The beauty of molecular surfaces as revealed by self-organizing neural networks

Johann Gasteiger, Xinzhi Li and Anders Uschold

Organisch-chemisches Institut, Technische Universität München, Germany

The Kohonen neural network is a self-organizing network that can be used for the projection of the surface properties of molecules. This allows one to view properties on a molecular surface, like the electrostatic potential in a single picture. These maps are useful for the comparison of molecules and provide a new definition of molecular similarity.

Keywords: molecular modeling, electrostatic potential, self-organizing map, nonlinear projection

INTRODUCTION

Chemists have developed a highly efficient international language for representing and communicating information on the structure of chemical compounds: the structural formula. The structural formula accounts for the atoms contained in a molecule and for the way they are connected by bonds. From a structural formula, a chemist can deduce many physical or chemical properties of a compound like solubility, dipole moment or its chemical reactivity. In recognition of the importance of a structural formula for expressing chemical information, the standard representation of chemical compounds for computer manipulation has been chosen as a connection table that is basically a one-to-one translation of a structural formula. However, it is clear that a more sophisticated representation of the structure of a molecule has to give account of the three-dimensional (3D) arrangement of the atoms in a molecule. Particularly, the biological activity of a compound is critically dependent on its 3D structure. Thus, modeling the 3D structure of molecules has become a cornerstone in research and development of pharmaceutical and agrochemical compounds.

The generation and manipulation of the 3D coordinates of the atom in a molecule is therefore of much interest.[1] However, it must be realized that molecules interact with each other, in a large part, at their surfaces. Studying molecular surface properties is therefore essential for understanding many physical, chemical or biological properties and, in particular, for understanding the binding of a substrate to its receptor.

Address reprints to Professor Gasteiger at Organisch-chemisches Institut, Technische Universität München, D–85747 Garching, Germany.
Received 2 August 1993; accepted 29 October 1993

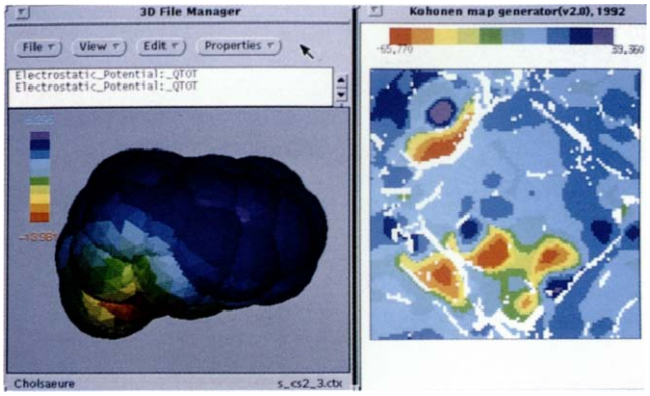
PROJECTION OF MOLECULAR SURFACE PROPERTIES

One of the most important properties governing the interaction between molecules is the electrostatic potential exerted by molecules. Therefore, it is not surprising that the electrostatic potential on the surface of molecules plays a central role in molecular modeling studies. However important the molecular electrostatic potential (MEP) on a surface and in spite of all the powerful computer graphics tools developed, the visualization of the MEP still presents problems.

The MEP on a surface is a 3D object and has to be viewed from various sides to obtain an impression of its full extent. The approach usually taken is to select an observation point and then show the part of the MEP that can be seen from this point on the graphics screen. This corresponds to a linear projection of the MEP on the screen. Various observation points have to be chosen to obtain an impression of the entire MEP. Color Plate 1 (left-hand side) shows such a linear projection of the MEP of cholic acid (Figure 1). Although this view shows that there is a side with a positive MEP (blue) and a side with a negative MEP (yellow and red), in cholic acid it is not sufficient to show fine details such as the individual sites with a negative potential that can function here as sites of attraction for a proton and are, therefore, sites for hydrogen bonding. With three OH- and one COOH group, there must be four such sites.

We will present here a method that allows one to visualize the entire MEP on a surface in a single picture. In the case of cholic acid, this leads to the picture shown in Color Plate 1 (right-hand side). Here, the four sites with negative electrostatic potential prone to hydrogen bonding can clearly be distinguished. (A detailed analysis of this map will be given in the discussion of Color Plate 12.)

The problem of representing a 3D object on a two-dimensional (2D) map is nothing new. In fact, our own brain has solved this problem. In the somatosensory cortex of the brain, the entire human body is represented in such a way that the sensory signals resulting from the excitation of a certain part of the body will always fire a neuron in a clearly defined position. Closely adjacent parts of the body will fire



Color Plate 1. Electrostatic potential on the van der Waals surface of cholic acid. At left, a linear projection; at right, a Kohonen map.

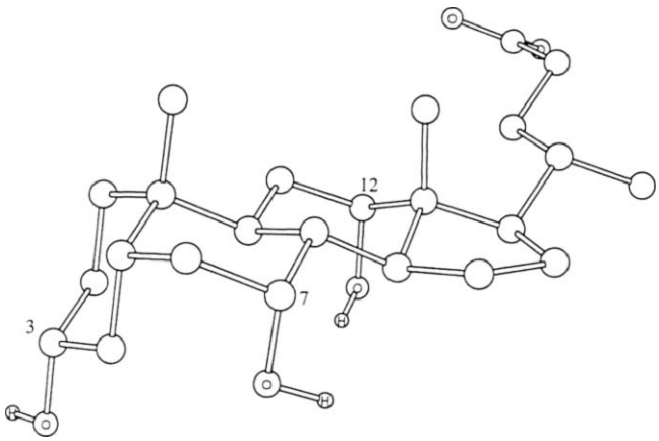


Figure 1. Ball-and-stick model of cholic acid.

