Physicochemical Effects in the Representation of Molecular Structures for Drug Designing

Johann Gasteiger*[1,2]

[1]*Computer-Chemie-Centrum and Institute of Organic Chemistry, University of Erlangen-Nuremberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany*

[2] *Molecular Networks GmbH, Nägelsbachstrasse 25, 91052 Erlangen, Germany*

Abstract: After the identification of a biological target, drug design is to analyze the relationships between the structure of potential ligands and their biological activity. A hierarchy of structure representation is presented here considering either the constitution of a molecule, its 3D structure, or the molecular surface. At each level, a variety of physicochemical effects can be accounted for. Furthermore, the special requirements of learning algorithm, such as neural networks, are taken into consideration. Application to problems from combinatorial chemistry, lead identification, high-throughput screening, and prediction of ADME-Tox properties are given.

Keywords: constitution, 3D structure, conformational flexibility, molecular surfaces, charge distribution, neural networks, lead discovery, high-throughput screening.

1. INTRODUCTION

Orally administered drugs will always be limited to relatively small molecules. This is largely dictated by the requirement for intestinal absorption. Thus, sooner or later, the development of a new drug will require analysis of the relationships between molecular structure and biological property. Combinatorial chemistry and high-throughput screening generate a huge bulk of data that has to be analyzed by automatic learning and data mining methods in order to extract knowledge that can be directed to a more focused synthesis of new lead structures and to the optimization of such lead compounds. On a wider level, molecular structures have to be analyzed also to predict their physical and chemical properties to evaluate ADME-Tox properties, and thus such as solubility, intestinal absorption,and metabolism already in the early stages of drug development (see article by Lombardo et al. in this issue). The relationships between the molecular structure and biological activity are too complex to be calculated from first principles. Therefore, resort has to be taken to automatic learning by algorithms, such as statistical or pattern recognition methods or neural networks. These methods establish a relationship between an object and its properties by being presented with a series of training data gained from experiments. Establishment of a structure-property relationship critically hinges on the representation of molecular structures. Chemists have developed a variety of methods for representing and communicating structure information. The most widely used, international language is a structure diagram; it is still the method of choice when representing chemical reactions. For a more in-depth analysis, three-dimensional molecular models are built, either by mechanical molecular model kits, or, increasingly, by computer modeling. A variety of representations is available, from framework, through ball and stick, to spacefilling models. An even more refined analysis of molecules, particularly when studying biological activity, has to consider molecular surfaces (see Figure **1**).

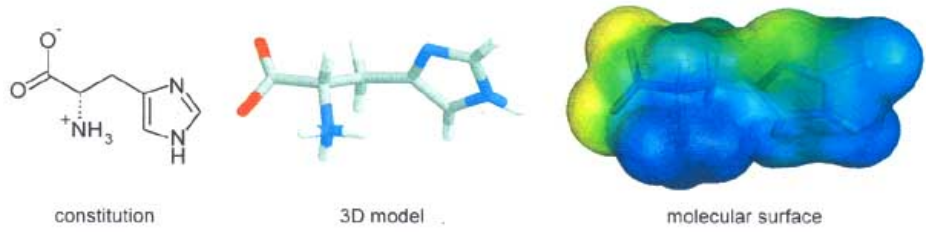

Fig. (1). A hierarchy of structure representation: constitution, 3D structure, molecular surfaces.

*Address correspondence to this author at the Computer-Chemie-Centrum and Institute of Organic Chemistry, University of Erlangen-Nuremberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany; E-mail: gasteiger@chemie.uni-erlangen.de; WWW: http://www2.chemie.uni-erlangen.de

[2]Molecular Networks GmbH, Nägelsbachstrasse 25, 91052 Erlangen, Germany; E-mail: info@mol-net.de; WWW: http://www.mol-net.de

1389-5575/03 $41.00+.00

Whatever the level of approximation, it is clear that at each level a variety of physicochemical properties, such as electron distribution, polarizability, or hydrogen bonding, have to be considered.

All these various representations of chemical structures have to be translated into a form amenable to computer manipulation. In the following, we will present various techniques for encoding these different forms that the chemists use for structure representation, from the constitution of a molecule, through 3D structures to molecular surfaces. These different encoding methods have been developed for the different requirements made by the intended applications. Furthermore, the size of the datasets under consideration strongly influences which coding method can be applied. Datasets of hundreds of thousands or millions of structures have to rely on rather rapid encoding procedures in order to be handled in a reasonable amount of time.

