Journal of Medicinal Chemistry

Volume 49, Number 22 November 2, 2006

2006 American Chemical Society Award for Computers in Chemical and Pharmaceutical Research

Of Molecules and Humans

Johann Gasteiger†

Computer-Chemie-Centrum, University Erlangen-Nuremberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany

Received July 28, 2006

The understanding and modeling of the relationships between the structure of a molecule and its biological activity are the central themes of medicinal chemistry. Clearly, the biological activity of a compound cannot directly be calculated by first principles from the molecular structure. In this situation, an indirect approach has to be taken to find a relationship between the molecular structure of a compound and its biological activity (Figure 1). First, the structure of a molecule has to be represented by structure descriptors. Then these structure descriptors are used for modeling the biological activity. This second step asks for a set of molecules and their associated biological activities in order to apply inductive learning methods such as statistical or pattern recognition methods or artificial neural networks to establish a relationship between the structure descriptors and the biological activity.

In this article we will concentrate on the first step, the representation of molecules. In the past few decades a wide variety of methods have been developed to derive structure descriptors for a molecule.[1−3] We propose here an ordering scheme that allows one to determine the level of sophistication in structure representation. In this endeavor we will compare molecules with humans, both being three-dimensional objects having surfaces, having different types of surfaces, having chiral parts, and being flexible (Figure 2). This comparison will allow us to decide what results we can expect from a certain type of molecular structure representation and thus give guidelines on which structure representation to choose for a certain problem at hand.

Before discussion of various structure representations, another task has to be addressed. The objective of using data analysis

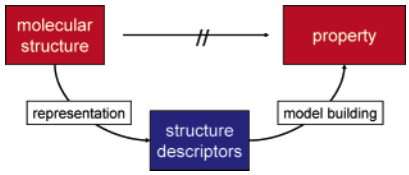


Figure 1. Construction of structure−property relationships.

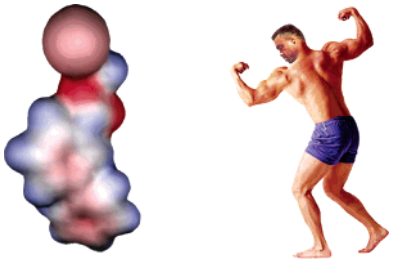


Figure 2. Molecules have shape and surfaces.

methods for relating structure descriptors with activities requires that each structure of the data set has to be represented by the same number of descriptors. This becomes already clear with one of the simplest data analysis methods, a multilinear regression analysis (MLRA) relating a set of independent variables (descriptors) x_{ij} for a molecule i with a property y_i (i.e., the activity) of this molecule:

$$y_i = c_0 + c_1 x_{i1} + c_2 x_{i2} + \ldots c_n x_{in} \quad (1)$$

In this case, it has to be decided from the very beginning how many descriptors, n, will be used to represent each molecule. For a congeneric set of molecules having the same skeleton/scaffold, this might be an easy task by choosing descriptors for the various substituents on the skeleton as was used in linear

† Phone: 0049-9131-8526570. Fax: 0049-9131-8526566. E-mail: gasteiger@chemie.uni-erlangen.de.

10.1021/jm0608964 CCC: $33.50
Published on Web 10/07/2006

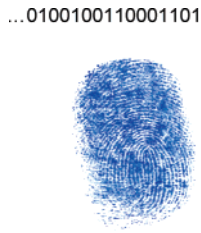


Figure 3. Fingerprint representation of a chemical structure.

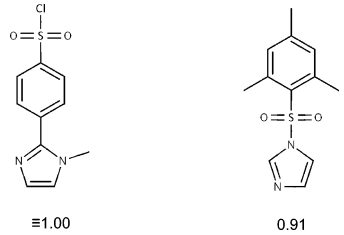


Figure 4. Two structures with a high Tanimoto similarity index.

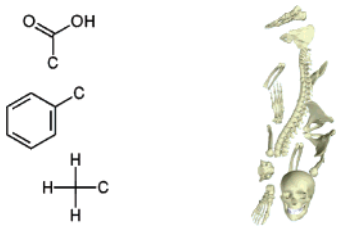


Figure 5. Fragmentation of chemical structures.

free energy models such as the Hammett or Taft type of approaches. However, this requirement that all molecules have to be represented by the same number of descriptors also applies to data sets that comprise quite diverse molecules of different sizes, having different number of atoms. And it also applies to any data analysis method not only to MLRA but to any pattern recognition method or artificial neural network. We will address this point of requiring a fixed-length representation for all molecules of a data set in the following sections.