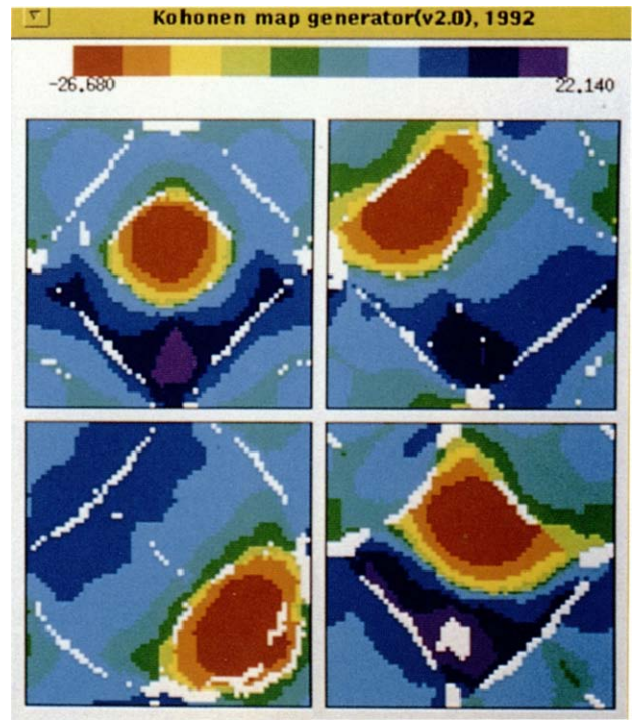
closely adjacent neurons in the somatosensory cortex (Figure 2).

The mapping of the human body onto the somatosensory cortex is such that those parts of the body having many sensory receptors (hands, lips, etc.) have reserved correspondingly large areas in the somatosensory cortex.

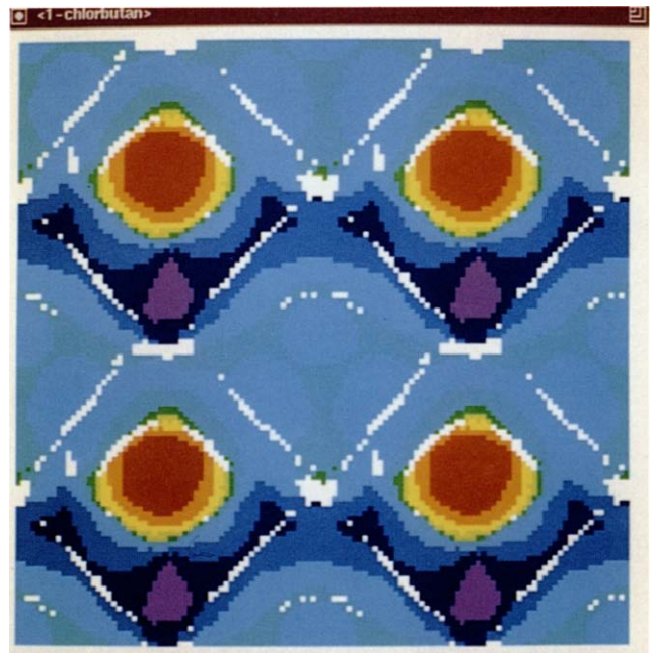
Kohonen has developed an artificial neural network that tries to model this feature of the human brain.[2,3] This method is also called a self-organizing topological feature map because it generates in an unsupervised process a 2D map of an information space while conserving as much as possible of the topology of the information.

The Kohonen neural network has been presented elsewhere[2,3] and its potential for the investigation of chemical information has already been stressed.[4,5] Details of the method for mapping MEPs on surfaces by a Kohonen network have also been given.[6]

It may suffice here to say that the inputs to the 2D Kohonen network are the Cartesian coordinates of a point on the molecular surface (Figure 3). As Kohonen learning is an unsupervised process, the information to be investigated (in our case, the property on the surface of a molecule) is not used in the learning process. Learning in a Kohonen network



Color Plate 2. Kohonen maps of the MEP of 1-chlorobutane (upper left), 2-chlorobutane (upper right), 2-chloro-2-methylpropane (lower left), and 1-chloro-2-methylpropane (lower right).



Color Plate 3. Kohonen map of the MEP of 1-chlorobutane (fourfold).

is a competitive process during which only one neuron finally becomes exited by each data point being input to the network. In our case, we have chosen the winning neuron to be the one having three weights that are closest in value to the input Cartesian coordinates.