2. CONSTITUTION, 3D STRUCTURE, MOLECULAR SURFACES

2.1. Constitution

It has become common usage to represent and store chemical structure information as connection tables (CT). A connection table provides lists of the atoms and of the bonds in a molecule. Thus, a connection table reflects the constitution of a molecule. In effect, it reproduces one valence bond structure of a molecule. Such a representation has also been called a topological representation as it shows the relationships (bonds) between the atoms of a molecule. Quite often, connection tables are called 2D representations because they should reflect the 2D structure diagram. However, it should be kept in mind that a CT carries no genuine two-dimensional information. A CT gives the molecular graph, with atoms being the nodes and bonds being the edges. Thus, a variety of problems in processing chemical structure information can be operated with methods from graph theory.

However, software has been developed that allows the construction of a 2D structure diagram from a CT. On top of that, we will see in the next section (2.2) that even a 3D molecular model can be constructed from the information contained in a CT.

The important point is that much information on chemical compounds is contained in databases with the chemical structures coded as connection tables. Typically, pharmaceutical companies have in-house databases with several hundred thousands to sometimes more than a million chemical structures stored as CT. This opens the door to a wide range of structure coding methods and thus also to a variety of information processing and chemical data analysis applications.

All structure coding methods and applications reported in this paper only need a connection table as input. Section 2.2 shows how a 3D structure can be obtained from a CT and, building on that, molecular surface properties can be calculated as shown in Section 2.3. All the methods for the calculation of physicochemical effects presented in Chapter 3 also need only the connection table of a molecule as input.

2.2. 3D Structures

Presently, about 250,000 experimentally obtained 3D structures from X-ray crystallography are stored in the Cambridge Structural Database (CSD) of the Cambridge Crystallographic Data Center [1]. Large as this number might seem on first sight, it is minute in comparison to the more than 25 million known organic compounds and the many orders of magnitudes larger number of conceivable organic molecules. Thus, automatic 3D structure generators are needed to fill the gap between experimentally known 3D structures and the many more molecules of interest in drug design, particularly, the large number of compounds studied in combinatorial chemistry.

Organic molecules have rather clearly defined construction principles. The prevalent atoms, C, N, O, H, S, P, F, Cl, Br, I, are bound together in molecules with bond lengths of very limited variability and bond angles that attain standard values. Any changes in bond angles, such as those inflicted in ring systems, can be achieved only by an increase in energy introducing strain. The larger the deviations from the standard values in bond angles, the higher the strain energy. Nevertheless, bond angles can be distorted by up to 30°, particularly so in polycyclic systems. On the other hand, rotations around bonds can usually be achieved quite easily making a multitude of conformations attainable within a narrow energy window.

The construction principles of organic molecules have been implemented in a rule and data based approach to generate a 3D model of an organic molecule from information on the constitution, the set of atoms and bonds, only (see Figure **2**) [2-4].

This system CORINA (COoRdINAtes) is quite general, automatically building a 3D model of any organic and many organometallic molecules from the constitutional information, from the information stored in a connection table. Table 1 shows the results obtained in the conversion of the public database (see article by Tetko in this issue) of the National Cancer Institute (NCI) [5].

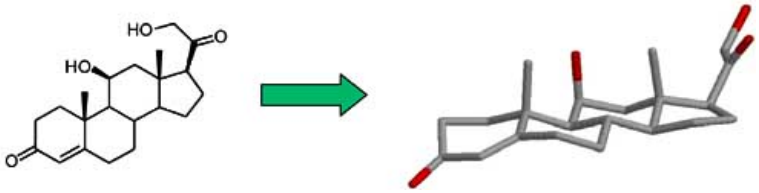

Fig. (2). The generation of a 3D molecular model from a chemical graph.

Table 1. Conversion of the NCI Database into 3D Structures by CORINA

number of compounds	structures converted	CPU time[a]
249,081 (100%)	247,821 (99.5%)	2:40 h (0.04 s/mol)

[a] on a Linux workstation, Intel Pentium III, 600 MHz.

CORINA generates a single low energy conformation of a molecule. It has already been said that molecules with open chain portions can attain many low energy conformations. The receptor bound conformation of a ligand neither must be the conformation with lowest energy from quantum mechanical calculations (in the gas phase) nor the one obtained from X-ray crystallography (in the solid state). Thus, the treatment of conformational flexibility is presently in the center of much attention in drug design. We have built a system, ROTATE, that preferentially generates multiple conformations giving preferences to those conformations observed in X-ray structures and - as established in an independent study [6] - in receptor bound states [7]. This is achieved by using a library of torsional angle patterns derived from a statistical analysis of the CSD file. Figure **3** shows conformations obtained for the HIV-1 protease inhibitor VX-478 and compares one of those conformations with the receptor bound conformation.