Fingerprints

Methods have been developed to search molecules for the presence or absence of certain predefined functional groups and other substructures and to compress this information into a bit string of given length. Such a representation is called a fingerprint of a molecular structure.[4] Initially, fingerprints were developed for rapidly searching databases of structures for determining the presence or absence of certain chemical structures (Figure 3). The same purpose is fulfilled by hashcodes[5] that are constructed in a way similar to fingerprints.

With the advent of combinatorial chemistry and thus the need to represent large sets of compounds, fingerprints were offered for the representation of chemical structures to model their biological activity. Figure 3 clearly indicates that fingerprints may be quite appropriate for identifying a molecule; however, they are hopelessly inadequate for representing the finer details of a molecule.

Fragment Codes

Whereas the presence or absence of certain substructures or fragments is compressed with fingerprints into a rather concise representation that does not allow any more the identification of the individual fragments, fragment codes explicitly retain the information about the presence or absence of a certain substructure. In this case, a predefined set of fragments is used and each fragment corresponds to a certain position of a bit string having a length equal to the number of fragments in the predefined set. In some cases, vectors of integers are used, thus allowing us not only to report the presence of a substructure but also to count how often a substructure is present. Fragment codes have been used for modeling a variety of properties such as predicting biological activities or simulating infrared spectra.

Moreover, fragment codes are often used for defining the similarity of structures by calculating the Tanimoto index I from the number n_A of substructures present in structure A but not contained in structure B, the number n_B of substructures present in structure B but absent in structure A, and the number n_C of substructures in common between structure A and structure B:

$$I = \frac{n_C}{n_A + n_B + n_C} \tag{2}$$

However, it has to be realized that fragment codes and lists of fragments only report the presence or absence of certain fragments and do not give information on how these fragments are arranged in a given molecule, what the distance between these substructures is. Thus, the two structures shown in Figure 4 have a high similarity based on the Tanimoto index, although clearly for many types of problems, such as questions on how to synthesize them, these two molecules have to be considered to be rather different.

The comparison of a fragment code with the corresponding level of information on a human being shows the strength and limitations of a fragment code (Figure 5). Yes, a fragment code can tell us whether a certain substructure (the skull) is present in a molecule, but it gives no information on the distance between two substructures (the skull and the foot). Clearly, such information is essential when modeling the biological activity of a compound because the distance between two atoms will be important for deciding whether a ligand can bind to one or two sites of the receptor.

Topological Distances and Atomic Properties

Thus, let us consider in our structure representation the distance between the atoms of a molecule. In the simplest case, the topological distance can be taken, which corresponds to the number of bonds between two atoms. Clearly, it is important to consider not only the distances between two atoms but also the identity of these atoms, in particular their physicochemical properties such as partial charges or hydrogen-bonding potentials.

One approach to simultaneously considering atomic properties and distances between atoms is topological autocorrelation as expressed in the following equation:[6,7]

$$A_{top}(d) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} a_i a_j \, \delta(d - d_{ij}) \tag{3}$$

In this equation, a_i and a_j are properties of atoms i and j, respectively, and d_{ij} is the topological distance between atoms i and j. δ is having a value of 1 when the running variable, the distance d, is equal to the distance d_{ij} between the two atoms; otherwise, its value is zero. The summation is made over all combinations of atoms i and j.

As atomic properties (a_i), any property of an atom such as atomic number or its mass can be taken. However, to represent the electronic properties of atoms, we have developed methods for calculating such important physicochemical effects such as partial charges[8,9] and inductive,[10] resonance,[9] or polarizability

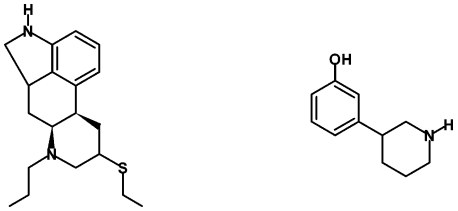


Figure 6. Two dopamine agonists having different number of atoms: 48 atoms for structure on the left and 27 atoms for structure on the right.