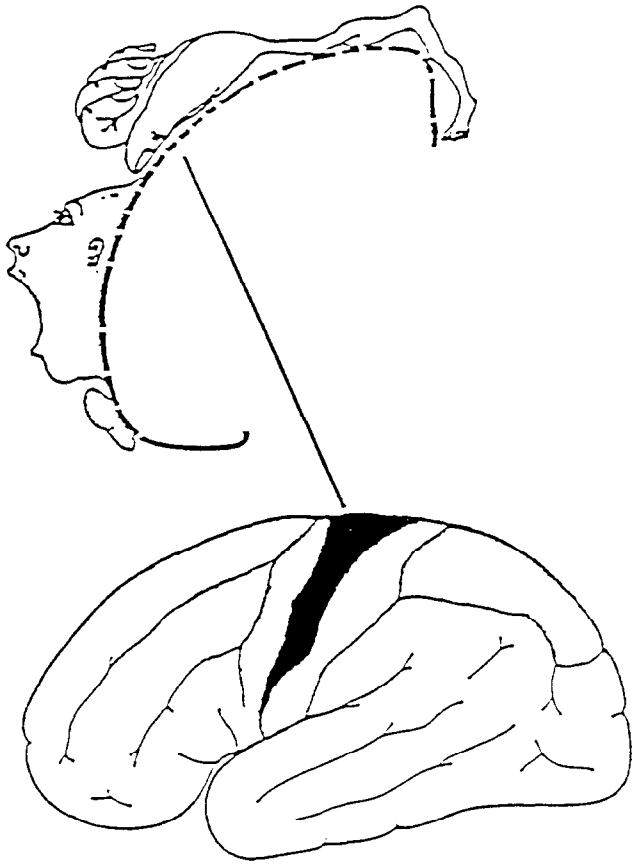


Figure 2. The mapping of the human body onto the somatosensory cortex of the brain.

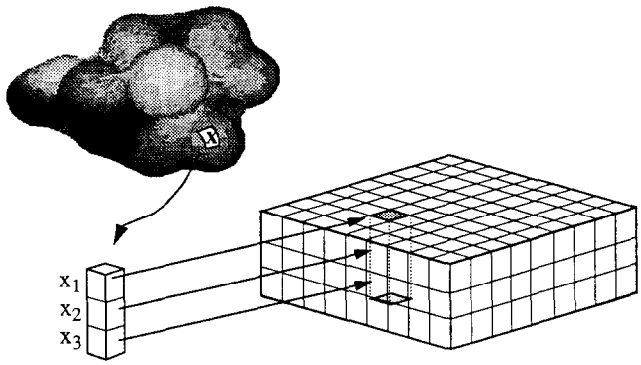


Figure 3. Mapping of a molecular surface onto a Kohonen network.

In order to have a plane without beginning and without end, the top of the 2D array of neurons is connected to the bottom and the right-hand side of the plane is connected to the left-hand side. This gives the surface of a torus. Mapping of molecular surfaces is thus performed 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.

It is important to note that these cuts can be made at arbitrary lines and, therefore, the Kohonen maps of molecular surfaces can be shifted into any direction—the part of the map that is shifted to the left is again appended at the right-hand side and the same is true for vertical movements.

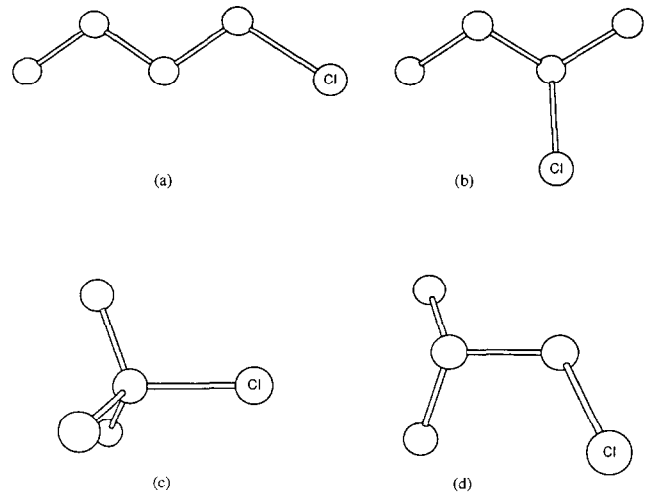


Figure 4. Ball-and-stick models of 1-chlorobutane (a), 2-chlorobutane (b), 2-chloro-2-methylpropane (c), and 1-chloro-2-methylpropane (d) used for generating the maps in Color Plate 2.

In summary, a Kohonen network can be used as a nonlinear projection method to spread molecular surfaces onto a plane. Another approach for nonlinear mapping of MEPs has been made by Dean and Chan.[7] In their gnomonic projection method, they emanate vectors from the center of a molecule onto a tesselated icosahedron around the molecule. The value of the electrostatic potential at the point where the vector penetrates the molecular surface is then shown on the icosahedron. Problems may arise in strongly folded molecules where the vector might penetrate the molecular surface several times.

In our approach, the following procedure is pursued in all studies:

(1) A molecule is input graphically as a structural formula giving information on the stereochemistry, where necessary, by wedges and dotted lines by the molecule editor medtool.[8]

(2) Three-dimensional atomic coordinates are calculated for this molecule by the 3D structure generator CORINA.[9,10]

(3) Partial atomic charges are calculated by the empirical PEOE method[11] and its extension to π systems.[12]

(4) The electrostatic potential is calculated for the van der Waals surface by Coulomb's law from a point charge and the partial atomic charges.

(5) A Kohonen network is trained by randomly sampling points on the van der Waals surface using their Cartesian coordinates as input to the network.

(6) When learning in the Kohonen network has stabilized, it is found that the individual neurons have obtained points from closely adjacent points on the molecular surface. The neurons are then colored according to the surface property observable at these points. In most cases, the MEP is used to color the neuron map. In cases where the manner of mapping has to be identi-

fied, the points are distinguished by an atomic surface assignment (ASA). In this procedure, the original contribution of an atom to the entire molecular surface is monitored and a point can then be assigned to a specific atomic surface. This assignment of a point to the surface of a specific atom is used for coloring.