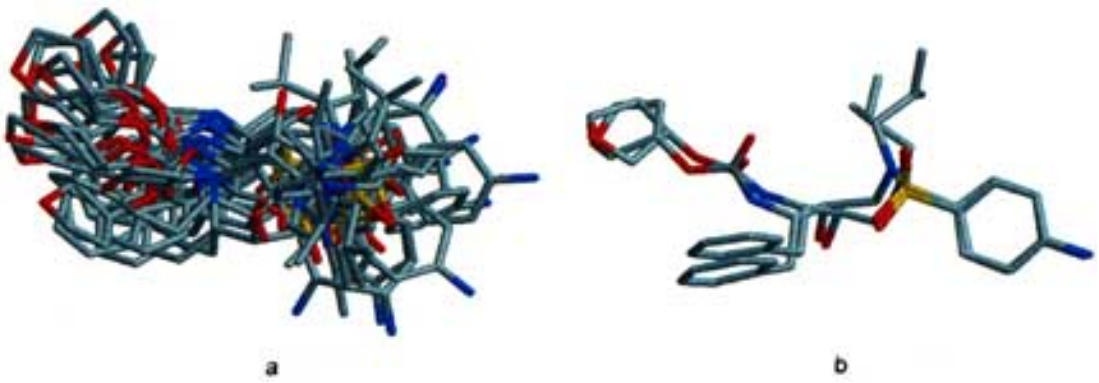

Fig. (3). Multiple conformations obtained with ROTATE for the HIV-1 protease inhibitor VX-478 (a) and comparison of one of these conformations with the receptor-bound conformation (b).

2.3. Molecular Surfaces

Once a 3D structure of a molecule is available, the surface of a molecule can be generated. Various types of surfaces, such as the van der Waals-surface or the solvent accessible surface, have been defined. In drug design, such surfaces that are obtained by moving a probe sphere with a specific radius (e.g., that of a water molecule) across a molecule have found most wide-spread use.

The generation of a molecular surface is a prerequisite for the calculation of molecular surface properties (*vide infra*). It has to be emphasized that molecules interact with each other through molecular surfaces and that any detailed understanding of the interaction of a ligand with its biological receptor has to take account of molecular surfaces and their properties.

2.4. Molecular Fields

Molecules will exert influences on other molecules, which reach even beyond the envelope of a molecular surface. A molecule generates different types of fields, such as an electrostatic field, a hydrophobic field, a lipophilicity field, or a polarizability field, that reach out into space and fall off with different distance dependences.

In fact, a description of a molecule, which is quite widely used in drug design, is a Comparative Molecular Field Analysis (CoMFA, see article by Migliavacca in this issue). In this approach, a molecule is put into a cubic box and a molecular field is calculated for the points in this box that are outside of the molecule. A Partial Least Squares (PLS) analysis is then performed to reduce the high number of molecular field values to a manageable number.

3. PHYSICOCHEMICAL EFFECTS

A detailed view into the electronic structure and various physicochemical properties of a molecule can be obtained by quantum mechanical calculations at various levels of sophistication. However, many applications in drug design ask for the processing of large datasets of hundreds of thousands, or millions of structures. For such problems, quantum mechanical calculations are unfeasible because they require large computation times.

We have therefore developed simple empirical methods for the estimation of fundamental physicochemical effects, such as charge distribution, inductive, resonance, polarizability, or steric effect. Most of these methods only consider the constitution of a molecule in order to allow the rapid processing of large datasets of molecules.

3.1. Charge Distribution

The electronic structure of a molecule has profound influences on many properties. A rather simple, nevertheless quite useful and still wide-spread picture assigns the valence electron distribution to the individual atoms of a molecule arriving at partially charged atoms. A simple empirical model has been developed [8] that is quite general and has such short computation times to allow the processing of

large datasets of millions of structures. (cf. Table 2 *vide infra*).

