Journal of Computer-Aided Molecular Design, 10 (1996) 521–534
ESCOM

J-CAMD 369

The comparison of geometric and electronic properties of molecular surfaces by neural networks: Application to the analysis of corticosteroid-binding globulin activity of steroids

Soheila Anzali[a], Gerhard Barnickel[a], Michael Krug[a], Jens Sadowski[b], Markus Wagener[b], Johann Gasteiger[b,*] and Jaroslaw Polanski[c]

[a]*Merck KGaA, Department of Medicinal Chemistry/Drug Design, D-64271 Darmstadt, Germany*
[b]*Computer-Chemistry-Center, Institute of Organic Chemistry, University of Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany*
[c]*Institute of Chemistry, University of Silesia, Ul. Szkolna 9, PL-40006 Katowice, Poland*

Received 4 April 1996
Accepted 22 July 1996

Keywords: Shape analysis; Kohonen maps; Molecular electrostatic potential; QSAR

Summary

It is shown how a self-organizing neural network such as the one introduced by Kohonen can be used to analyze features of molecular surfaces, such as shape and the molecular electrostatic potential. On the one hand, two-dimensional maps of molecular surface properties can be generated and used for the comparison of a set of molecules. On the other hand, the surface geometry of one molecule can be stored in a network and this network can be used as a template for the analysis of the shape of various other molecules. The application of these techniques to a series of steroids exhibiting a range of binding activities to the corticosteroid-binding globulin receptor allows one to pinpoint the essential features necessary for biological activity.

Introduction

The shape of a molecule and electronic properties such as the electrostatic potential strongly influence many chemical and biological properties of the molecule. The intuitively appealing picture of a key fitting into a lock for describing substrate–receptor interactions emphasizes the role of shape in biological activity. However, not only must the geometry of a molecule be appropriate for it to fit into a receptor, but also certain properties on the molecular surface such as hydrophobicity, the electrostatic or the hydrogen-bonding potential must fit to ensure an effective binding. Several approaches based on shape analysis have been developed to study structure–activity relationships within a series of individual compounds. In particular, the Comparative Molecular Field Analysis (CoMFA) approach [1] is widely used for such a comparison on the basis of molecular fields and properties like electrostatic potential. In the last decade, the three-dimensional comparison of molecular electrostatic fields and volumes was successfully applied in many cases to reveal effective structure–activity relationships [2].

The human brain is highly efficient in performing an analysis of three-dimensional objects and of properties on the surface of these objects such as shape, color, and texture. This is achieved by the generation of sensory maps of the environment in the visual, auditory, or somatosensory cortex. Artificial neural networks are models for the information processing in the human brain. The application of neural networks in chemistry has increased dramatically in recent years [3–5]. The neural network method developed by Kohonen [6–8] is rather efficient in modeling the generation of sensory maps in the brain. In the Kohonen network, the artificial neurons self-organize in an unsupervised learning process and thus can be used to generate topological feature maps. This property of a Kohonen neural network (KNN) has already been used for the mapping of molecular surface properties such as the molecular electrostatic potentials (MEPs) into two dimensions [9–11]. We will show here that a KNN can

*To whom correspondence should be addressed.

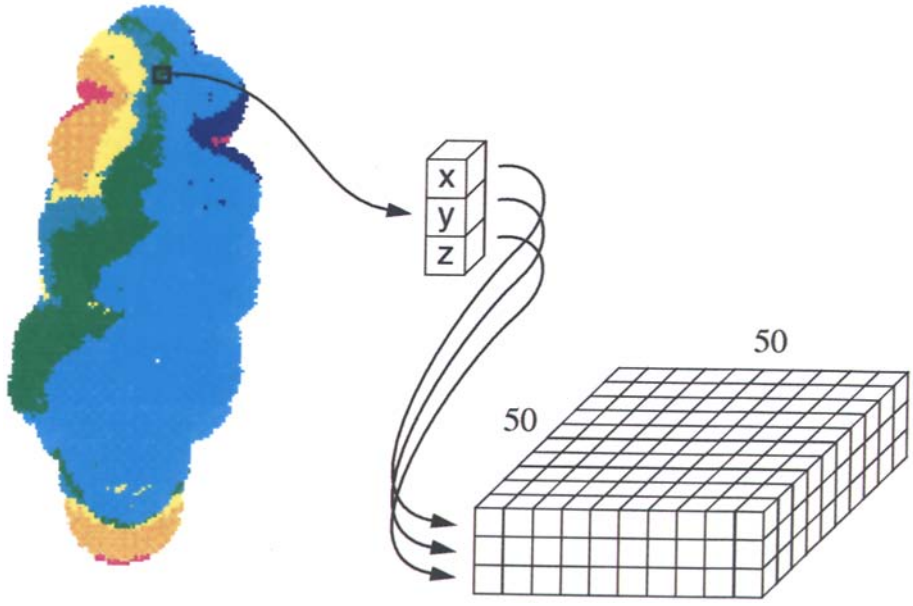

Fig. 1. Mapping of a molecular surface into a Kohonen neural network.

also be used to analyze the shape of molecules and give a quantitative definition of similarity for the geometry of molecular surfaces. In addition, the mapping of molecular surface properties by Kohonen networks [6–8] is further extended to allow a quantitative comparison of the molecular electrostatic potentials for a series of compounds. The combined application of a Kohonen network for both the analysis of the shape of a molecule and molecular surface properties provides deep insights into the foundations of biological activity within a set of compounds. We show this here with a data set that has already been studied by other groups and methods, a data set of 31 steroids exhibiting a range of activity for binding to the corticosteroid-binding globulin (CBG) receptor.

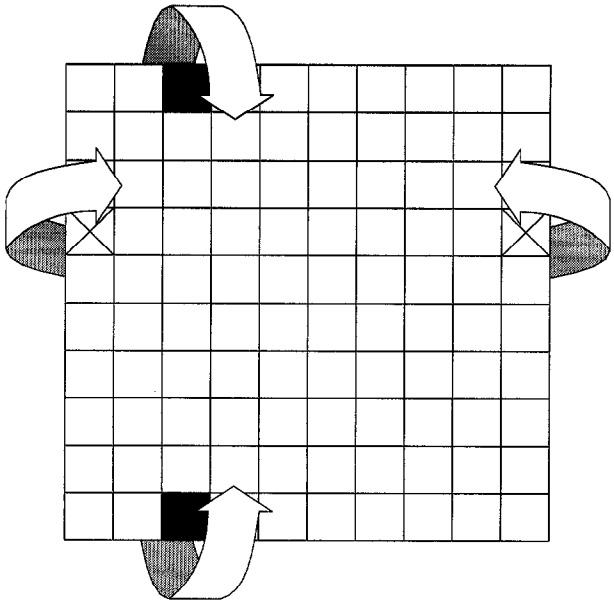

Fig. 2. Plane obtained by making two cuts into a torus. The cuts can be shifted into any direction. Thus, the two neurons marked by a cross are direct neighbors, as are the two neurons marked by black squares.

Materials and Methods

Mapping of surface properties

The KNN is a self-organizing network, which can be used to generate a nonlinear projection of objects from a high-dimensional space into a lower dimensional space, usually a two-dimensional plane. This method enables the decrease of the dimensions of the information space while conserving the topology of the information as best as possible. Details on Kohonen networks can be taken from the literature [5–8]; thus only the major features are repeated here. Learning in a Kohonen network is an unsupervised competitive process such that an object is mapped into a neuron. Each neuron has as many weights w_{ji} as there are input variables x_{si} for an object s. The winning neuron, c_s, will be the one that has weights that come closest to the input variables (Eq. 1):

$$c_s \Leftarrow \min\left[\sum_{i=1}^{m}(x_{si} - w_{ji})^2\right] \qquad (1)$$

The learning process adjusts the weights such that they become even closer to the input variables, but it does so with a factor that decreases with distance from the winning neuron.