It should be emphasized that Kohonen mapping can be applied to other molecular surfaces such as the Connolly[14] or the solvent-accessible surface.[15] Other ways of calculating electrostatic potentials by calculating quantum mechanical methods of various degrees of sophistication can be used. In fact, any molecular surface property can be used to identify the points mapped onto the individual neurons.

The calculation of one typical Kohonen map of the size 100 × 100 takes 10–15 minutes on a SPARCstation-10 computer. It should be pointed out that the simulation of a Kohonen network is quite suitable for parallel computation. Without much effort, our simulator has been ported to a parallel computer with 16 processors; an acceleration by a factor of ten has been achieved.

EFFECTS OF ELECTRONEGATIVITY AND TOPOLOGY

As a simple entry, we present in Color Plate 2 the results of the Kohonen maps of the MEP of the four isomeric monochlorobutanes: 1-chlorobutane, 2-chlorobutane, 2-chloro-2-methylpropane, and 1-chloro-2-methylpropane. Figure 4 shows the 3D models used for producing these Kohonen maps.

The large negative MEP at the chlorine atom is clearly seen as a red spot in each of the four maps. However, each map shows quite distinctive features, which indicate that the spatial arrangement of atoms has a profound effect on the appearance of the maps and that isomeric compounds can be clearly distinguished.

We have already noted in the preceding section that the mapping is performed on the surface of a torus and that the maps can therefore be moved in any direction. We have found it quite helpful to express this feature of the maps to be without beginning and without end by putting several identical maps together, like tiles. Color Plates 3 and 4 show the tiling of four maps of the MEP of 1-chlorobutane and 2-chloro-2-methylpropane, respectively.

It may now be noted that the Kohonen maps show distinguished patterns of white lines and spots. These white spots correspond to empty neurons—neurons to which no point of the molecular surface is mapped. We have shown in a mathematical analysis that these empty neurons are the result of topological distortions.[12] The topology of a molecular surface is more like a sphere and thus quite distinct topologically from a torus. This different topology can only be accommodated by a distortion. Figure 5 attempts to visualize this topological distortion. The torus is squeezed together at two positions and the sign of the surface (the inside and the outside of the surface) changes.

Projecting the Kohonen maps back onto the molecular surface shows where the topological distortions are located. We have found that the sites where the topological distortions occur provide information on the shape of a molecule and these distortions try to avoid cutting through atoms.[6]

Color Plate 3 shows that the white lines of empty neurons generate two regular rhomboids. This indicates that the distortion splits the molecule into two parts of equal size—one containing the polar part of the molecule resulting from the chlorine atom and the other part containing the other end of the molecule that has an evenly distributed electrostatic potential (an area predominantly shown in light blue).

Color Plate 4 shows that in 2-chloro-2-methylpropane, the distortion cuts through the molecule in a manner that results in two halves of unequal size by avoiding a cut through the carbon atom in the center of the molecule.

In order to investigate the exact location of topological distortion in a molecule, we have projected the Kohonen net back onto the molecular surface. This is shown in Color Plates 5, 6 and 7 for 1,2-ethanediol; 1,3-propanediol and phenole in the conformations given in Figure 6.

Color Plate 5 shows, on the right-hand side, the Kohonen map of the MEP of 1,2-ethanediol (in quadruplicate form) and, on the left-hand side, the projection of this map onto the space-filling model of this molecule. The Kohonen map consists of two identical parts. This is an indication that the topological distortion coincides with the plane of symmetry of this molecule cutting perpendicularly through the C-C bond. This is shown in the projection of the Kohonen map onto the molecular surface. An inversion occurs at the distortion that is indicated by leaving one part of the molecule uncolored.

The Kohonen map of 1,3-propanediol (Color Plate 6, right-hand side) shows rhomboids of unequal size. This is an indication that the distortion avoids the plane of symmetry that cuts through the central carbon atom, but passes next to this atom through the molecule. This is clearly shown in the projection of the Kohonen map onto the molecular surface.

In phenole (Color Plate 7), the distortion isolates the OH group, two carbon atoms and one hydrogen atom from the rest of the molecule. Incidentally, the Kohonen map shows quite distinctly in one rhomboid the oxygen atom and the