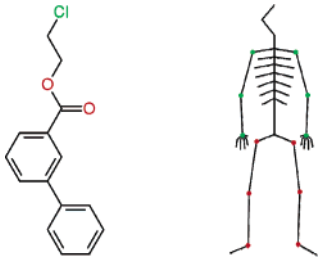


Figure 7. Topology of molecules as expressed by the relative arrangement of the atoms in a molecule.

effects[11] that are based on simple and rapid algorithms that allow the processing of large sets of molecules comprising hundreds of thousands or even millions of structures.

The merits of topological autocorrelation of electronic properties of atoms have been shown in studies distinguishing molecules having different biological activities[7] for finding new lead structures, for lead hopping, and for the comparison of libraries of compounds. This kind of structure representation by topological autocorrelation is able to perceive the similarity of the two structures contained in Figure 6, both being dopamine agonists.[7] With all its successes it should not be overlooked that topological autocorrelation only considers the constitution of a molecule, its set of atoms, and how they are bonded (Figure 7).

3D Structure Representation

Molecules are three-dimensional objects, and any in-depth representation of a molecule should take into account its 3D structure. The first step is then to gather 3D information on the structures of molecules. Presently, for about 250 000 organic and organometallic molecules their 3D structures have been determined by X-ray diffraction or NMR studies and have been stored in the Cambridge Structure Database (CSD). Large as this number might seem at first, it is minute in comparison with the number of known compounds, which exceeds 30 million. Thus, we know the 3D structures of less than 1% of the known compounds. The question is then, can we learn enough rules from the known 3D structure of organic compounds that would allow us to predict the 3D structure of the other 99% of organic compounds? The answer is clearly yes. Several automatic 3D structure generators have been developed that can generate a 3D molecular model from information on the constitution of a molecule only.[12] In our group, the 3D structure generator CORINA has been developed.[13] CORINA has incorporated data and rules on the construction principles of organic compounds that allows the generation of a 3D model for basically any organic molecule.[12,14] Thus, the publicly available database of the National Cancer Institute, containing 250 251 structures could automatically be converted into 3D molecular models in a single run requiring 1.1 h on a PC (1.6 GHz, Linux) and providing 3D models for 99.4% (248 795) of the structures. CORINA produces a single low-energy conformation of a

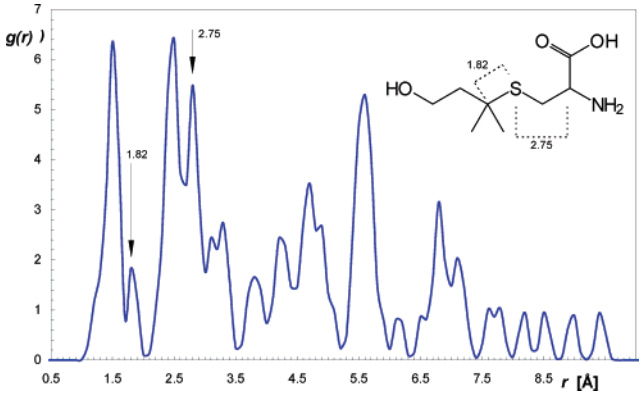


Figure 8. Radial distribution function of a 3D molecular structure.

molecule. Comparison with experimental 3D structures from X-ray structure determination has shown the high quality of the 3D structures.[15]

With automatic 3D structure generators being able to produce 3D molecular models for basically any organic molecule, the question is then, "How can the 3D structures be represented for data analysis methods requiring the same number of descriptors irrespective of the size, the number of atoms in a molecule?" Clearly, the Cartesian coordinates cannot be used because then the number of descriptors would be directly related to the number N of atoms in a molecule requiring $3N$ coordinates. A fixed-length representation of the 3D structure can again be obtained by autocorrelation in an analogous manner as shown by eq 3 with the distance d_{ij} being binned into ranges.

As an alternative, radial distribution functions (RDF) originating in powder X-ray diffraction or electron diffraction studies for the representation of the 3D structure of molecules can be used as shown in the following equation:[16]

$$g(r) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} a_i a_j \, e^{-b(r-r_{ij})^2} \qquad (4)$$

In eq 4 the radial distribution function $g(r)$ is obtained from the product of the properties a_i and a_j of atoms i and j and by considering the distances r_{ij} between those two atoms. The parameter b is the so-called temperature factor, "fuzzifying" the distances. The value of r is a distance and is the running variable of the function.

