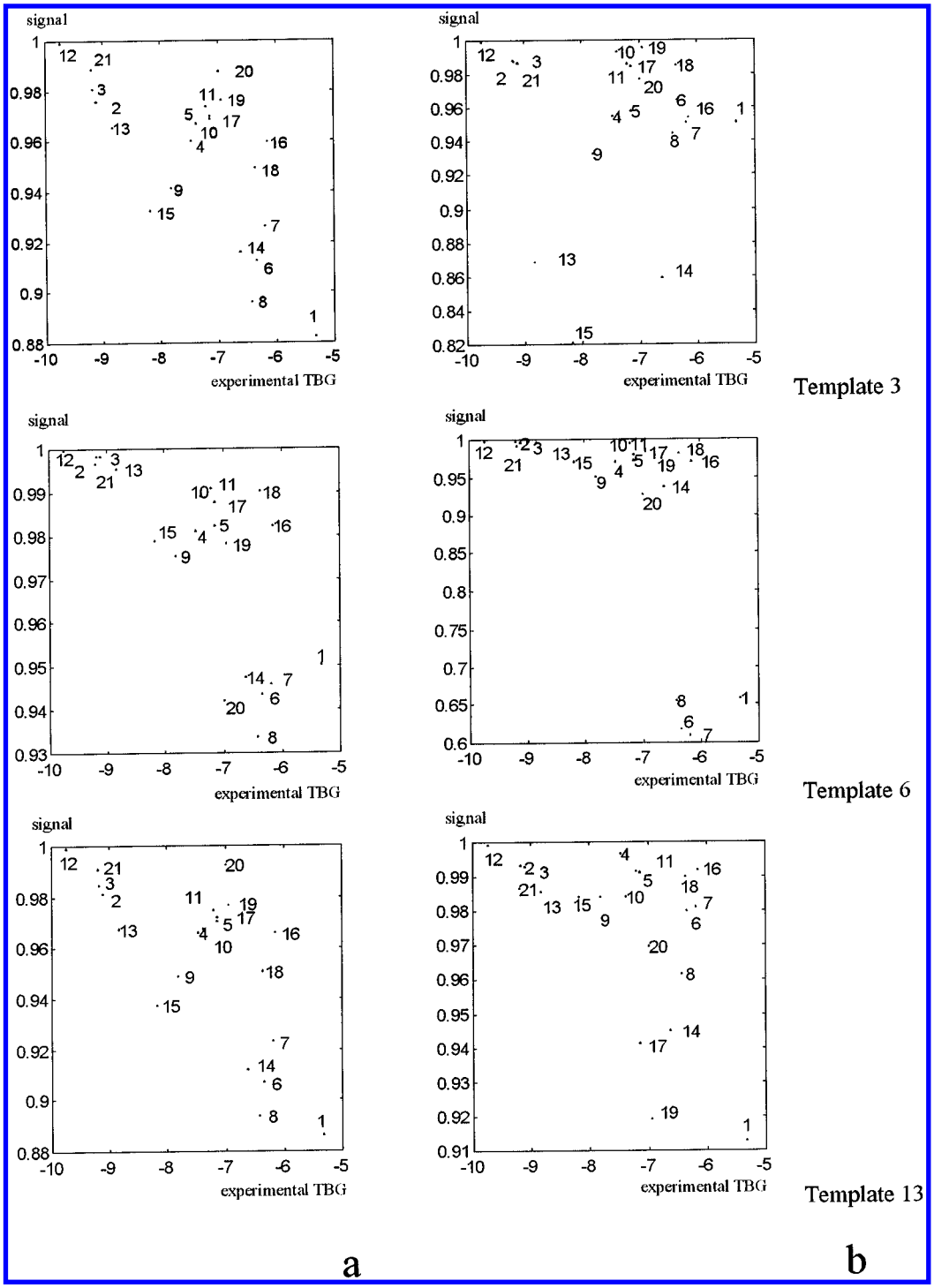



**Figure 9.** Signals generated by the N neuron and plotted against the experimental TBG affinities for molecular maps (matrices) given by the comparison of the individual $HM_i$ with the templates of **3**, **6**, and **13**, respectively. The maps are colored with charge (a) sum of signals and (b) an average signal. Steroid molecules **1**−**21** are shown only, as the affinity of other anologs (**22**−**31**) is unknown.