We have shown that a Kohonen network can be used for the generation of maps of molecular surface properties by projecting these surfaces of three-dimensional objects into two dimensions [9–11]. For example, maps of

the molecular electrostatic potential (MEP) can thus be obtained. Input to the Kohonen network consists of the Cartesian coordinates of points on the molecular surface. With three input variables, each neuron also has three weights. The neurons of such a Kohonen network are arranged in two dimensions. In the projection of the molecular surfaces of the molecules studied here, an arrangement of 50×50 neurons was used. The surface of a molecule is continuous, without a beginning and ending. Thus, it seemed desirable to also have an arrangement of neurons without a beginning and ending. This can be achieved by arranging the neurons on the surface of a torus. For visualization, this torus is cut along two perpendicular lines and then the torus surface is spread into a plane. Since these cuts can be made at arbitrary lines, the maps can be shifted into any direction.

The method for obtaining two-dimensional maps of molecules consists of training the Kohonen network with the three Cartesian coordinates of points randomly sampled on the van der Waals surface of the molecule with a density of 100 points per Å^2. With a Kohonen network of 50×50 neurons, there are $3 \times 50 \times 50 = 7500$ weights. The architecture of the Kohonen network is shown in Fig. 1.

As the size of the network is much smaller than the size of the data set (number of points from the molecular surface), a neuron is normally excited by several points of the surface. In this case, the neuron obtains the average value of the information of those points that excite it.

The information on the molecular surface – the properties on the surface; in our case, the individual electrostatic values of each coordinate of the van der Waals surface – is not used in the learning process. Thus, during processing by unsupervised learning the points with similar coordinates are put close together by the network, into the same or adjacent neurons.

It is emphasized again that the neurons are actually arranged on the surface of the torus. In the planar maps of molecular surfaces shown in the figures that follow, it has to be remembered that they have been obtained by spreading this toroidal surface into a plane. Thus, neu-

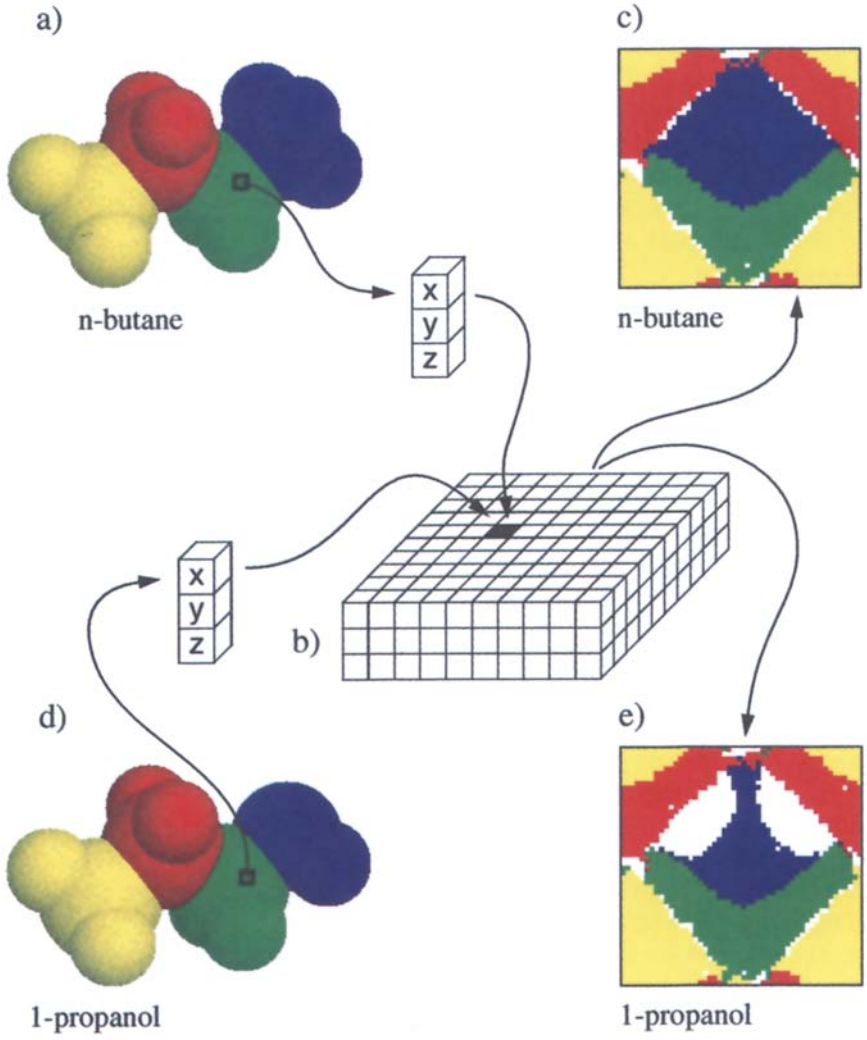

Fig. 3. The surface of a reference molecule (*n*-butane) is stored in a Kohonen network that can be visualized by a map. This network can be used as a template for comparison with the surface of a second molecule (propanol) leading to a compared Kohonen network.

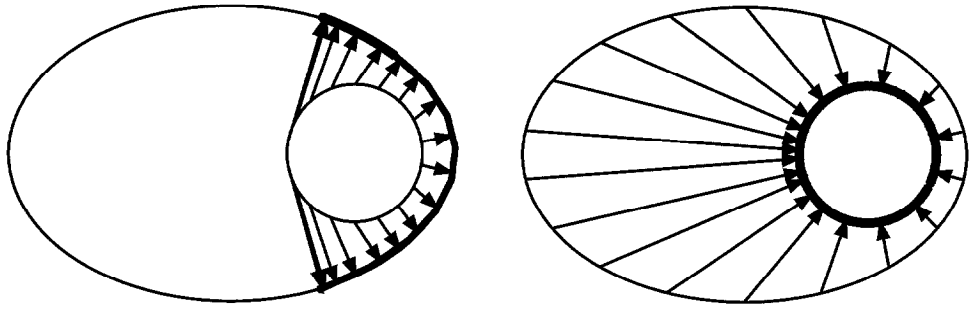

Fig. 4. Comparison of a circle and an ellipse using either the points on the ellipse (left-hand side) or the points on the circle (right-hand side) to train a reference Kohonen network. Points on the other shape, respectively, are sent through the reference network, resulting in different Kohonen maps.

rons at the right-hand column are directly adjacent to neurons at the left-hand column (cf. the neurons crossed in Fig. 2) and neurons in the top line are directly adjacent to neurons in the bottom line (cf. the two filled neurons in Fig. 2).

Clearly, the topology of a molecular surface, which is more like a sphere or an ellipsoid, is different from the topology of a torus. This must necessarily lead to topological distortions during the projection of the surface of a molecule onto the surface of a torus. During learning in a Kohonen network, the torus adjusts itself as best as possible to the molecular surface. Nevertheless, the topological distortions manifest themselves in neurons that are not used during projection [12]. These empty neurons show up as white areas in the resulting Kohonen maps as can be seen in the following figures.

Comparison of the shape of molecules

It should be realized that the weights of a Kohonen network trained by the Cartesian coordinates of points on the molecular surface basically store the geometry of the molecular surface. This fact offers the foundation for a new method of comparing molecular surfaces:

(1) A Kohonen network is trained with points of the surface of a molecule taken as a template (cf. *n*-butane in Fig. 3). The network thus obtained stores the coordinates of points on this surface and therefore can be taken as a template to compare the surfaces of other molecules.

(2) Points of the surface of another molecule (1-propanol in Fig. 3) are sent into this template network. If the points find a neuron with weights that are quite similar to the Cartesian coordinates of this point (cf. Eq. 1), they are mapped into this neuron. Points of the surface of the reference molecule that have no counterparts in the compared molecule stay empty and lead to white areas in the maps of the surface of the molecule to be compared with the reference molecule. The more the surfaces of the two molecules are different, the larger the number of empty neurons. Thus, the number of empty neurons can be taken as a quantitative measure of the difference between the molecule to be compared and the reference molecule.