Figure 8 shows the RDF of the 3D structure of the given molecule. The peak at 1.5 Å results from all bonding distances involving C−C, C−O, and C−N bonds. The peak at 1.82 Å is caused by the longer C−S distance. Next come the 1,3 distances involving two bonds, etc. The last peak in Figure 8 results from the largest distance in this molecule between two oxygen atoms. To obtain a fixed-length representation, the function is only calculated at discrete values of r and only up to a certain distance, e.g., every 0.1 Å up to a value of 12.8 Å resulting in a vector of 128 values.

An RDF code has successfully been used for the simulation of infrared spectra because it encodes the entire 3D structure of a molecule and thus can model the vibrations of an entire molecule, both individual bonds and the entire skeleton.[16−18] RDF codes have a bright future in studies of the effects of the 3D structure on biological activity because they have a quite clear physicochemical interpretation. Thus, recently an RDF code has been used to analyze the NF-κB binding affinity of a series of sesquiterpene lactones.[19] Valuable as 3D structure codes are for the representation of molecules in modeling their

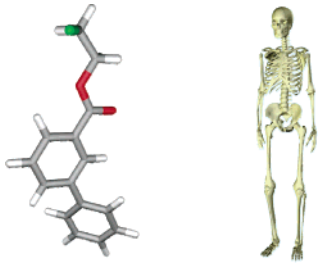


Figure 9. Molecular 3D skeleton.

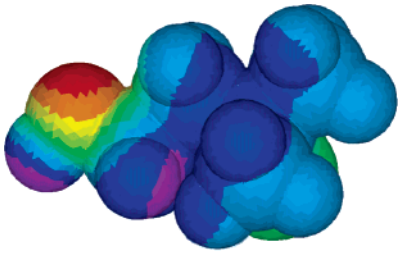


Figure 10. Electrostatic potential of a molecular surface.

biological activity, it should not be forgotten that we have only represented the skeleton of molecules (Figure 9).

Molecular Surface Properties

Molecules, however, have shape and have surfaces (see Figure 2). They interact with their environment through their surfaces and the properties on their surfaces (Figure 10). Again, we are faced with the task of representing the properties on the surfaces of a series of molecules with a fixed-length vector, with the same number of descriptors, irrespective of the size of the molecule. And again, autocorrelation can be used:

$$\mathrm{AC}_{\mathrm{surf}}(d) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} p_i p_j \, \delta(d - d_{ij}) \qquad (5)$$

In this case, properties p of points i and j taken from the molecular surface with a certain sampling density will be used and the distance d will be binned between a lower d_l and an upper bound d_u, $d_l \leq d < d_u$.

It has been shown that through autocorrelation of the molecular electrostatic potential a representation is obtained that is very well suited to modeling the binding affinity of a series of 31 steroids to the corticosteroid binding globuline receptor.[20] Autocorrelation of the molecular electrostatic potential has been used to define the similarity and diversity of combinatorial libraries consisting of amino acids attached to xanthene, cubane, and adamantane scaffolds.[21]

In another study, the representation of three different surface properties, molecular electrostatic potential, hydrophobicity potential, and hydrogen-bonding potential, by autocorrelation and three different fingerprint representations have been compared in their ability in separating hits from nonhits in a combinatorial library of hydantoins.[22] As it turned out, on the basis of autocorrelation of the hydrogen-bonding potential, a filter could be developed that was able to select 96% of the hits from a test set and allowed one to discard 92% of the nonhits.

In another attempt to represent molecular surface properties, two-dimensional maps of molecular surfaces have been produced by a nonlinear mapping procedure utilizing a self-organizing neural network.[23] In this approach, the Cartesian coordinates of points sampled from a molecular surface are used

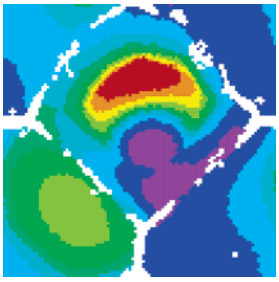
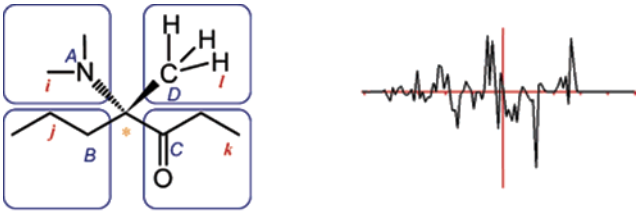


Figure 11. A 2D map of a molecular electrostatic potential.