For the p*K* values below ca. −8 the affinity is limited only by the distribution of charge within a molecule, while within the medium range, $-7.5 \div -6$, shape also contributes to the affinity. Figure 11 shows the model constructed according to the above mentioned rules. The affinity of highs **12**, **2**, **3**, **9**, **13**, **15**, **21** and low **1** (indicated with black filled circles) is controlled by the electrostatic factors, as predicted in Figure 8a, while that of the mediums (indicated with white circles) is ruled by models shown in Figures 9a, respectively. Although the high correlation obtained ($R = 0.92$, $s = 0.48$) supports the conclusions, it cannot be used for prediction purposes, because one cannot determine a priori whether compounds interactions could be limited by steric or electrostatic factors. The analysis performed gives a clear insight into the molecular mechanism which explains the complexing by the TBG. If one compares the actual compounds structures, it is no steric hindrance at $C_{17}$; $\beta$-OH at $C_{17}$ and not α-OH, but $\beta$-OH or =O at $C_3$ that are the most visible determinants of high TBG affinity.

**Pseudoreceptor Modeling.** Unlike many other comparison methods, the one presented here not only provides the analytical model but also gives useful information for studying recognition processes taking place at the receptor sites. Since all information is analyzed without knowledge

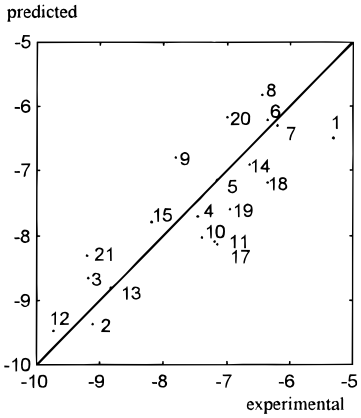



**Figure 10.** Signals generated by the N neuron plotted against the experimental TBG affinities. The signals are selected from the models shown in Figures 8 and 9, depending upon the stereochemistry of the main steroid skeleton, and scaled linearly in TBG units. Molecular maps (matrices) given by the comparison of the individual $HM_i$ with the templates of **3**, **6**, **12**, and **13**, respectively, for the analogs of **3: 3**, **9**, **17**, **18**; **6: 1**, **4**, **6**, **7**, **8**, **10**, **11**, **19**, **20**, **21**; and **12: 2**, **5**, **12**; **13: 13**, **14**, **15**. The maps result from coloring by charge (a) sum of signals (b) an average signal. Steroid molecules **1**−**21** are shown only, as the affinity of other anologs (**22**−**31**) is unknown: $R = 0.84$; $s = 0.67$.

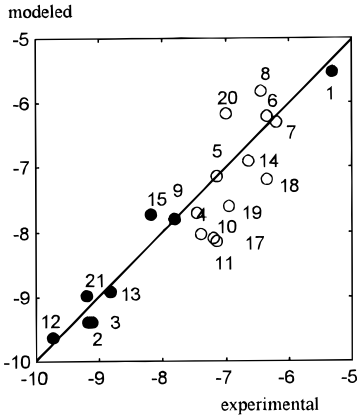



**Figure 11.** TBG affinities modeled by a receptor-like procedure. Black circles indicates the ranges limited (and modeled) by electrostatic effects only, while white circles, that are determined both by shape and electrostatic effects: $R = 0.92$; $s = 0.48$.

of the real receptor site, such techniques are sometimes called pseudoreceptor modeling.[23]

The CBG receptor can be shown as a well defined sterically restricted cavity that in fact was already postulated in the literature.[16] Such a site does not provide a good model for TBG. The pseudoreceptor should rather be seen as a unit capable of accomodating the shape (stereochemistry) of the individual compounds. Electrostatic effects are of the main importance, but steric effects, coming both from shape fit/misfit and steric hindrance (of the $C_{17}$ groups), also contribute to the recognition processes. Since the $P_i$ matrices described by the sum of charges provide slightly better correlations, it can be speculated that also electrostatic interactions are not strictly sterically defined. Consequently, it is rather a sum of charges within individual neurons (for similar molecules decided by the size of the molecule) that better describes the affinity. Although the adjustment of the ligands and pseudoreceptor can also be achieved by the manipulation with the steroid conformations, such a manipulation would evidently change the distribution of the atomic charges within stimulating molecules. This situation

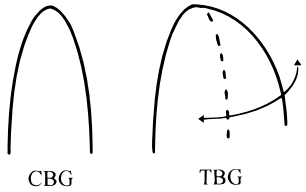



**Figure 12.** A scheme of the CBG and TBG pseudoreceptor sites. An arrow indicates the possibility of the conformational adjustment of the TBG receptor site.

can also be modeled by a presented method but if tried did not give the results complying with the TBG affinities observed.