When applying this technique, one has to take into account that the choice of the template structure largely influences the results. Even if only two compounds are compared, the outcome may be different depending on the chosen reference structure. This can easily be seen in Fig. 4. There are two shapes that are compared, a small circle and a much larger ellipse. On the left-hand side of the figure, a template Kohonen network was trained with points from the surface of the ellipse and then points from the circle were sent through this network. As they are all mapped into neurons that represent the right part of the ellipse, a large part of the corresponding Kohonen map will remain empty. A totally different result is obtained when the template network is trained with points from the surface of the circle and those from the ellipse are sent through the network afterwards. This results in a completely filled Kohonen map because every part of the ellipse corresponds to a different part of the circle. Therefore, changing the template structure may reveal details that would otherwise remain hidden.

Calculational procedure

The following procedure was applied in this study:

(1) The 3D structures of the molecules were built from the connection tables and stereochemical information derived from wedge and dot graphics using the 3D structure generator CORINA [13–15]. Because of the relative rigidity of the molecules studied here, steroids, each compound was represented by a single conformation, as produced by CORINA.

(2) Partial atomic charges were calculated by the empirical PEOE method [16,17].

(3) The molecular electrostatic potential (MEP) was calculated on the van der Waals surface by Coulomb's law using a unit positive charge probe at surface points randomly sampled with a density of 100 points per $Å^2$ and partial atomic charges as calculated by the PEOE method.

(4) A Kohonen network was trained with the Cartesian coordinates of the randomly sampled points.

(5) The trained neurons were then colored according to

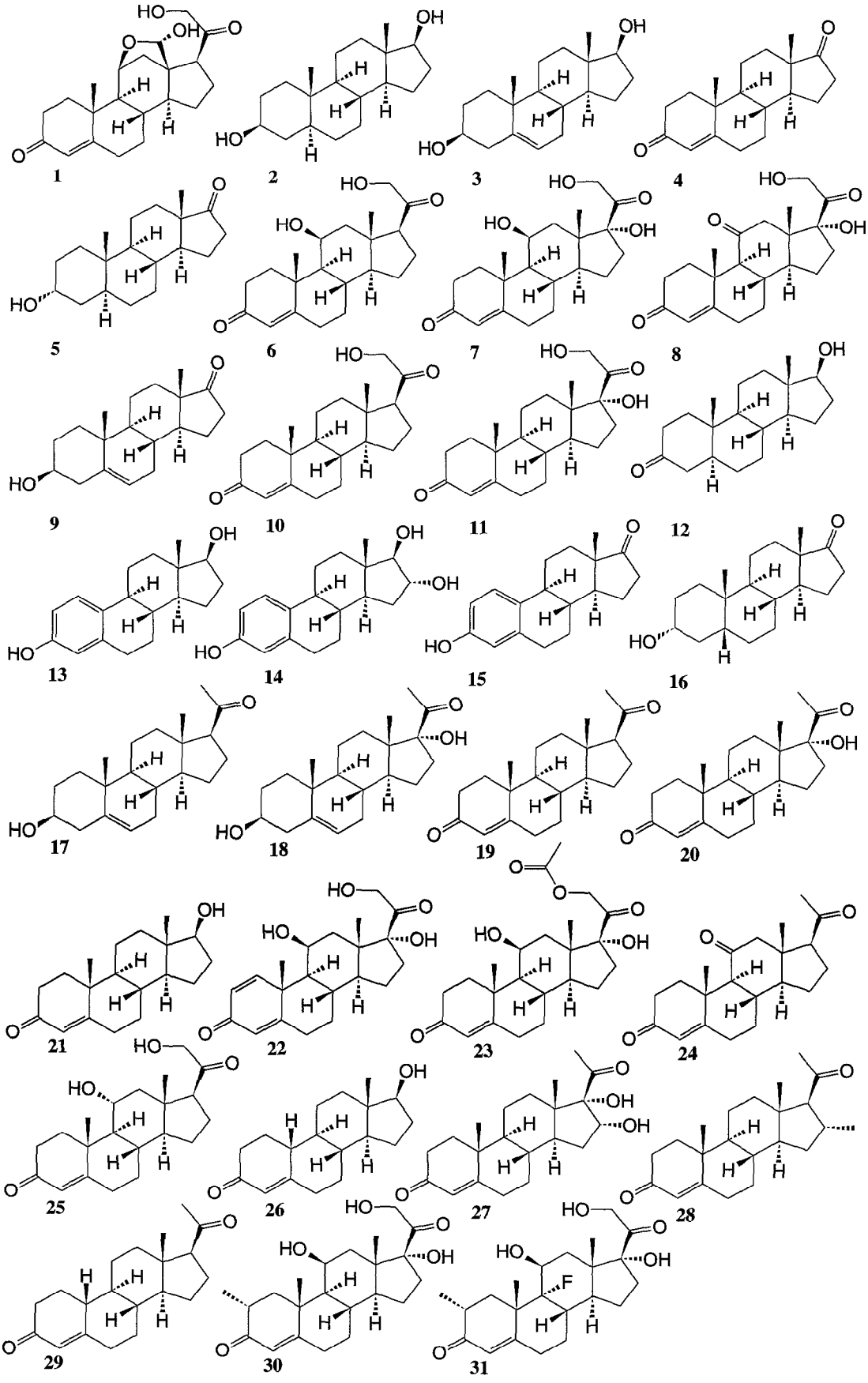

Fig. 5. Steroid structures of the data set used in this study.

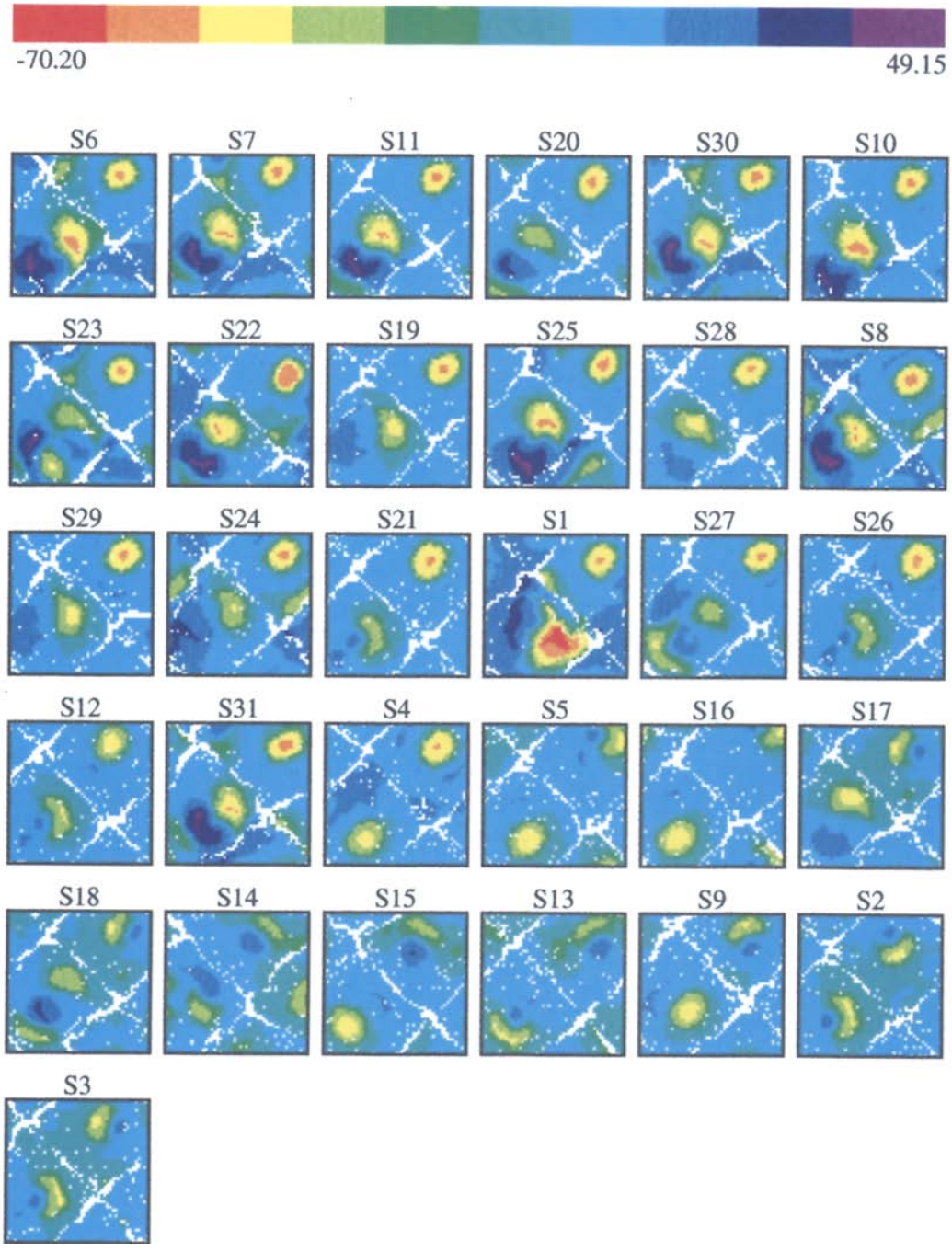

Fig. 6. The 31 Kohonen maps of the steroids of Fig. 5 colored by the MEP. The maps are ordered according to decreasing CBG affinity.