The method builds on the concept of electronegativity as defined by Mulliken on the basis of ionization potentials, IP, and electron affinities, EA (see eq 1).

$$\chi_{iv} = \frac{1}{2}\,(IP_{iv} + EA_{iv}) \qquad \text{(eq 1)}$$

In specific, an electronegativity value is assigned to each orbital of an atom in a particular hybridization state. Furthermore, the electronegativity of an orbital, χ_{iv}, is considered as depending on the partial charge of the considered atom, *i*. For this dependence, a polynomial of degree two is taken (eq 2).

$$\chi_{iv} = a + bq_i + cq_i^2 \qquad \text{(eq 2)}$$

On formation of bonds between different orbitals, their electronegativities equilibrate but they do so only partially because of changes in orbital electronegativities on incorporation of atoms into molecules. A simple iterative procedure has been developed for Partial Equalization of Orbital Electronegativities (PEOE). The PEOE procedure can be applied to σ-bonded systems.

Conjugated π-systems need additional precautions since π-electrons are quite movable and can be delocalized over more than two atoms. For such systems, first the charge distribution in the σ-bonded skeleton is calculated. Then, the π-electron distribution adjusts to the partial charges in the σ-skeleton by a modified Hückel Molecular Orbital Treatment. The partial charges on the atoms are then obtained from the σ- and the π-charges.

The charges calculated by the PEOE method and its extension to conjugated systems have been correlated with a variety of physical and chemical data thus showing their physicochemical significance. As an example, the dipole moment is taken as a measure of the quality of the calculated values (Figure **4**) [9].

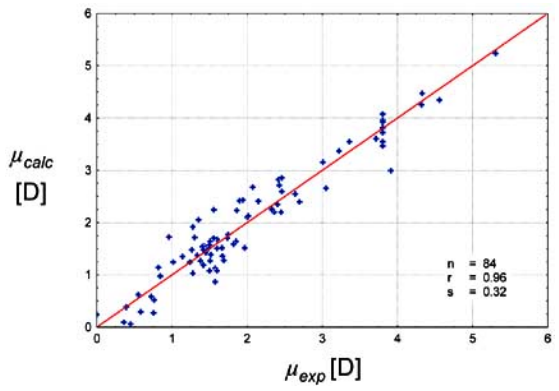

Fig. (4). Comparison of experimental dipole moments and those calculated from partial charges.

3.2. Polarizability Effect

Polarizability is a dominant factor in the stabilization of charges in molecules introduced by protonation, or deprotonation, or by hydrogen bonding. An additivity scheme can be used to estimate mean molecular polarizability. In order to account for distance dependence of the polarizability effect, a simple damping model was used when considering the contribution of each atom (eq 3) [10].

$$\alpha_{eff,j} = \sum_i \alpha_i . 0.5^{(n_{ij}-1)} \qquad \text{(eq 3)}$$

In equation 3 the effective polarizability $\alpha_{eff,j}$ of atom *j* is obtained by the sum of the atomic contributions α_i of each atom *i* in a molecule only to an extent that accounts for the number of bonds, n_{ij}, between atoms *i* and *j*.

The chemical significance of the effective polarizability values has been shown through correlations with quantitative data on chemical reactions, in particular, with proton affinities [10] and gas phase acidities. Intentionally, gas phase data were studied in order to analyze the inherent physicochemical effects of molecules, uncorrupted by the influence of solvents.

3.3. Additional Physicochemical Effects

Electronegativity is the inherent potential of an atom to attract electrons. This potential of an atom has to be modified when an atom is embedded in a molecular environment. Concomitant with the values of partial charges, the PEOE method also provides values of residual electronegativity for each orbital of an atom according to eq 2. These values of residual electronegativitiy can be taken as quantitative measures of the inductive effect as shown through correlation of reactivity data, in particular, those measured for gas phase reactions [11].

The methods for the calculation of charge distribution, polarizability, inductive and resonance effect, hydrogen bonding properties, as well as some additional descriptors have been collected in the program package PETRA (Parameter Estimation for the Treatment of Reactivity Applications), which is also accessible on the web [12]. Table 2 gives results for the calculation of all physicochemical properties implemented in PETRA for some medium-sized databases.

Table 2. Calculation of Physicochemical Properties Implemented in PETRA for Some Medium-Sized Databases

dataset	number of compounds	CPU time[a]	rejected compounds
Acros catalog	13,412	0:32 h (0.14 s/mol)	20 (0.1%)
Maybridge database	54,543	2:41 h (0.18 s/mol)	0
NCI database	249,081	17:12 h (0.25 s/mol)	822 (0.3%)

[a] on a Linux workstation, Intel Pentium III, 600 MHz.