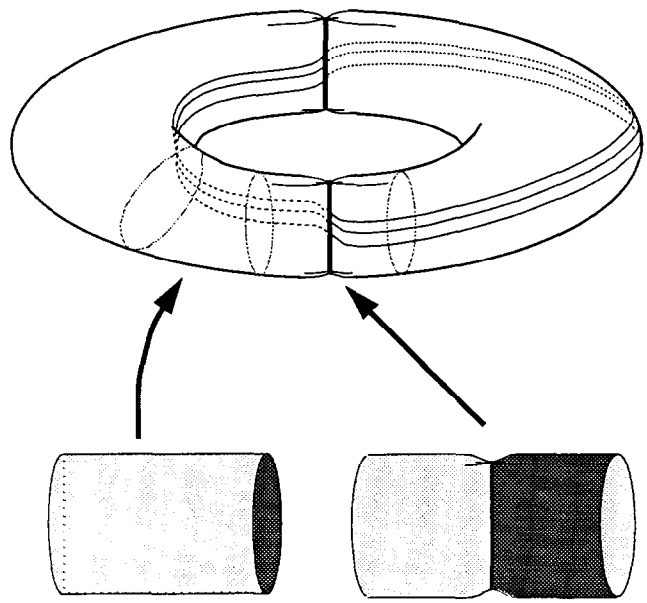
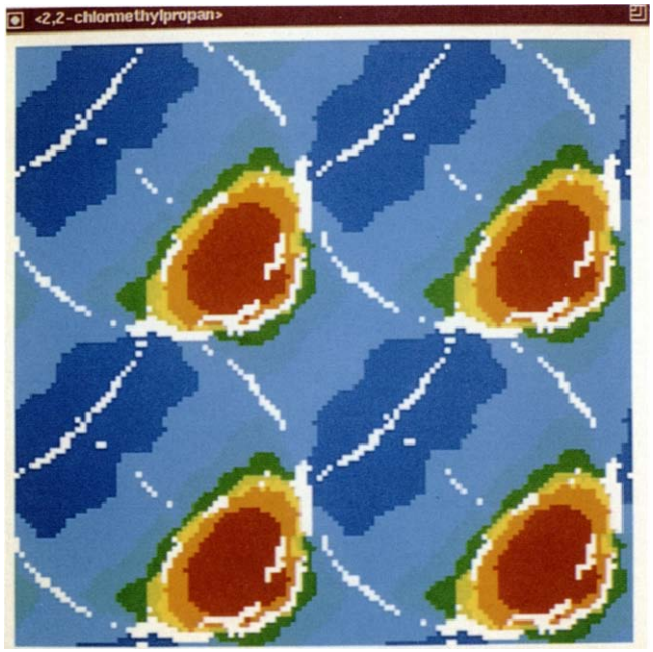
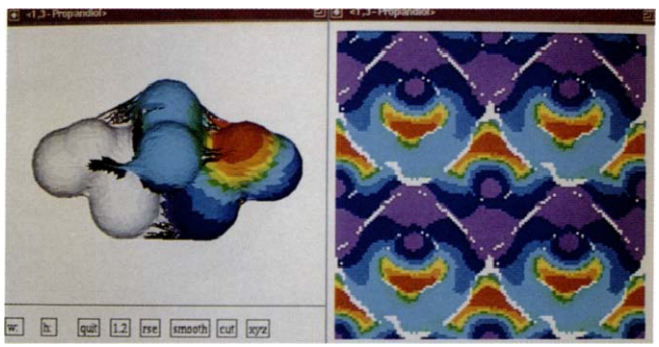


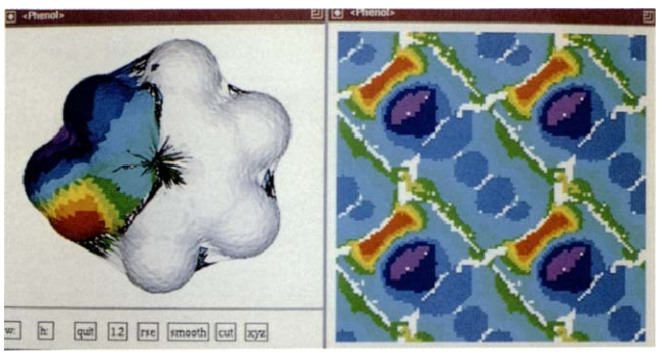
Figure 5. A typical kind of topological distortion of the torus in Kohonen mapping.



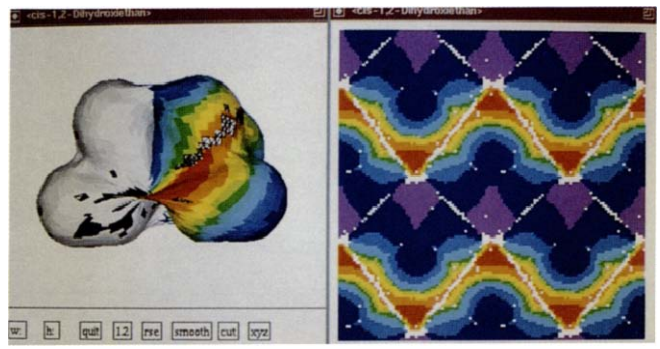
Color Plate 4. Kohonen map of the MEP of 2-chloro-2-methylpropane (fourfold).



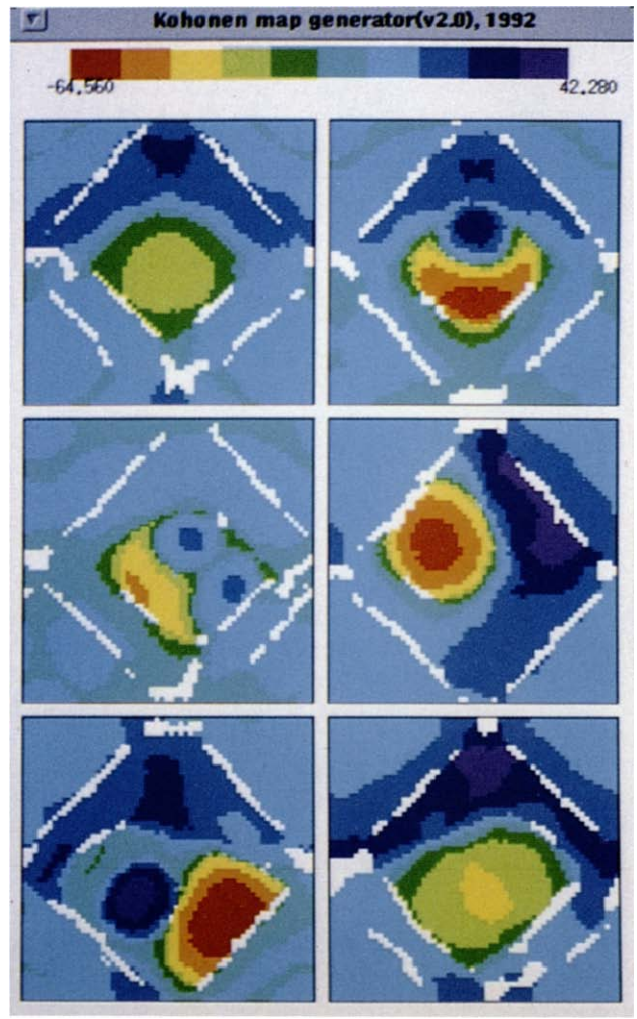
Color Plate 5. Kohonen map of the MEP of 1,2-ethanediol. On the left, as projected back onto the van der Waals surface; on the right, fourfold.