the MEP values existing at those points on the molecular surface that were mapped into a neuron. For this purpose, the arithmetic average of all the MEP values assigned to a neuron was calculated and one out of ten colors was given to that neuron according to the calculated average.

(6) The resulting Kohonen maps were aligned by manually shifting, mirroring, and rotating them by 90°.

(7) MEP patterns were compared within the Kohonen maps obtained from the individual molecules, the 31 steroids, of the data set.

(8) The steroids were divided into three different classes of activity: compounds with high affinity, intermediate affinity, and low affinity to the CBG receptor. For each class of compounds an averaged map was generated.

(9) Shapes of the molecules were compared with a reference molecule, the most active compound, by using the trained net of this reference molecule as a template.

(10) The weights of the Kohonen network were interpreted as Cartesian coordinates and the corresponding shape was plotted in 3D space. This procedure allows one to investigate in which way the areas in the 2D maps

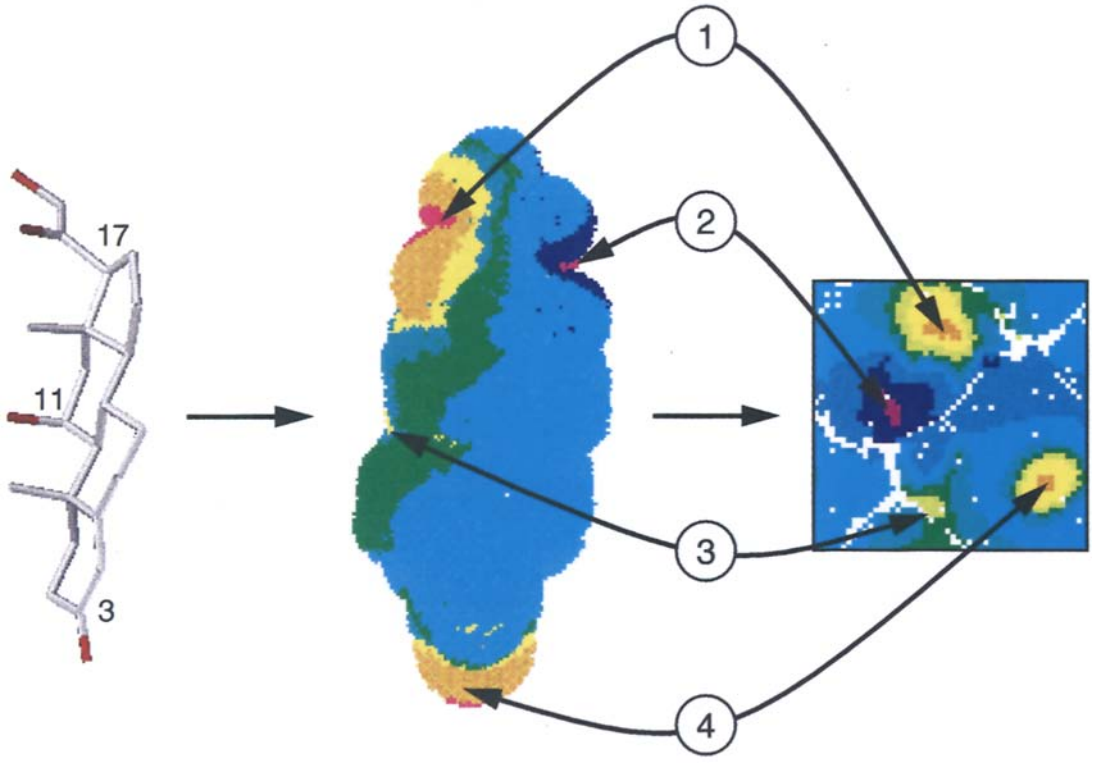

Fig. 7. Three-dimensional model of corticosterone, **6**; parallel projection of the electrostatic potential onto the van der Waals surface of corticosterone, and the corresponding Kohonen map indicating the MEP features corresponding to (1) the side chain at position 17 with a large negative value; (2) the area below the D-ring and the side chain at 17 with a large positive value; (3) the hydroxyl group at position 11b; and (4) the C3-carbonyl group with a conjugated double bond at C4.

correspond to the 3D shapes of the structures under consideration.

Data set

The methodology was tested with 31 steroids for which the CBG affinities were known [18–20]. This data set was chosen because it had been selected for the introduction of the widely used CoMFA method [1] and has also been studied with other methods [21–23].

A set of steroids and their binding affinities to human CBG (31 molecules) were extracted from the literature [1,21]. The structural formulas of the individual compounds are shown in Fig. 5. It has to be emphasized that most of the compilations of this data set widely distributed in printed or electronic form all contain coding errors. These errors have been eliminated in Fig. 5 by a reexamination of the original literature. The experimental binding affinity data are listed in Table 1. The distributions of the compounds in high, intermediate, and low affinity classes are defined as in Ref. 21.

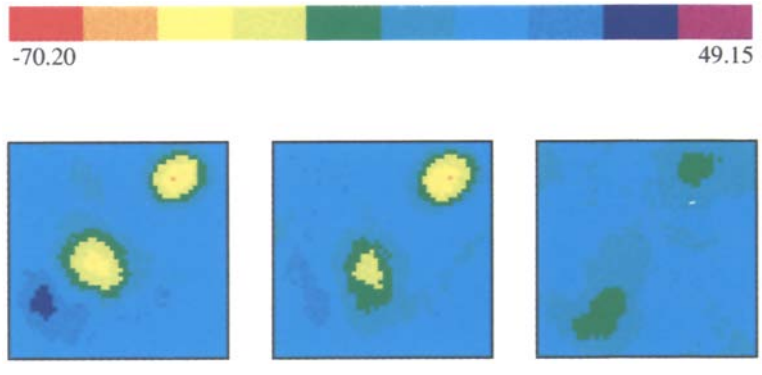

Fig. 8. The averaged maps of the electrostatic potential on the van der Waals surface of the sets of (a) high, (b) intermediate, and (c) low active compounds of the CBG series.

TABLE 1
CBG AFFINITY DATA FROM REF. 21

Compound	Affinity (pK)	Activity class[a]	Compound	Affinity (pK)	Activity class[a]
1	6.279	2	**17**	5.225	3
2	5.000	3	**18**	5.000	3
3	5.000	3	**19**	7.380	1
4	5.763	3	**20**	7.740	1
5	5.613	3	**21**	6.724	2
6	7.881	1	**22**	7.512	1
7	7.881	1	**23**	7.553	1
8	6.892	2	**24**	6.779	2
9	5.000	3	**25**	7.200	1
10	7.653	1	**26**	6.144	2
11	7.881	1	**27**	6.247	2
12	5.919	2	**28**	7.120	2
13	5.000	3	**29**	6.817	2
14	5.000	3	**30**	7.688	1
15	5.000	3	**31**	5.797	2
16	5.225	3			

[a] 1: high; 2: intermediate; 3: low; this classification was obtained by dividing the data set into three classes of comparable size.