Figure 12. Chirality of molecules.



$$f_{CIC}(u) = \sum_i \sum_j \sum_k \sum_l s_{ijkl} \cdot \exp\left[-b\left(u - e_{ijkl}\right)^2\right]$$

Figure 13. Chirality code of a molecular structure, derived from all combinations of the atoms of the four ligands.

to train a self-organizing (Kohonen) neural network. The mapping of the surface points into the neurons of the network can be visualized by any property these points had on the surface, e.g., the molecular electrostatic potential (MEP). Figure 10 shows the MEP of the surface of a molecule. Because this is a linear projection, only part of the surface can be shown. Figure 11, on the other hand, shows the self-organizing map of the entire MEP, as this method is a nonlinear projection method being able to map the entire molecular surface into a single plane. It has been shown that such maps of the MEP can be used to distinguish compounds that bind to the muscarinic receptor from those that bind to the nicotinic receptor.[24]

Chirality

All proteins are chiral, and therefore, many receptors and enzymes respond differently to enantiomers. Correspondingly, about 70% of all drugs are chiral. There is a strong tendency in the pharmaceutical industry to bring pure enantiomers into the market. Any more detailed modeling of the effects of structure on biological activity therefore has to represent chirality (Figure 12). In distance space, enantiomers cannot be distinguished. Thus, enantiomers will obtain the same 3D autocorrelation vectors or RDF codes. However, we have developed both a conformation-dependent and a conformation-independent chirality code that is based on the 3D structure of a molecule and that considers all the atoms of the ligands around a chiral center or chiral axis (Figure 13).[25,26]

It has been shown that such chirality codes can successfully be used to predict the major enantiomer in an enantioselective reaction caused by a chiral catalyst.[25] Furthermore, chirality codes were used to predict the first eluted enantiomer in enantioselective chromatography.[26] Thus, the door is open for using chirality codes in modeling the biological activity of different enantiomers.

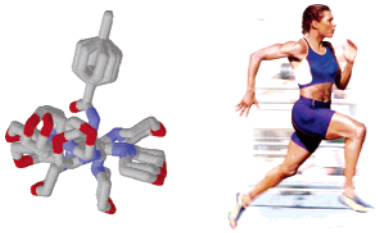


Figure 14. Flexibility of molecules.

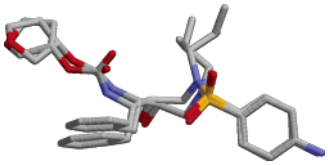


Figure 15. Superimposition of the receptor-bound structure of the HIV protease inhibitor: VX-478 with a conformation generated by ROTATE.

Molecular Flexibility

All structure representations mentioned until now have assumed molecules to be rigid. However, most molecules are quite flexible having single bonds that allow rotation yielding different torsional angles and thus providing different conformations (Figure 14). Then the quest for the biologically active conformation becomes of central importance. Lack of a knowledge of the biologically active conformation is also the reason, in quite a few situations, that topological or 2D descriptors outperform 3D descriptors in modeling biological activity. Clearly, molecules are three-dimensional, and thus, 3D descriptors should perform better than 2D descriptors. However, as soon as the 3D structure of a molecule is considered, the problem of finding the right conformation becomes imminent.

The generation of conformations is an easy task; even the generation of low-energy conformations is not that difficult. However, because of the large number of potential settings for torsional angles, one might soon end up with too many conformations to handle. The question is then how to generate not too many conformations while maintaining the biologically active conformation. Two approaches are conceivable: (1) the constrained generation of conformations, constrained so as not to generate too many conformations; (2) a direct search for the biologically active conformation. Attempts along both lines will be presented here.

An analysis of the distribution of torsional angles around single bonds in X-ray structures showed clear preferences and provided a statistical distribution of the incidences of torsional angles.[27] Such distributions are taken by the program ROTATE to preferentially generate those conformations that have a high incidence in the Cambridge Crystallographic Structure Database (CSD). In addition, conformations with small deviations in torsional angles are collected into families and each family is represented by one conformation only. This allows the generation of a limited but quite diverse set of conformations.[28,29] These sets of conformations also contain a conformation that is quite close to the receptor-bound, biologically active conformation (Figure 15).