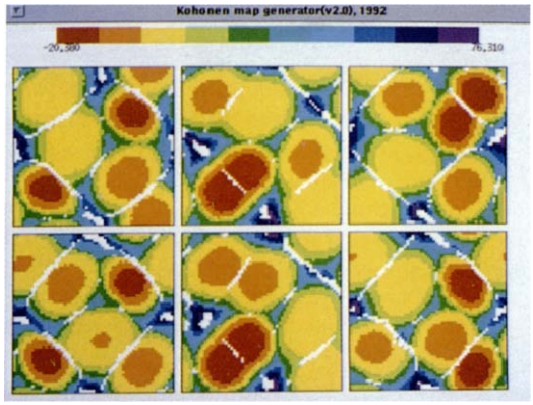
Color Plate 6. Kohonen map of the MEP of 1,3-propanediol. On the left, as projected back onto the van der Waals surface; on the right, fourfold.



Color Plate 7. Kohonen map of the MEP of phenole. On the left, as projected back onto the van der Waals surface; on the right, fourfold.



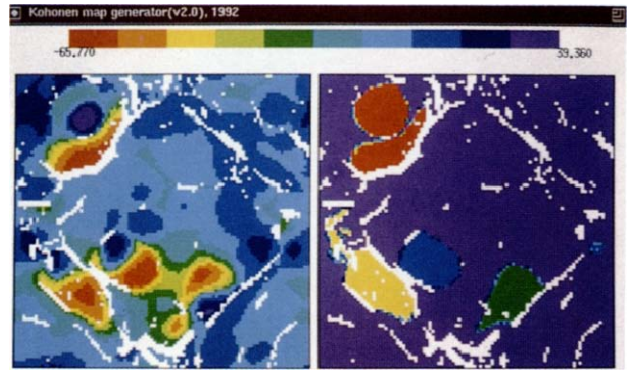
Color Plate 8. Kohonen maps of the MEP of 1-chlorobutane (upper left), butanol-1 (upper right), 1-aminobutane (middle left), butanal (middle right), butanoic acid (lower left) and 1-nitrobutane (lower right).



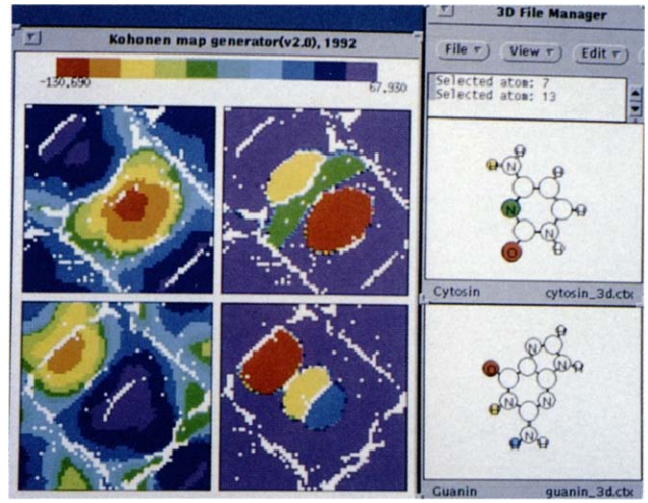
Color Plate 9. Kohonen maps of the MEP of 1,2-difluoro-; 1,2-dichloro- and 1,2-dibromoethane. Top row: S,S configuration; from the left: anti, eclipsed and gauche conformation. Bottom row: S,R configuration; from the left: anti, eclipsed and gauche conformation. (See also Figure 8.)

Color Plate 10. Sixfold Kohonen map of the MEP of the anticonfiguration of S,R 1,2-difluoro; 1,2-dichloro- and 1,2-dibromoethane.

Color Plate 11. Sixfold Kohonen map of the MEP of the eclipsed configuration of S,R 1,2-difluoro; 1,2-dichloro- and 1,2-dibromoethane.



Color Plate 12. Kohonen map of MEP of cholic acid (at left) and of its ASA (at right). The 3-, 7- and 12-OH groups are colored in blue, green and yellow, respectively.



Color Plate 13. Cytosine: MEP map (upper left), ASA map (upper middle) and structural formula (upper right). Guanine: MEP map (lower left), ASA map (lower middle) and structural formula (lower right).

two hydrogen atoms (as a merged spot of dark blue color) and in the other rhomboid (as spots of blue color) the other four hydrogen atoms.

Functional groups show quite distinct patterns in the Kohonen maps of the MEP. Color Plate 8 illustrates this on a series of butane derivates. The 3D models of the six molecules investigated are given in Figure 7.

The Kohonen maps of the MEP of the six molecules have been arranged in Color Plate 8 so that the electronegative atoms (site of negative MEP) are in the center of the maps.