4. UNIFORM STRUCTURE REPRESENTATION

4.1. Inductive Learning Methods

The analysis of a dataset of molecular structures by automatic learning methods, such as pattern recognition methods or neural networks needs to have all individual objects, in this case, the molecular structures, to be represented by the same number of descriptors (see Figure

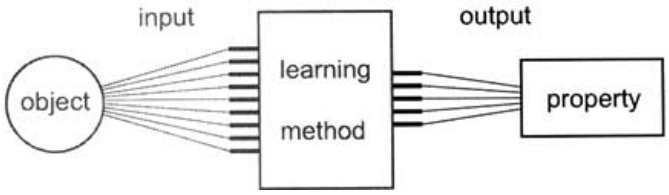

Fig. (5). Establishment of a structure-property relationship by an automatic learning method: all structures have to be represented by the same number of descriptors.

5). As a consequence, datasets comprising molecules of different size, i.e., with different numbers of atoms, cannot directly be investigated by atom-centered properties. Rather, mathematical transformations have to be invoked to arrive at a uniform structure representation having the same number of descriptors for each molecule irrespective of its size.

In the following three sections we will present such mathematical transformations amenable to the constitution, the 3D structure, and molecular surfaces. The use of these structure-encoding schemes will be illustrated with various problems encountered in drug design.

We analyze such problems usually with various artificial neural networks [13]. We are strongly convinced that the analysis of a dataset should be made by a sequence of unsupervised and supervised learning methods. This is true so irrespective of whether one employs statistical or pattern recognition methods or artificial neural networks. Investigations by unsupervised learning have to find out the specific structure coding particularly appropriate for the problem being studied. Only when one has established that the structure coding chosen can reflect the problem under investigation, should one switch to a supervised learning method to establish a quantitative model for predicting the property of interest. In the following examples we only use the unsupervised learning technique of a self-organizing neural network (Kohonen network) [14] to underline the importance of the various structure encoding schemes to reflect biological activity.

4.2. Coding the Constitution

The structural diagram can be considered as a mathematical graph; graph theory has therefore played a major role in the computer handling of structure information. In order to transform the information inherent in a structure diagram into a uniform fixed-length representation, an autocorrelation function was used (eq 4).

$$A(d) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} \delta_{ij}\,(d) p_j\, p_j \qquad \text{(eq 4)}$$

A value for the autocorrelation function A, at a certain topological distance (number of bonds), d, is calculated by summation over all products of a certain property, p, of atoms i and j having the required distance, d.

A range of properties, such as partial atomic charges, measures of the inductive, resonance, or polarizability effect calculated by the program package PETRA (see section 3), were used for autocorrelation. With seven such properties, p, and seven topological distances, d = 2...8, each molecule was represented by a 49-dimensional vector. It could be shown that such a representation can distinguish between dopamine and benzodiazepine agonists (Figure **6**) [15].

The separation of the two types of agonists can clearly be seen. The separation was even maintained after projection of this 49-dimensional space into two dimensions by a Kohonen neural network. Of even more importance is the fact that dopamine and benzodiazepine agonists could still be distinguished when contained in a dataset of more than 8,000 compounds of a chemical supplier catalog. These two types of compounds were found in limited and separated regions of a Kohonen map [15].

Thus, this study showed where benzodiazepine or dopamine agonists have to be sought and in which region of chemical space no such activity is to be anticipated. This methodology is used in industry for the comparison of large inhouse compound collections. It can also be used for the definition of similarity and diversity of combinatorial libraries [16].

4.3. Coding the 3D Structure

With a 3D structure accessible for practically any organic molecule (see section 2), the problem is then, how to encode the 3D structure under the restriction of having to come up with a fixed number of variables, independent of the number of atoms in a molecule. Clearly, again, autocorrelation of atomic properties as given by eq 2 now using genuine spatial distances could be used. However, we were seeking

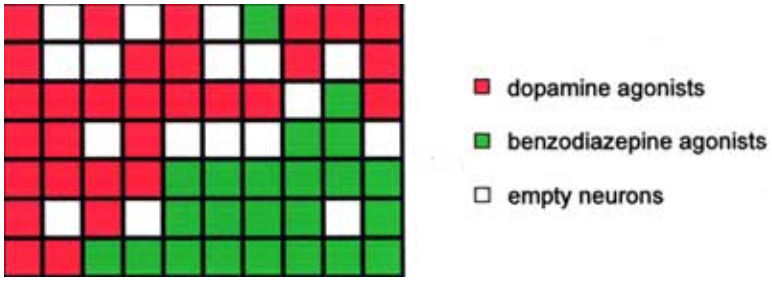

Fig. (6). Self-organizing map obtained from a dataset of 112 dopamine and 60 benzodiazepine agonists.