Results and Discussion

Molecular electrostatic potential

The 31 trained Kohonen maps of the molecular electrostatic potential (MEP) of the steroid molecules (cf. Fig. 5) are shown in Fig. 6. The maps are ordered according to decreasing CBG binding affinity. For each molecule of the series a Kohonen net was trained, using the three Cartesian coordinates of points on the molecular surface as input to the network. The values of the electrostatic potential on the surface were not considered in the learning process; thus training was unsupervised. Only after the map was trained, a MEP value was allocated to each neuron by taking the MEP values of those points on the molecular surface that were mapped into this neuron. These values of the MEP determine the colors of the map. The maps can be compared directly to one another by using the shared color-palette.

Depending on the random initialization of the Kohonen network and the chosen learning parameters and the way in which the data points are presented to the network, as well as to the sites chosen for making the cuts into the torus, the Kohonen maps can be oriented in a variety of ways [9]. Therefore, the maps were aligned in order to achieve similar positions for all patterns, particularly the red–yellow colored spots (the most negative value of the MEP) and the violet spots (the most positive value of the MEP). This was done by manually shifting, mirroring, and rotating them by 90°. Basically, this corresponds to cutting the torus at different positions.

Details of the mapping of the MEP on the van der Waals surface of corticosterone, **6**, are shown in Fig. 7. This representation permits one to realize the correspondence between the spatial arrangement of the functional groups of the steroid and the colors in the 2D Kohonen map. Corticosterone, **6**, has two sites with a large negative value of the MEP, the carbonyl group at position 3 (4 in Fig. 7) and the side chain $COCH_2OH$ at position 17 (1 in Fig. 7). Consistent with this, the Kohonen map shows two areas with a red–yellow color for these sites. The spatial distance of these groups is reflected by two different shapes of the projection of the MEP into the Kohonen network. The third site with a negative value of the MEP stems from the hydroxyl group at position 11 (3 in Fig. 7). For this group, an area with a light green color is reserved in the map. Furthermore, the large positive MEP area of corticosterone is below the D-ring and the side chain at position 17 (2 in Fig. 7). The projection of the MEP into the Kohonen map indicates the location of this violet area close to the area of negative MEP caused by the $COCH_2OH$ side chain at position 17.

It can be noted that the Kohonen maps of Figs. 6 and 7 contain white spaces (white lines and spots), which

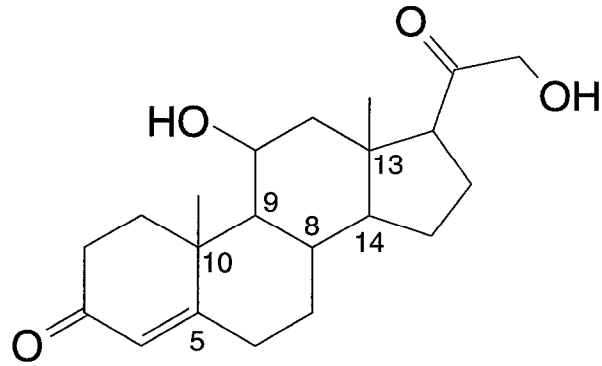

Fig. 9. The chemical structure of corticosterone, indicating the ring junction atoms used for the superposition of the steroid molecules.

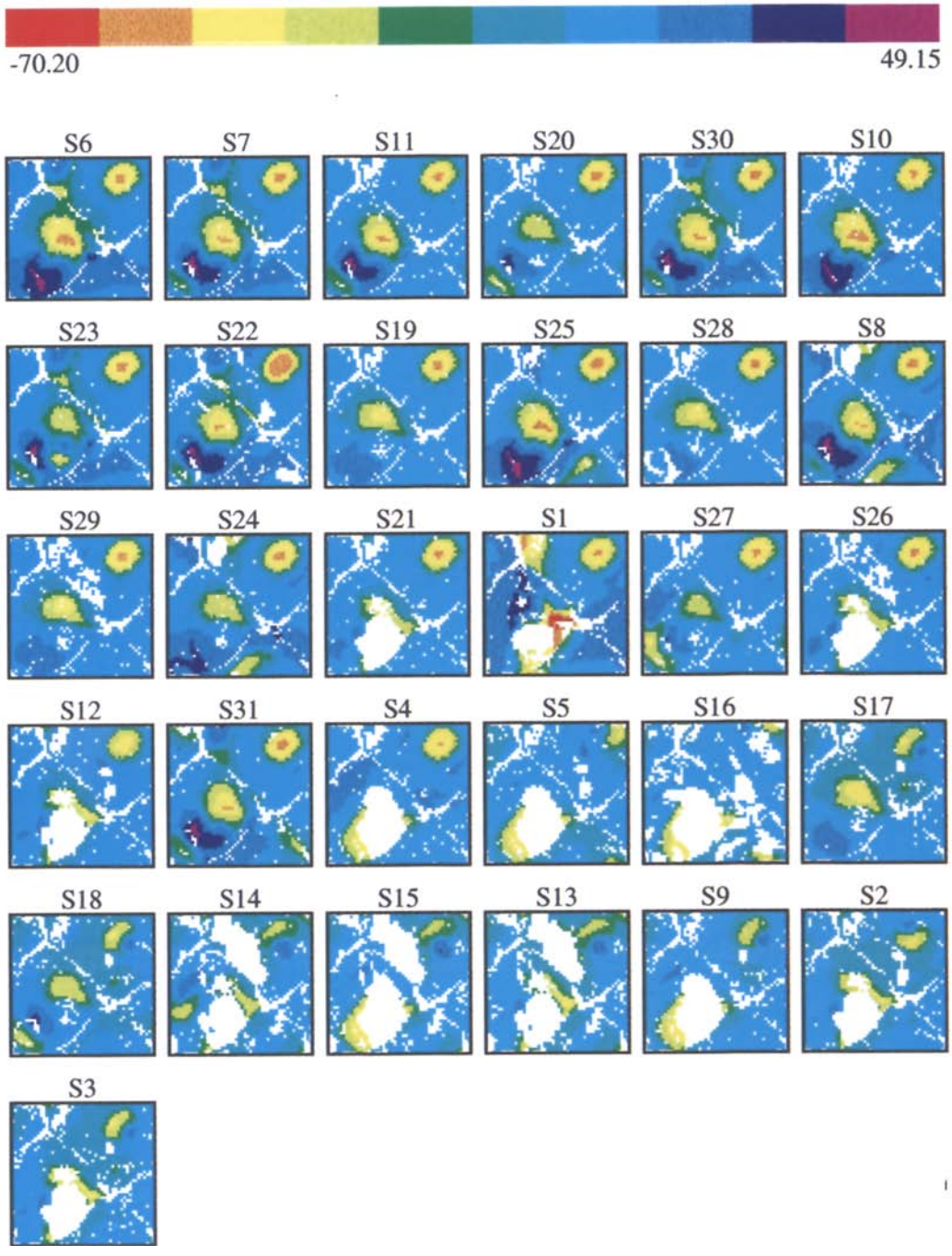

Fig. 10. Compared c-KNNs of steroids **1**–**31** (identified here as S1–S31) obtained by processing through the reference KNN of **6**, the most active CBG analogue.

correspond to empty neurons. These spots are the result of topological distortions [12] which are due to the difference in topology between a sphere and a torus and the fact that the molecular surface is more globular (sphere-like). It was found that these distortions try to avoid cutting through atoms [9].

As shown in Fig. 6 (see the upper row of the panel), the maps of highly active molecules usually contain these two red–yellow spaces (projections of the negative values of the MEP) and one violet space (positive values of the MEP) as well. A difference of the MEP pattern between the most active in the top rows and the least active in the bottom rows is clearly visible. Compounds with higher affinity show larger negative (red–yellow) and positive values (violet) of the MEP compared to compounds with lower affinity.

For a more objective analysis, the averaged maps for the sets of high, medium, and low active compounds (see Table 1) were generated (cf. Fig. 8). For this purpose, each neuron in the Kohonen maps of the single compounds

was assigned a color index in a range of ten values representing the color the neuron obtained during the training process. Then, the colors of the neurons in the averaged maps were obtained by averaging the color indices of the neurons in the single maps.