Starting at the left of the top row, the map of 1-chlorobutane shows the site of the negative potential of the chlorine atom (yellow-green) and the positive potential of the adjacent carbon and two hydrogen atoms (blue and dark blue), resulting from the polarization by the chlorine atom. The next map (top row right) of 1-butanol indicates the oxygen atom (half moon in red) and the hydrogen atom of the OH group (in dark blue) as well as the polarization on the adjacent CH_2 group. In 1-aminobutane (middle left), the nitrogen atom (yellow and orange) and its two hydrogen

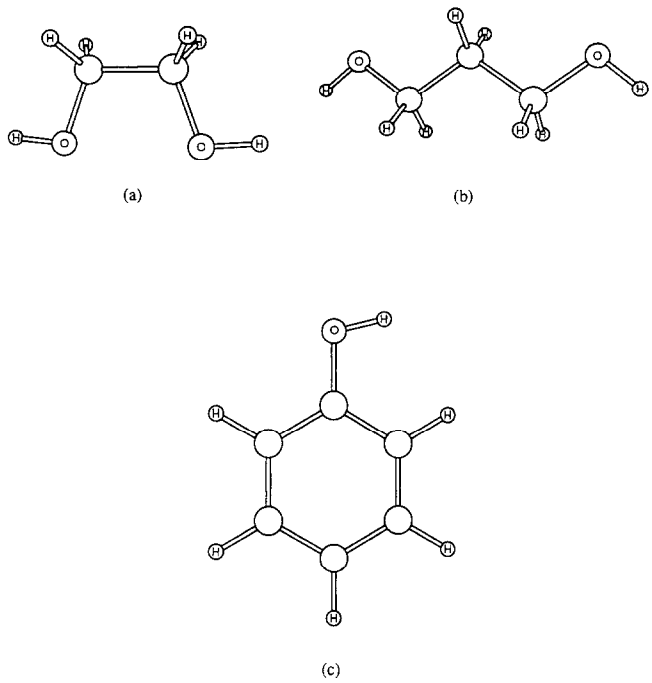


Figure 6. Ball-and-stick models of 1,2-ethanediol (a); 1,3-propanediol (b) and phenole (c) used for generating the maps in Color Plates 5, 6 and 7.

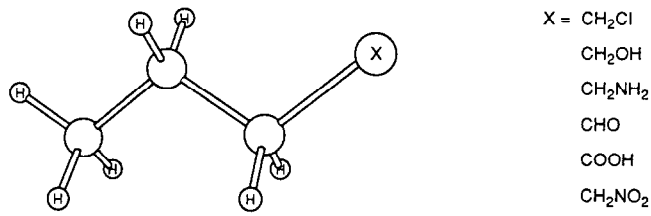


Figure 7. Three-dimensional models of the butane derivates that have their Kohonen maps presented in Color Plate 8.

atoms (two blue spots) can be clearly distinguished. The aldehyde (round red spot for the oxygen atom), the carboxylic acid (red spot for the carbonyl oxygen atom), blue-green half moon for the oxygen, purple spot for the hydrogen of the OH group and the nitro compound (two yellow-green spots for the two oxygen atoms) all show quite distinct patterns that can be used in identifying these groups in the Kohonen maps of the MEP of larger polyfunctional molecules.

To illustrate this, let us have a look at Color Plate 1 with the Kohonen map of the MEP of cholic acid. In the top left corner of Color Plate 1 (right-hand side), the features of the carboxylic acid group as shown in Color Plate 8 (lower left) can clearly be identified. The other spots with a negative electrostatic potential in the Kohonen map all show the features, although sometimes slightly distorted, of the OH group as represented in Color Plate 8 (top right) and thus allows us to locate the three OH groups.

EFFECTS OF CONFORMATION, CONFIGURATION AND SYMMETRY

The next study is concerned with the way the 3D arrangement of atoms is expressed in the Kohonen maps. To this effect, we have taken ethane and substituted each methyl group with three atoms that can be clearly distinguished—a fluorine, a chlorine and a bromine atom. This gives rise to a chiral center at each carbon atom. Both the S,S and the S,R forms have been investigated. For each of three two diastereomeres, three conformations (the anti, the eclipsed and a gauche conformation) have been studied. Altogether this gives six structures as shown in Figure 8.

Color Plate 9 shows the Kohonen maps of these structures in the same order as the structures in Figure 8. We refrain from a detailed discussion as a closer inspection quite clearly reveals how the 3D arrangement of the atoms is mapped onto the Kohonen maps. (The bromine atoms are easily identified as yellow spots, the chlorine atoms as light brown, and the fluorine atoms as red spots.) Even a cursory investigation already shows that all six structures lead to quite distinctive maps, making the point that the 3D structure is reflected in the Kohonen maps. This point is stressed even more when the tiling procedure is used to show fully the Kohonen maps. Color Plate 10 shows this with the sixfold map of the anticonfiguration of the S,R form and Color Plate 11 shows this for the eclipsed conformation of the S,R form.