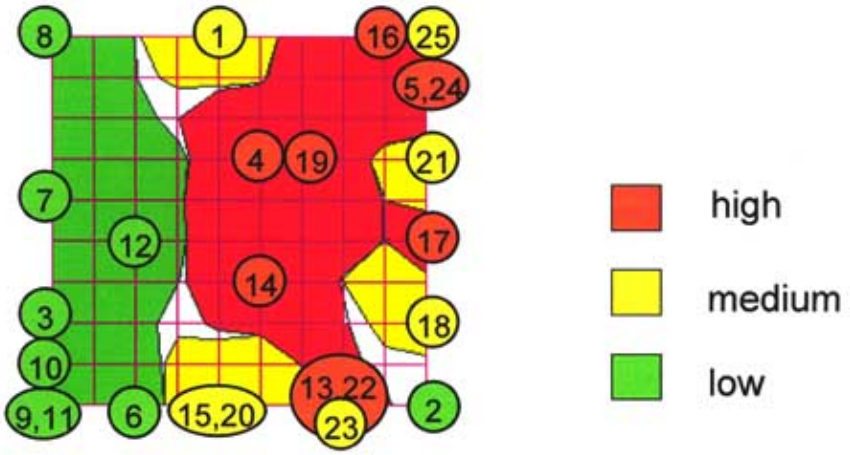

Fig. (7). Clustering of 25 steroids binding to the corticosteroid binding globulin receptor in a Kohonen network.

for structure representations that offer the possibility of regaining the 3D structure of a molecular code.

Building on equations used for obtaining the 3D structure of a molecule from electron diffraction experiments, the encoding procedure embodied in eq 5 was developed [17].

$$g(r) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} A_i A_j e^{-B(r-r_{ij})^2} \qquad \text{(eq 5)}$$

In this equation, $g(r)$ is the radial distribution function (RDF), A_i and A_j are atomic properties such as atomic number, or partial charges, and r_{ij} is the distance between the atoms i and j; N is the number of atoms in the molecule.

This RDF code was mainly used for the simulation of infrared spectra. However, it could also be demonstrated that this code shows great promise for correlating structure with biological activity. Thus, a dataset of 31 steroids showing different binding affinities for the corticosteroid binding globulin (CBG) receptor could be clustered into those having high, intermediate, and low affinity [18] (see Figure **7**).

4.4. Coding of Molecular Surface Properties

Molecules interact with each other at molecular surfaces. This is particularly true for the interaction of a ligand binding to its receptor. The investigation of molecular surfaces, the coding of surface properties, is therefore of primary importance.

A rapid access to the molecular electrostatic potential (MEP) can be gained through the use of the partial atomic charge values. A unit positive point charge is moved across the molecular surface and at each point the electrostatic potential is calculated in a classical manner through Coulomb's law by calculating the interactions between the point charge and the partial charges on the various atoms. Furthermore, simple empirical methods for the calculation of the hydrogen bonding potential (HBP) and of the hydrophobicity potential (HPP) have been developed.

Autocorrelation can also be used to encode surface properties. Equation 2 is now modified such that the properties, p, are sampled on a molecular surface; for the distance parameter, d, all distances within a certain range, e.g., between 3 and 4 Å, are collected in one autocorrelation coefficient.

In a study, results from a high-throughput screening experiment were used to develop a filter that could separate hits from non-hits. Data were obtained from an assay investigating 5,513 hydantoins obtained from a library built from a set of amino acids, aldehydes, and isocyanates. This data set contained 185 hits. Six different structure-encoding schemes were investigated. Encoding of the structures by

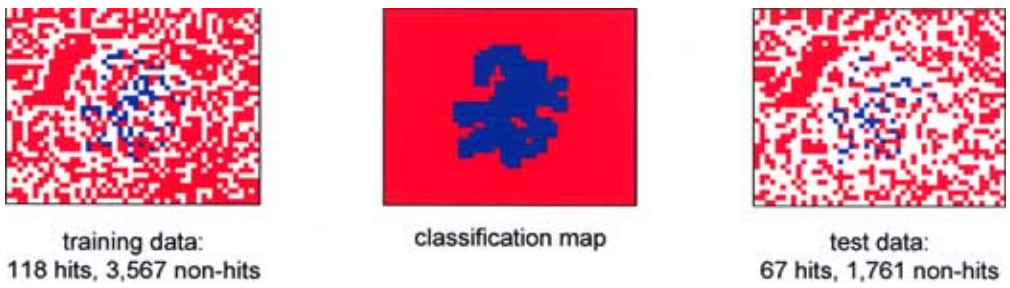

Fig. (8). Self-organizing maps obtained from a dataset of 5,513 hydantoins investigated by a HTS. Assay with the molecules encoded by autocorrelation of the hydrogen bonding potential.