The MEP pattern of the most polar area in the averaged map of the highly active compounds is the most pronounced. In the three averaged maps, the destination of the polar spaces decreases according to decreasing activity of the compounds. The averaged map of the highly active compounds can be used to build a pharmacophore model. Therefore, a comparison of the maps of steroids with the averaged map allows one to establish whether a molecule belongs to the active or inactive CBG compounds.

Shape analysis

In the first part of the discussion, the 31 steroids were compared for their MEP similarities without consideration of the molecular shape. In the template approach, the shape will be considered by using a reference molecule within a series of molecules to prepare a template network which then forms a basis for the comparison of the surface of the other molecules. A reference network was trained with the van der Waals surface coordinates of corticosterone, **6**, the compound having the highest CBG activity. This compound supplies the reference Kohonen neural network (r-KNN), a template, while the analogues are filtered through this r-KNN to produce a series of individual compared KNNs (c-KNNs). In contrast to the previous approach, the template approach requires a superposition of all molecules onto the template molecule. The ring junction atoms 5, 8, 9, 10, 13 and 14 of the steroid system (see Fig. 9) were used to produce the superposition.

The resulting compared Kohonen maps can be marked with any molecular surface property. In our case, again the values of the MEP were used, specifically the MEP values on the surface of the steroid that is compared with corticosterone. The maps obtained in this procedure using the most active compound in this series, corticosterone, **6**, as the template molecule are shown in Fig. 10. The patterns with blank (white) areas represent a spatial mismatch, while non-blank areas indicate spatial similarities of the molecules. The maps of Fig. 10 show that the compounds with low CBG activity have rather large white areas of empty neurons. In fact, it has been shown that the number of empty neurons can be taken as a quantitative measure of the similarity of the surface of two molecules. Table 2 gives the number of empty neurons for the maps shown in Fig. 10; again the molecules are arranged in order of decreasing activity (cf. Fig. 6 and Table 1).

The results of the template approach can be better understood using a backprojection of the generated maps onto the molecular surface of the reference compound. This method clearly shows those parts of the surface where the two molecules that are compared are different. As an example, two maps of Fig. 10, the first, **6**, and the last one, **3**, of that series, corresponding to the compounds with the highest and the lowest CBG activity, are taken and projected onto the shape that is obtained by plotting the weights of the reference Kohonen network in 3D space. Figure 11 shows the backprojection of map 1 and map 31 of Fig. 10 (corresponding to molecules **6** and **3**)

TABLE 2
NUMBER OF EMPTY NEURONS FOR THE MAPS OF CBG COMPOUNDS

Compound	Number of empty neurons[a]	Compound	Number of empty neurons[a]
6	0	**12**	324
7	15	**31**	28
11	52	**4**	350
20	61	**5**	417
30	6	**16**	749
10	50	**17**	108
23	13	**18**	130
22	149	**14**	636
19	58	**15**	700
25	46	**13**	644
28	113	**9**	378
8	75	**2**	344
29	152	**3**	332
24	79		
21	296		
1	225		
27	69		
26	357		

Total number of neurons = 2500.
[a] The number of empty neurons is given relative to the reference compound, **6**.

TABLE 3
NUMBER OF EMPTY NEURONS IN THE COMPARATIVE KOHONEN MAPS SHOWN IN FIG. 13

Comparative map no.	Number of empty neurons
1	918
2	1112
3	1195
4	1663
5	826
6	1484
7	2011
8	739
9	1089
10	561

Total number of neurons = 2500.

onto the shape of **6**, the reference molecule. Figure 11 clearly indicates that part of the surface of the reference molecule corticosterone, **6**, that leads to the empty neurons in the Kohonen maps of 5-androstenediol, **3**, because no corresponding points are found on the surface of 5-androstenediol. Since in **3** (5-androstenediol), the long side chain of **6** (corticosterone) is replaced by the hydroxyl group, the surface of this area is empty. In the non-blank areas, the differences of the MEP are clearly visible.

In Fig. 10, up to map 21, **4**, the large red spotted area comes from position 3 of the steroid system (the carbonyl group), which in the other maps is predominantly yellow or totally absent. From map 22, **5**, onward all compounds have a hydroxyl group in this position. The second red spot of the maps comes from position 17 occurring up to map 14, **24**. These maps come from the steroids with the side chain $COCH_2R$ (R = H or OH) at position 17. From map 15, **21**, onward all maps except maps 16, 20, 25 and 26 have blank spaces in this area. The violet spotted area, which indicates the positive electrostatic potential below the D-ring and the side chain at position 17 (a $COCH_2R$ group with R = H or OH), is more or less apparent up to map 14, **24**. With a few exceptions (maps 20, 24, 25 of molecules **31**, **17**, **18**), this area is not visible in the low active compounds. The exceptions **31**, **17**, **18** also possess this side chain at position 17. Thus one can find this pattern again.

In general, a considerable amount of information concerning related compounds can be extracted from each map. The analysis performed here shows that the compounds with high, intermediate, and low activity for CBG binding can be distinguished in different clusters due to effects determined by shape and the MEP. It can be concluded that for CBG affinity both the negative potential areas at position 17 and at the carbonyl at position 3 with a conjugated double bond at C4-C5 in the A-ring are important, as is the positive potential area below the side chain at position 17 and the D-ring.

The largest deviation was found for compound **31**, a compound of intermediate activity. The corresponding map (see Fig. 10, S31) shows a good spatial match and an MEP pattern similar to the template map. From this Kohonen map, we cannot find any shape and/or MEP difference to the most active compound **6** (the first map). In order to find any differences between the shapes of **31** and **6**, we turn the investigation around and use molecule **31** as the template molecule to produce a reference net. Compound **6** can then be compared with the template molecule using KNN. Figure 12a shows the obtained Kohonen maps. Now the map of **6** contains patterns with additional blank areas. The projection of these two maps onto the shape that is obtained by plotting the weights of **31** (see Fig. 12b) reveals that the blank space in the map of **6** corresponds to the region close to position 3 of the A-ring. Furthermore, the regions below the ring junction in these maps show different patterns of MEP values. Compared to **6**, **31** has an additional methyl group at position 2a and a fluorine atom at position 9a (see Fig. 5). These results suggest, as a reason for the low affinity of **31**, a sterically unfavorable effect near the A-ring and/or an unfavorable electrostatic interaction below the ring junction.

These results show that the selection of an appropriate reference molecule is of importance in the template approach to properly analyze existing differences. No significant differences in the features of the maps **6** and **31** could be realized when **6** was used as the template molecule. However, producing a template network with **31** as the reference revealed differences in shape and MEP between the two molecules. This shows that, in certain cases, finer details in the shape differences can be discovered by interchanging the reference and the compared molecule.

Optimization of overlap

The template approach can also be used for finding the best alignment of two molecules. This is demonstrated with two steroids from the data set, 5-androstenediol, **3**, and corticosterone, **6**. A series of 10 different spatial arrangements of the two molecules was manually generated by random transformations of each of the three coordinate axes. These different alignments are shown in Fig. 13.

One of the two molecules, the highly active corticosterone, **6**, is taken as the reference structure and points from the surface of this molecule are used for training a Kohonen network of size 50×50 neurons. Points of the surface of 5-androstenediol, **3**, are sent through this reference neural network for each of the 10 different positions of **3** against **6** (see Fig. 13). These 10 experiments result in 10 different compared maps, as shown in Fig. 14.

The goodness of fit is evaluated by the number of empty neurons, N_e, in the compared maps, which are given in Table 3.