ATOMIC SURFACE ASSIGNMENT

Color Plate 1 presents the Kohonen map of the MEP of cholic acid. Four sites of negative electrostatic potential associated with the four functional groups can be clearly distinguished. A closer inspection and a comparison with Color Plate 8 allows one to identify the carboxylic acid group with the spot in the top left-hand corner of Color Plate

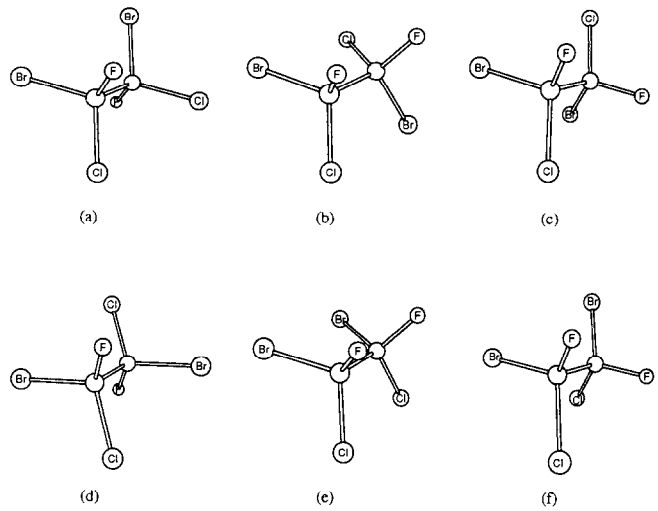


Figure 8. Ball-and-stick models of 1,2-difluoro-; 1,2-dichloro- and 1,2-dibromoethane: anti (a), eclipsed (b) and gauche (c) conformation of the S,S form and the same for the S,R form (d, e and f).

1b. However, the question of which of the other three red spots are assigned to which hydroxy group of cholic acid is left open. An answer can be given by coloring a neuron not according to the MEP observed at the points that are mapped onto the neuron, but by identifying individual atoms. This is achieved, as we briefly mentioned in a previous section, by dissecting the molecular surface into parts that are contributed by the individual atoms. Each point can then be assigned to a specific atomic surface. This ASA is then used for identifying the points mapped onto a neuron and thus coloring this neuron. Color Plate 12 compares the Kohonen map of the MEP with that of the ASA. The coloring of the ASA map clearly exhibits the mapping of the individual oxygen atoms. It should be emphasized that both the MEP and the ASA maps were obtained from the *same* Kohonen network. The only difference lies in the method used for coloring the individual neurons of the network. This underscores the merits of an unsupervised learning method.

DNA BASE PAIRING

Important as MEPs are in investigating and understanding the interactions of molecules, particularly in biological systems, Kohonen maps of MEPs have many applications. We are actively exploring this potential.[6,16] Here, only one rather transparent investigation is given.

Color Plate 13 shows the MEP map, the ASA map, and the ball-and-stick models of cytosine (top row) and guanine (bottom row). The ASA maps allow one to identify easily how the atoms are mapped onto the Kohonen network and thus understand the details of the MEP maps. The MEP maps of cytosine and guanine match each other very nicely. Sites of negative potential in cytosine have their counterparts in guanine and vice versa. The three atoms in cytosine and the three in guanine responsible for forming the three hydrogen bridges in the pairing of these two bases in DNA are clearly distinguished.

CONCLUSIONS

Kohonen's self-organizing map is a powerful method for mapping molecular surfaces. Because of the unsupervised character of this learning method, the resulting 2D neural networks can be used for visualizing any property on a molecular surface. Maps of the MEP show the important features of this property in a single picture. These MEP maps offer many possibilities for studying the interactions between molecules, particularly in biological systems. They provide new measures of molecular similarity and molecular correspondence.

ACKNOWLEDGMENTS

We appreciate the many helpful discussions with Prof. Jure Zupan during our joint venture into the exciting world of neural networks. We thank the Bundesminister für Forschung und Technologie for supporting this work.

REFERENCES

1 Sadowski, J. and Gasteiger, J. *Chem. Reviews*, in press
2 Kohonen, T. *Biol. Cybern.* 1982, **43,** 59–69
3 Kohonen, T. *Self-Organisation and Associative Memory;* Springer: Berlin, 1989
4 Gasteiger, J. and Zupan, J. *Angew. Chem.* 1993, **105,** 510–536 and *Angew. Chem. Int. Ed. Engl.* 1993, **32,** 502–527
5 Zupan, J. and Gasteiger, J. *Neural Networks for Chemists: An Introduction;* VCH-Publishers: Weinheim, FRG, 1993
6 Gasteiger, J., Li, X., Rudolph, C., Sadowski, J. and Zupan, J. *J. Am. Chem. Soc.* (submitted for publication)
7 Chau, P.L. and Dean, P.M. *J. Mol. Graph.* 1987, **5,** 97–100
8 Witzenbichler, W. *Title.* Dissertation, TU München, 1993
9 Gasteiger, J., Rudolph, C. and Sadowski, J. *Tetrahedron Comput. Method.* 1992, **3,** 537–547
10 Sadowski, J., Rudolph, C. and Gasteiger, J. *Anal. Chim. Acta.* 1992, **265,** 233–241
11 Gasteiger, J. and Marsili, M. *Tetrahedron* 1980, **36,** 3219–3228
12 Gasteiger, J. and Saller, H. *Angew. Chem.* 1985, **97,** 699–701 and *Angew. Chem. Int. Ed. Engl.* 1985, **24,** 687–689
13 Li, X., Gasteiger, J. and Zupan, J. *Biol. Cybern.*, in press
14 Connolly, M.J. *Science* 1983, **221,** 709–713
15 Lee, B. and Richards, F.M. *J. Mol. Graph.* 1971, **55,** 379–400
16 Gasteiger, J. and Li, X. In *Software-Development in Chemistry 7* (D. Ziessow, Ed.), GDCh-Fachgruppe CIC, (1993)