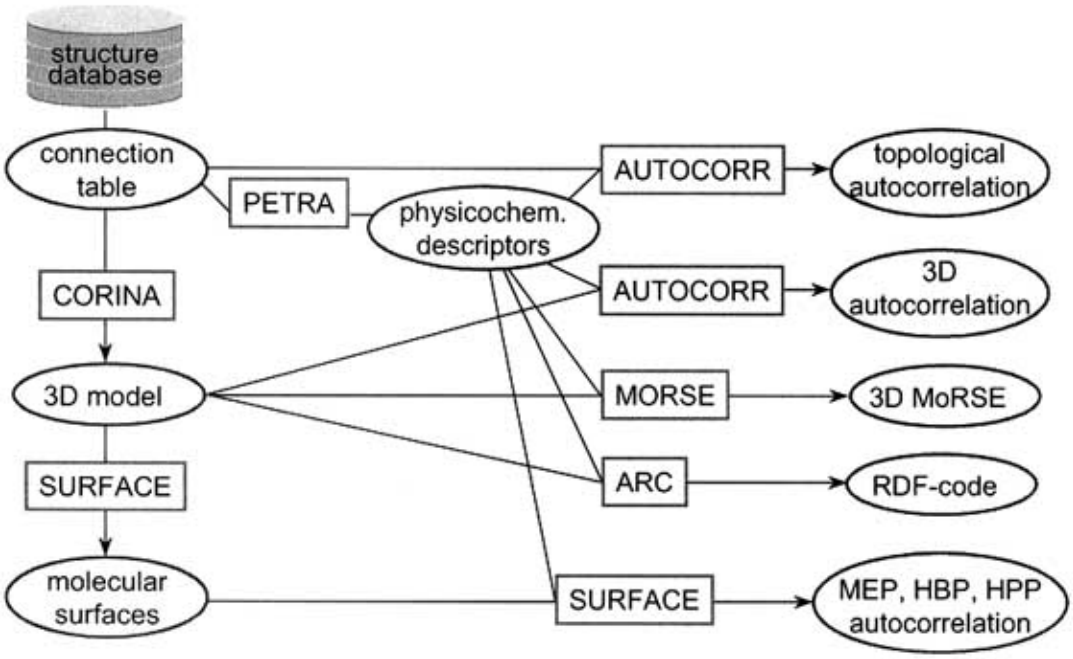

Fig. (9). Suite of programs for deriving a hierarchy of structure coding methods (see text).

Daylight fingerprints of length 256, 572, or 1024 hits proved to be totally inadequate for separating the hits of the assay from the non-hits. Then, three different molecular surface properties, the electrostatic potential, the hydrogen bonding potential, and the hydrophobicity potential were encoded by autocorrelation. The best results were obtained when the molecules were represented by autocorrelation of the hydrogen bonding potential (Figure **8**).

Two thirds of the molecules were used for training a Kohonen network. This network was converted into a filter that assigned 66 (96%) of the 67 hits of the test dataset to active compounds and 1,619 (92%) of non-hits to the non-active compounds. Based on these results a more focused library could be synthesized.

5. SUMMARY

Approaches to the encoding of molecular structures have been developed that allow the investigation of datasets of diverse molecules by learning methods. These structure representations form a hierarchy of increasing sophistication. The methods developed here have been encased in computer programs that can be invoked in different sequences and set-ups. Figure **9** shows the various modules and the way they can be combined.

On the left-hand side, one can see how the geometry of a molecule can be considered in increasing sophistication. From the constitution of a molecule as encoded in a connection table and stored in a database, CORINA [4] generates a three-dimensional molecular model. This 3D model then allows SURFACE to calculate molecular surfaces.

The constitution of a molecule is sufficient for allowing the calculation of the physicochemical effects presented here, such as charge distribution, effective polarizability, residual electronegativity, the resonance effect. These methods have been collected in the program package PETRA [12].

The values of the physicochemical effects calculated with PETRA can be transformed by autocorrelation, either on the level of the constitution to give an encoding of a molecule by topological autocorrelation, or on the level of the 3D structure of a molecule to give an encoding by 3D autocorrelation. Alternative to autocorrelation, an encoding of the physicochemical values obtained with PETRA and accounting for the 3D molecular structure can be performed by the MORSE code [18] or by radial distribution functions with the program ARC [17].