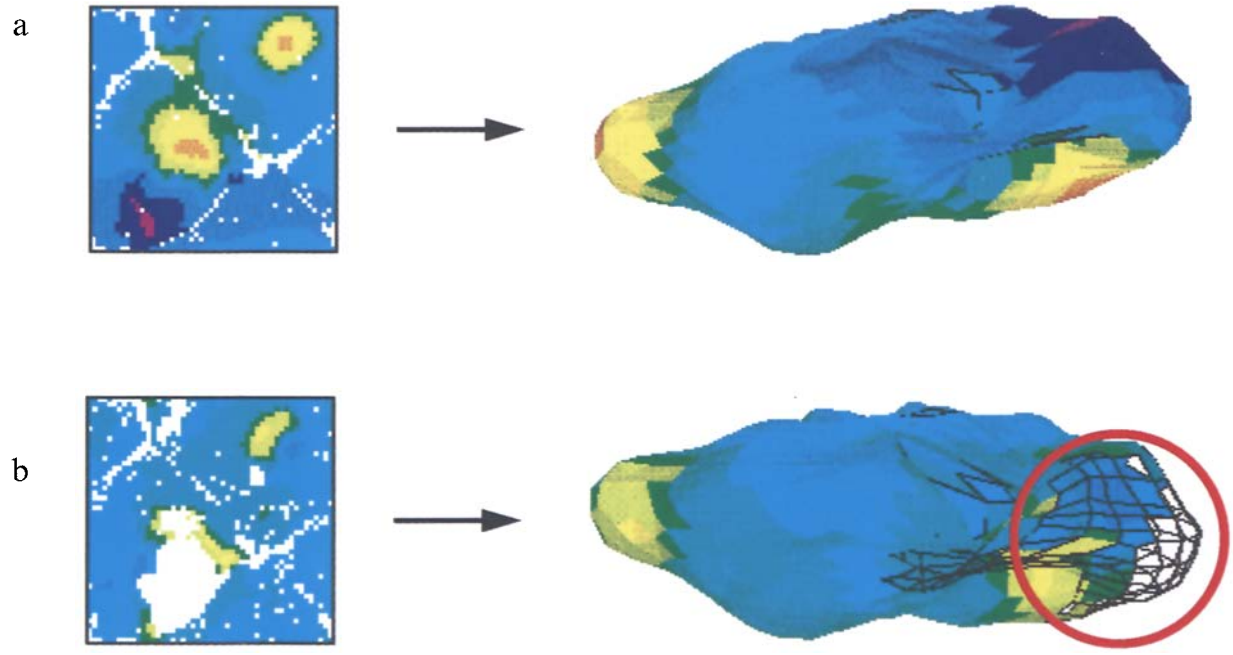

Fig. 11. Backprojection of the maps of (a) corticosterone, **6**, and (b) 5-androstenediol, **3**, onto the shape obtained by plotting the weights of the reference network trained with the corticosterone data.

The best superposition can be deduced from the lowest number of empty neurons indicating the largest correspondence in the geometry of the two surfaces, that of the reference structure, **6**, and the superimposed structure **3**. Orientation 10 shown in Fig. 13 has the lowest number of empty neurons (cf. Fig. 14 and Table 3) and therefore can be taken as the best alignment from the 10 tested alignments of the two molecules shown in Fig. 13.

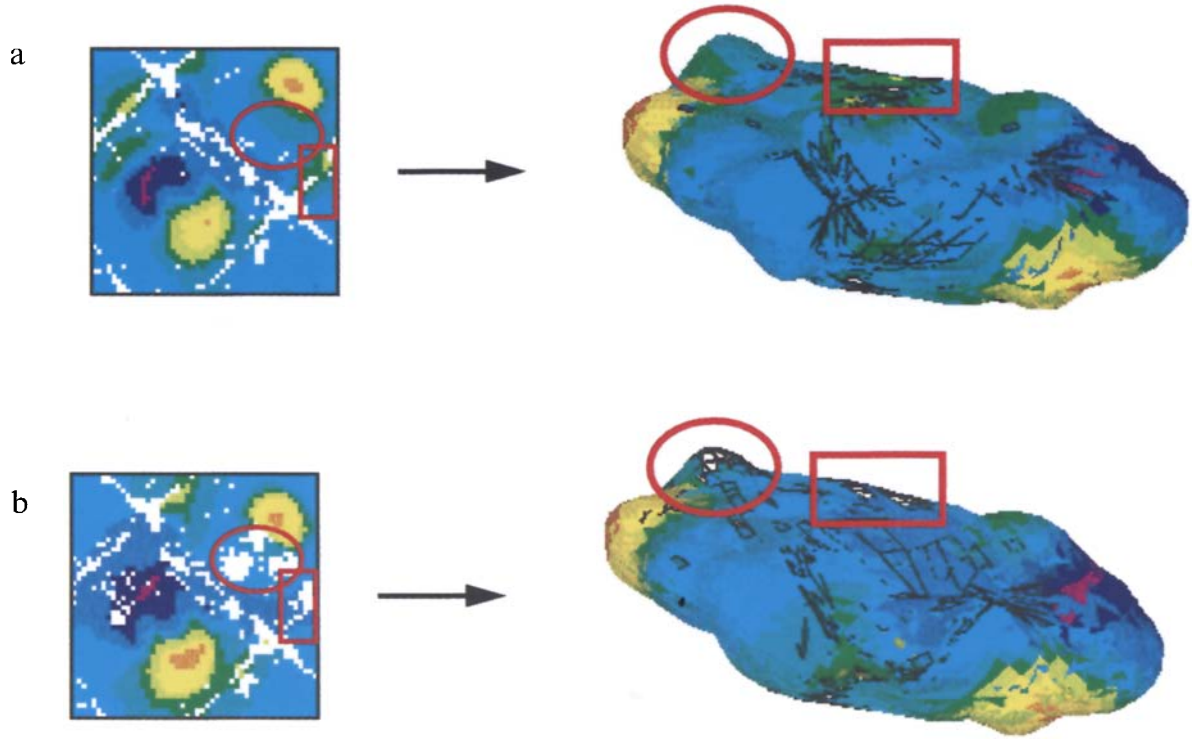

Fig. 12. Backprojection of the Kohonen maps of (a) **6** and (b) **31**. Molecule **31** was used to obtain the reference KNN.