The approach of trying to directly access the bioactive conformation rests on the idea that a set of ligands binding to the same receptor must have common spatial features. Thus, a search for the three dimensional maximum common substructure (3D-MCSS) of a set of ligands is initiated by superimposing the 3D molecular models of these ligands to maximize the number of atoms of the different ligands that can be superimposed. In this process, rotations around single bonds of the ligands are allowed, thus introducing conformational flexibility. To manage this optimization problem, a genetic algorithm such as a stochastic optimizer is used.[30] Figure 16 shows the superimposition of three nicotinic allosterically potentiating ligands emphasizing their 3D structural similarity.

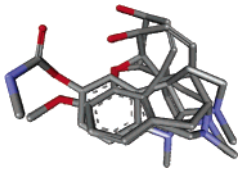


Figure 16. Superimposition by GAMMA of the three nicotinic allosterically potentiating ligands galanthamine, codeine, and physostigmine.

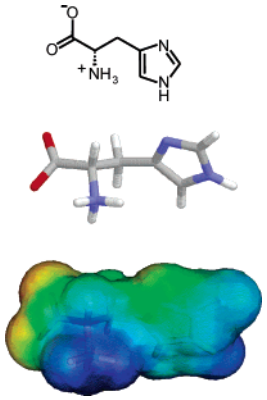


Figure 17. Hierarchy of structure representation: a 2D model, a 3D model, and a molecular surface.

Summary and Conclusions

The comparison of a molecule with a human being allows one to assess the level of structure representation needed (and chosen) for solving the various problems encountered in drug design and development. Here, we have largely concentrated on the geometric aspects of structure representation. However, we want to emphasize that the proper consideration of physicochemical effects exerted by the atoms in a molecule is of equal importance.[2,31] The equations presented here allow their transparent incorporation into the various structure coding methods, starting from the constitution through the 3D structure to molecular surface properties (Figure 17). These methods combining molecular geometry of increasing resolutions with physicochemical properties have been integrated into the package ADRIANA.Code.[32]

Although the discussion here has centered around problems encountered in drug design, the methods for the representation of molecular structures can be used in all areas of chemistry. Because of the need for the prediction of a wide range of physical, chemical, or biological activities of compounds, we will see in the future increasing use of structure coding methods in many fields of chemistry.

Clearly, despite the enormous progress that has been made in recent years in the area of molecular structure representation, there is still a lot of space for further improvement. Particularly, the quest for the biologically active conformation is still a challenging problem being open for new ideas and approaches. It is our belief that the development of new structure representations should rest on clearly defined levels of resolution of the geometry of molecules and on considerations of a variety of physicochemical effects.

Acknowledgment. Our work in this area has been pushed forward by a series of able and dedicated co-workers mentioned in the references. Support of our research came from the German Minister of Science and Technology (BMFT and BMBF), from the German Research Council (DFG), and from several pharmaceutical companies, in particular, Pfizer. I also highly appreciate the recognition of our work by the 2006 American Chemical Society Award for Computers in Chemical and Pharmaceutical Research.