Molecular surface properties, such as the molecular electrostatic potential (MEP), hydrogen bonding potential (HBP), or the hydrophobicity potential (HPP) can be transformed by autocorrelation with the program SURFACE.

The level of structure representation chosen will largely be dictated by the size of the dataset to be investigated. Representations of the constitution will be applied to datasets comprising millions of structures, whereas representations of molecular surface properties can still be chosen for datasets comprising 100,000 and more structures. Even with large datasets these methods are rapid enough to be performed on small workstations with computation times of a few hours. Consideration of conformational flexibility is presently limited to smaller sets of structures.

ACKNOWLEDGEMENTS

All the coworkers whose names appear in the references developed the methods. The Federal Minister of Education and Research (BMBF) funded our research in perticular.

Support from Merck KGaA (G. Barnickel, S. Anzali, M. Krug), BASF AG, and the Universities of Marburg (G. Klebe) and Tuebingen (A. Zell) is gratefully acknowledged.

REFERENCES

[1] (a) Allen, F.H.; Davies, J.E.; Galloy, J.J.; Johnson, O.; Kennard, O.; Macrae C.F.; Mitchell, E.M.; Mitchell, G.F.; Smith, J.M.; Watson, D.G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187-204. (b) Allen, F.H.; Hoy, V.J. Cambridge Structural Database. In Encyclopedia of Computational Chemistry; Schleyer, P.V.R.; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer, III, H.F.; Schreiner, P.R., Eds.; John Wiley & Sons, Inc.: Chichester, UK, **1998**, 155-167.

[2] Gasteiger, J.; Rudolph, C.; Sadowski, J. *Tetrahedron Comput. Method,* **1992**, *3*, 537-547.

[3] Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000-1008.

[4] CORINA Version 3.0 is available from Molecular Networks GmbH, Nägelsbachstr. 25, 91052 Erlangen, Germany (http://www.mol-net.de) and can be tested via the Internet at http://www2.chemie.uni-erlangen.de/software/corina.

[5] The open part of the NCI database is accessible via the Internet at http://cactus.nci.nih.gov/ncidb2/download.html.

[6] (a) Klebe, G.; Mietzner, T. *J. Comput. Aided Mol. Des.* **1994**, *8*, 583-606. (b) Klebe, G.; Mietzner, T.; Weber, F. *J. Comput. Aided Mol. Des.* **1999**, *13*, 35-49.

[7] Schwab, C.H.; 3D Structure Generation and Conformational Searching, in Computational Medicinal Chemistry and Drug Discovery, Bultinck, P.; De Winter, H.; Langenaeker, W.; Tollenare J.P., Eds., Dekker Inc., New York, In Press (**2003**).

[8] Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3228.

[9] Gasteiger, J.; Kleinöder, T.; Fato, M. Unpublished results.

[10] Gasteiger, J.; Hutchings, M.G. *Chem. Soc. Perkin 2* **1984**, 559-564.

[11] Hutchings, M.G.; Gasteiger, J. *Tetrahedron Lett.* **1983**, *24*, 2541-2544.

[12] PETRA Version 3 is available from Molecular Networks GmbH, Nägelsbachstr. 25, 91052 Erlangen, Germany (http://www.mol-net.de) and can be tested via the Internet at http://www2.chemie.uni-erlangen.de/software/petra.

[13] Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*; Second Edition; Wiley-VCH-Verlags GmbH: Weinheim, **1999**.

[14] (a) Kohonen, T. *Biol. Cybern.* **1982**, *43*, 59-69. (b) Kohonen, T. *Self-Organization and Associative Memory*; third Edition; Springer Verlag: Berlin, **1989**. (c) Kohonen, T. *Self-Organizing Maps*; Huang, T. S., Kohonen, T., Schröder, M.R.; Eds., Springer Verlag: Berlin, 1995.

[15] Bauknecht, H.; Zell, A.; Bayer, H.; Levi, P.; Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1205-1213.

[16] Sadowski, J.; Wagener, M.; Gasteiger, J.; *Angew. Chem.*, **1995**,*107*, 2892-2895 and *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2674-2677.

[17] Hemmer, M.C.; Steinhauer, V.; Gasteiger, J. *Vibrat. Spectroscopy* **1999**, *19*, 151-164.

[18] Schuur, J.H.; Selzer, P.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334-344.