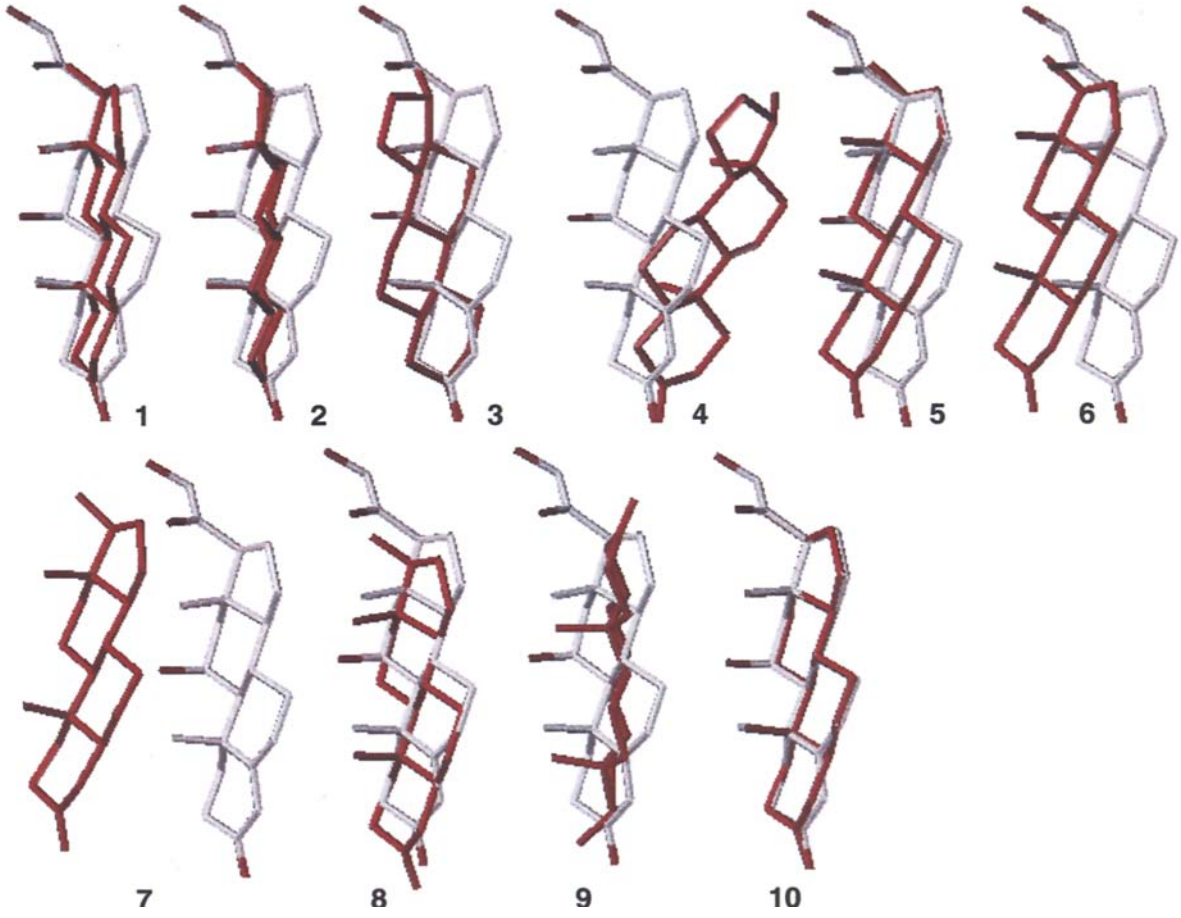

Fig. 13. Ten random alignments of 5-androstenediol, **3**, and corticosterone, **6**.

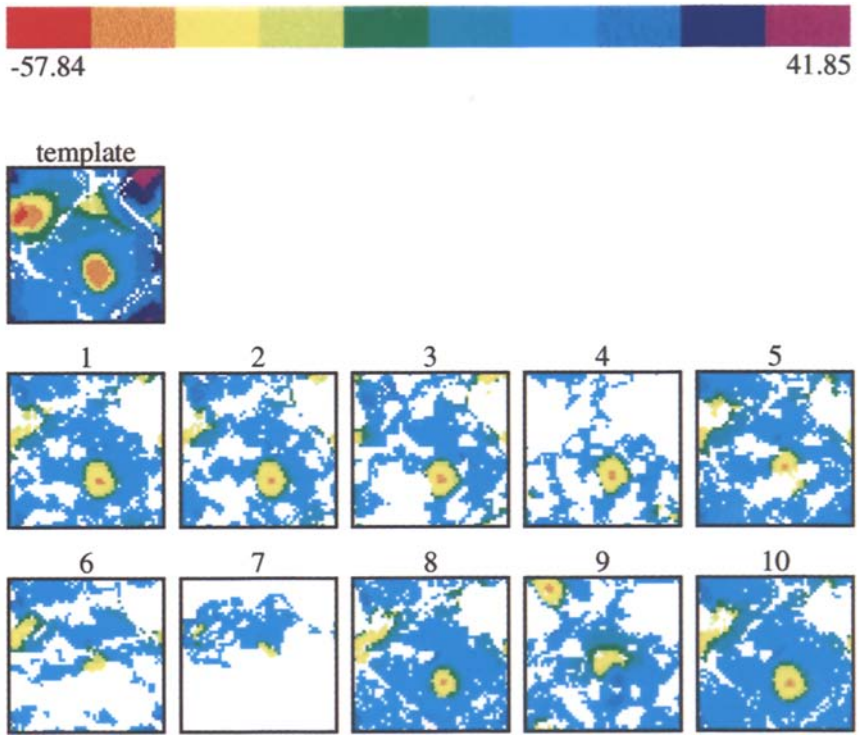

Fig. 14. The reference KNN of corticosterone, **6**, and 10 compared KNNs of random orientations of 5-androstenediol, **3**, obtained by processing through the reference KNN.

Conclusions

In the discussion of structure–activity relationships in the steroid series, the 2D Kohonen maps serve as a versatile tool to elucidate the structural factors responsible for biological activity. In cases like the CBG example, the comparison of trained maps of each molecule directly allows a classification whether a molecule is active or not. Therefore, the approach supplies a straightforward tool to qualitatively predict the activity of an unknown molecule. Due to its simplicity, it might even be used for screening large sets of molecules in order to find potential de novo compounds. In this sense, it supplements methods aiming at quantitatively modeling biological activity such as the CoMFA method [1] or the coding of the molecular electrostatic potential by autocorrelation vectors [22].

The technique of averaging the maps of active compounds enables one to elucidate the essential features of active compounds. Thus, these maps can be used to define a 2D representation of a pharmacophore model. The projection of the 2D Kohonen maps onto the shape that is obtained by plotting the weights of the template network in 3D space allows one to address areas of the molecular surface where active and inactive compounds differ significantly. Such information can be used for designing new substituents at the positions found in the molecular scaffold. Thus, the backprojection technique is the link between the nonintuitive form of information encoded in the Kohonen maps and the 3D world of molecular structures. The Kohonen map approach is not limited to the electrostatic potential and the shape factor as presented here. Hydrophobic, hydrogen bonding, dipole and multipole features, etc. can be mapped on the surface of molecules and then be used to study the interplay of these different factors for structure–activity relationships.

Acknowledgements

Financial support of this work by the German Minister of Research and Technology (BMBF) through the scheme 'Molekulare Bioinformatik' (ÄBAV project, No. 01IB305D) is gratefully acknowledged. We also wish to thank Dr. H. Bruns (Merck KGaA) for comments on the manuscript.

References

1 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
2 Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993.
3 Zupan, J. and Gasteiger, J., Anal. Chim. Acta, 248 (1991) 1.
4 a. Gasteiger, J. and Zupan, J., Angew. Chem., 105 (1993) 510.
b. Gasteiger, J. and Zupan, J., Angew. Chem. Int. Ed. Engl., 32 (1993) 502.
5 Zupan, J. and Gasteiger, J. (Eds.) Neural Networks for Chemists: An Introduction, VCH, Weinheim, Germany, 1993.
6 Kohonen, T., Biol. Cybernet., 43 (1982) 59.
7 Kohonen, T. (Ed.) Self-Organisation and Associative Memory, 3rd ed., Springer, Berlin, Germany, 1989.
8 Kohonen, T., Proc. IEEE, 78 (1990) 1460.
9 Gasteiger, J., Li, X., Rudolph, C., Sadowski, J. and Zupan, J., J. Am. Chem. Soc., 116 (1994) 4608.
10 Gasteiger, J., Li, X. and Uschold, A., J. Mol. Graph., 12 (1994) 90.
11 a. Gasteiger, J. and Li, X., Angew. Chem., 106 (1994) 671.
b. Gasteiger, J. and Li, X., Angew. Chem. Int. Ed. Engl., 33 (1994) 643.
12 Li, X., Gasteiger, J. and Zupan, J., Biol. Cybern., 70 (1993) 189.
13 Sadowski, J., Rudolph, C. and Gasteiger, J., Anal. Chim. Acta, 265 (1992) 233.
14 Sadowski, J. and Gasteiger, J., Chem. Rev., 93 (1993) 2567.
15 Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
16 Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
17 a. Gasteiger, J. and Saller, H., Angew. Chem., 97 (1985) 699.
b. Gasteiger, J. and Saller, H., Angew. Chem. Int. Ed. Engl., 24 (1985) 687.
18 Dunn, J.F., Nisula, B.C. and Rodbard, D., J. Clin. Endocrinol. Metab., 53 (1981) 58.
19 Mickelson, K.E., Forsthoefel, J. and Westphal, U., Biochemistry, 20 (1981) 6211.
20 Westphal, U. (Ed.) Steroid–Protein Interaction II, Springer, Berlin, Germany, 1986.
21 Good, A.C., So, S.S. and Richards, W.G., J. Med. Chem., 36 (1993) 433.
22 Wagener, M., Sadowski, J. and Gasteiger, J., J. Am. Chem. Soc., 117 (1995) 7769.
23 Jain, A.N., Koile, K. and Chapman, D., J. Med. Chem., 37 (1994) 2315.