Thus, a model satisfying these requirements is shown in Figure 12. It seems that a pseudoreceptor site is capable of accomodating the shape of ligands, which is schematically shown by a receptor with allosteric effect. It is clear now that a neural procedure presented completely mimics an operation of such a unit. The operation defined by the HM imitates the adjusting of the ligand and the receptor, while the comparison of the template molecules and HM estimates steric fit/misfit for the individual subseries.

## CONCLUSIONS

An assembly of two unsupervisedly learned neural nets was designed to mimic the operation taking place within two different receptor sites of the CBG and TBG. The architecture applied correctly analyzes the similarities between the reference and tested molecules, providing good models of affinities. The method gives clear insight into the molecular mechanism of the recognition processes. Thus, the contribution of the electrostatic and steric effects can be nicely separated, and reasonable CBG and TBG pseudoreceptors can be concluded from the models obtained.

## ACKNOWLEDGMENT

The author thanks Prof. Johann Gasteiger from the University of Erlangen, BRD for supplying the coordinates of the steroids, which makes possible verification of the structures modeled, and Dr. Andrzej Polański from the Institute of Automatic Control of the Silesian Technical University Gliwice for providing the possibility of performing some computational operations.

**Supporting Information Available:** The MATLAB "m-files" that perform the procedures discussed. See any current masthead page for Internet access instructions.

## REFERENCES AND NOTES

(1) *Molecular Similarities I, Topics in Current Chemistry*; Sen, K., Ed.; Springer: Berlin, 1995; Vol. 173.
(2) *Molecular Similarities II, Topics in Current Chemistry*; Sen, K., Ed.; Springer: Berlin, 1995; Vol. 174.
(3) Kier, L. B. Indexes of Molecular Shape from Chemical Graphs. *Med Res. Rev.* **1987**, *7*, 417−440.
(4) 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.
(5) 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 Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1983**, *26*, 990−996.
(6) 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.

(7) Kohonen, T. *Self-Organization and Associative Memory*; Springer: Berlin, 1988.
(8) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists—An Introduction*; VCH: 1993; pp 285−291.
(9) 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.
(10) Gasteiger, J.; Li, X. Representation of the electrostatic potentials of muscarinic and nicotinic agonists with artificial neural networks. *Angew. Chem*. **1994**, *106*, 671−674.
(11) Polański, J.; Gasteiger, J. The Comparison of Molecular Surfaces by an Assembly of Self-Organizing Neural Network. *Computers in Chemistry '94*, Technical University of Wrocław, Wrocław, National Institute of Standards and Technology: Gaithersburg, MD, 1994; p 88.
(12) Barlow, T. W. Self-organizing Maps and Molecular Similarity. *J. Mol. Graphics* **1995**, *13*, 24−27.
(13) Polański, J. Neural Nets for the Simulation of the Molecular Recognition within MS-Windows Environment. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 694−705.
(14) Simon, Z.; Szabadai, Z. Specific Interactions. Intermolecular Forces, Steric Requirements, and Molecular Size. *Angew. Chem*. **1974**, *86*, 802−811.
(15) Motoc, I. Molecular Shape Descriptors. In *Steric Effects in Drug Design*; Charton, M., Motoc, I., Eds.; Spriger Verlag: Berlin, 1983; pp 93−105.
(16) Mickelson, K. E.; Forsthoefel, J.; Westphal, U. Steroid-Protein Interactions. Human Corticosteroid Binding Globulin: Some Physicochemical properties and Binding Specificity. *Biochemistry* **1981**, *20*, 6211−6218.
(17) 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.
(18) 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.
(19) Good, A. C.; So, S. S; Richards, W. G. Structure−Activity Relationships from Molecular Similarity Matrices. *J. Med. Chem.* **1993**, *36*, 433−438.
(20) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Atomic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(21) Demuth, H.; Beale, M. *Neural Network Toolbox. For Use with MATLAB*; The MathWorks, Inc.: Natick, MA, 1994; pp 8.1−8.22.
(22) Melssen, W. J.; Smits, J. R. M.; Buydens, L. M. C.: Kateman, G. Using Artificial Neural Networks for Solving Chemical Problems. Part II. Kohonen Self-Organizing Feature Maps and Hopfield Networks. *Chemom. Intell. Lab. Syst.* **1994**, *23*, 267−291.
(23) Koehler, K. F.; Rao, S. N.; Snyder, J. P.; Modeling Drug-Receptor Interactions. In *Guidebook on Molecular Modeling in Drug Design;* Cohen, N. C., Ed.; Academic Press: San Diego, 1966; p 235.

CI960105E