References

(1) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2001.
(2) Gasteiger, J. A Hierarchy of Structure Representations. In *Handbook of Chemoinformatics. From Data to Knowledge*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, Germany, 2003; pp 1034−1061; ISBN 3-527-30680-3.
(3) *Chemoinformatics. A Textbook*; Gasteiger, J., Engel, T., Eds.; Wiley-VCH: Weinheim, Germany, 2003; ISBN 3-527-30681-1.
(4) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(5) Ihlenfeldt, W.-D.; Gasteiger, J. Hash Codes for the Identification and Classification of Molecular Structure Elements. *J. Comput. Chem.* **1994**, *15*, 793−813.
(6) Moreau, G.; Broto, P. *Nouv. J. Chim.* **1980**, *4*, 359−360.
(7) Bauknecht, H.; Zell, A.; Bayer, H.; Levi, P.; Wagener, M.; Sadowski, J.; Gasteiger, J. Locating Biologically Active Compounds in Medium-Sized Heterogeneous Datasets by Topological Autocorrelation Vectors: Dopamine and Benzodiazepine Agonists. *J. Chem Inf. Comput. Sci.* **1996**, *36*, 1205−1213.
(8) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity. A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(9) Gasteiger, J.; Saller, H. Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept. *Angew. Chem.* **1985**, *97*, 699−701; *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687−689.
(10) Hutchings, M. G.; Gasteiger, J. Residual Electronegativity. An Empirical Quantification of Polar Influences and Its Application to the Proton Affinity of Amines. *Tetrahedron Lett.* **1983**, *24*, 2541−2544.
(11) Gasteiger, J.; Hutchings, M. G. Quantification of Effective Polarisability, Applications to Studies of X-ray Photoelectron Spectroscopy and Alkylamine Protonation. *J. Chem. Soc., Perkin Trans. 2* **1984**, 559−564.
(12) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(13) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1992**, *3*, 537−547.
(14) CORINA can be used via the Internet at http://www2.chemie.uni-erlangen.de/software/corina and is available from Molecular Networks GmbH, Erlangen, Germany (http://www.mol-net.com).
(15) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 100−1008.
(16) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. The Prediction of the 3D Structure of Organic Molecules from Their Infrared Spectra. *Vib. Spectrosc.* **1999**, *19*, 151−164.
(17) Schuur, J. H.; Selzer, P.; Gasteiger, J. The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334−344.
(18) Selzer, P.; Gasteiger, J.; Thomas, H.; Salzer, R. Rapid Access to Infrared Reference Spectra of Arbitrary Organic Compounds: Scope and Limitations of an Approach to the Simulation of Infrared Spectra by Neural Networks. *Chem.—Eur. J.* **2000**, *6*, 920−927.
(19) Wagner, S.; Hofmann, A.; Siedle, B.; Terfloth, L.; Merfort, I.; Gasteiger, J. Development of a Structural Model for NF-κB Inhibition of Sesquiterpene Lactones Using Self-Organizing Neural Networks. *J. Med. Chem.* **2006**, *49*, 2241−2252.
(20) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.
(21) Sadowski, J.; Wagener, M.; Gasteiger, J. Assessing Similarity and Diversity of Combinatorial Libraries by Spatial Autocorrelation. *Angew. Chem.* **1995**, 107, 2892−2895; *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2674−2677.
(22) Teckentrup, A.; Briem, H.; Gasteiger, J. Mining High-Throughput Screening Data of Combinatorial Libraries: Development of a Filter To Distinguish Hits from Nonhits. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 626−634.
(23) Gasteiger, J.; Li, X.; Rudolph, Ch.; Sadowski, J.; Zupan, J. Representation of Molecular Electrostatic Potentials by Topological Feature Maps. *J. Am. Chem. Soc.* **1994**, *116*, 4608−4620.
(24) Gasteiger, J.; Li, X. Mapping the Electrostatic Potential of Muscarinic and Nicotinic Agonists with Artificial Neural Networks. *Angew. Chem.* **1994**, *106*, 671−674; *Angew. Chem., Int. Ed. Eng.*, **1994,** *33*, 643−646.
(25) Aires de Sousa, J.; Gasteiger, J. New Description of Molecular Chirality and Its Application to the Prediction of the Preferred Enantiomer in Stereoselective Reactions. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 369−375.
(26) Aires de Sousa, J.; Gasteiger, J. Prediction of Enantiomeric Selectivity in Chromatography. Application of Conformation-Dependent and Conformation-Independent Descriptors of Molecular Chirality. *J. Mol. Graphics Modell.* **2002**, *20*, 373−388.
(27) Klebe, G.; Mietzner, T. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583−606.
(28) Schwab, C. H. Conformational Analysis and Searching. In *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, Germany, 2003; pp 262−301.
(29) Renner, S.; Schwab, C. H.; Gasteiger, J.; Schneider, G. Impact of Conformational Flexibility on Three-dimensional Similarity Searching Using Correlation Vectors. *J. Chem. Inf. Model.*, published online August 8, 2006, http://dx.doi.org/10.1021/ci050075s.
(30) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 200−232.
(31) Gasteiger, J. Physicochemical Effects in the Representation of Molecular Structures for Drug Designing. *Mini-Rev. Med. Chem.* **2003**, *3*, 789−796.
(32) *ADRIANA.Code*; Molecular Networks GmbH: Erlangen, Germany; http://www.mol-net.com.

JM0608964