Computer-Assisted Molecular Design (CAMD)—An Overview

By Horst Frühbeis,* Robert Klein, and Holger Wallmeier

Dedicated to Professor Heinz Harnisch on the occasion of his 60th birthday

A new instrument, long established as CAD in engineering, is beginning to make its presence felt in chemical research laboratories: Computer-Assisted Molecular Design (CAMD). The combined use of computer graphics and theoretical chemistry is opening up new perspectives in molecular research. Structures and properties of molecules such as spacefilling, charge distribution, or dynamic behavior can be determined and used for comparison. For research on complex systems like biomolecules (protein engineering), this new approach turns out to be indispensable.

1. Introduction

Computer-assisted molecular design is a new approach to molecular research using methods from theoretical chemistry. The essential feature is the use of models and their realization on a computer. This may be regarded as the continuation of an old tradition. Many important concepts of chemistry have been worked out using models.[1] As an example take *van't Hoff* and *Le Bel*'s tetrahedron model of the carbon atom, which gave rise to the terms valency and isomerism in organic chemistry. With his model of the benzene molecule, *Kékulé* likewise prepared the way for the concepts of delocalization, conjugation, and resonance. In developing the model of the DNA double helix, *Watson* and *Crick*[2] opened up a new dimension in the understanding of specificity.[3]

Concomitant with the increasing interest in large systems more and more detailed problems are being encountered, and there is a growing tendency to transfer molecular models from the chemist's desk to the computer. This offers the possibility of providing models with well-defined properties and boundary conditions, and thus obtaining quantitative information. The mathematical instruments of theoretical chemistry and many-particle physics can be used. Computer graphics as a tool for handling models has played a decisive role in this development.

The interest of industrial research in computer-assisted molecular design stems from the desire to make the search for new drugs and plant protection agents more efficient. Theoretical methods can help to keep research costs within limits and are therefore being used to an increasingly greater extent. At present, there are two complementary strategies. The one which is based on structure/activity relationships takes an indirect route by comparing known active compounds. For the large number of cases where site and mechanism of action are not known, comparative methods are indispensable. In contrast, the other strategy requires knowledge of the target of action. Active compound and receptor are both included in the model. The computer becomes the locus of an experiment.

From a large number of different applications it has become obvious that both strategies can be used for more than just drug design. In the present review, we therefore try to outline the spectrum of theoretical, molecular research by presenting a representative selection of the large variety of methods used and their possibilities. We regard computer-assisted molecular design as a new *additional* tool for research—combination with experiment still remains an essential precondition for successful molecular design.

2. Interactive Modeling

Computer-assisted molecular design (CAMD) was introduced into chemistry only a relatively short time ago, whereas computer-assisted design in engineering (CAD) has been well established for several years. One of the reasons for this is that interactive work with three-dimensional structures—a typical operation in molecular design—requires highly developed computer graphics systems, which have become available only in the last few years. They generally consist of a minicomputer, a graphics system, and very extensive software. Their development began with the MAC project at the Massachusetts Institute of Technology. Probably the first computer graphics system for molecular modeling was developed in the course of this project. *Cyrus Levinthal*, director of this project, also published the first papers on computer-assisted molecular graphics.[4] Since then a large number of modeling systems have been developed at universities and industrial research institutes,[5–7] but almost exclusively in the USA and Great Britain, only one important contribution coming from Germany.[8] For this reason, the major part of commercially available software now comes from the Anglo-Saxon countries.[9–13] This also applies to the hardware. A good review of the development of molecular graphics is given in Ref. [14]. Molecular graphics systems have now become a useful instrument in molecular research. They make it possible to employ substantially more complicated models than has been possible in the past. At the beginning, computer graphics was predominantly used in pro-

[*] Dr. H. Frühbeis, Dr. R. Klein, Dr. H. Wallmeier
Zentralforschung, Hoechst Aktiengesellschaft
Postfach 800320, D-6230 Frankfurt am Main 80 (FRG)

  0570-0833/87/0505-0403 $ 02.50/0

tein crystallography. As a consequence, the molecular modeling program, which is probably the one most widely used, was developed by a crystallographer.[8] Other areas followed, including drug and plant protection research.

The modeling systems mentioned above have more or less convenient user interfaces in their current versions and differ considerably in the number of features available. Some have been designed primarily for modeling large biopolymers,[11] while others are more suitable for medium-sized organic molecules.[10, 13] However, there are also program systems for handling any type of molecule.[6]

The important features of computer-assisted molecular design, to be found in most of the modeling systems, will be described in the following. To start modeling, chemical structures can be generated either by building them up from atoms and/or molecular fragments or by altering an existing structure, for example one taken from a data base.

Molecules can be displayed e.g. as vectors, which correspond to the bonds (Dreiding models), as overlapping spheres (space-filling CPK models), or as vectors together with a transparent space-filling representation, consisting of a dotted surface or a polygonal network (chicken wire model). The object displaced on the screen can be rotated, zoomed, and clipped (3D clipping). All operations work in real time. Furthermore, a number of techniques are available which give a three-dimensional impression of the object, for example by rotation and wagging, perspective representation, and depth cueing (the intensity of the vectors represented decreases with increasing distance between the object and the observer). The most elegant solution, however, is provided by the liquid crystal 3D viewer.[15] A detailed description of virtually all the techniques available is given in Ref. [16].

To visualize properties of three-dimensional structures, it is often helpful to superimpose them. Thus, for example, the space occupied, or atomic charge patterns can be compared (see Section 4.1). A typical problem in modeling is the analysis of the conformational space of a molecule. All the possible geometrical arrangements for a given structure at a specified level of energy have to be determined. This is performed interactively by connecting the relevant torsion angles to dials and rotating them, and monitoring the energy in real time. Such an analysis can also be carried out systematically by means of special algorithms (see Section 5). Conformation analysis was one of the first applications of computer modeling.[17]

The docking of substrate and inhibitor molecules into enzymes is used to determine whether or not a proposed structure will fit into the active site cavity of a particular enzyme. Some stages of such a docking manoeuvre are shown in Figure 1. One of the first programs for carrying out interactive docking was developed by a crystallography group.[18] Docking programs were generally designed so that enzyme and substrate can be subjected to rotations or translations, both together and separately. Moreover, any desired conformation of the substrate can be generated by rotation through the appropriate torsion angle. In the simplest case, it is possible to check by inspection whether or not an arrangement produced in this way is physically realizable. Overlap of van der Waals radii of the enzyme and inhibitor should be avoided. This can be checked easily with the above mentioned graphics facilities of a modern modeling system. In addition, the energy of the complex under consideration can be calculated. It has to be minimized in the course of the docking manoeuvre.

A more elegant tool for interactive docking is the display of 'hot spots' (Fig. 1b), which indicate sites of strong repulsion between enzyme and substrate. Such sites must be eliminated by means of the usual geometrical operations on the substrate. Interactive docking is very time-consuming and requires a lot of intuition; nowadays it should only be used for relatively rigid inhibitor or substrate molecules. For flexible molecules, automatic docking is the method of choice (see Section 5.5).

Apart from these basic functions of CAMD, the structure data banks at Cambridge (Cambridge Crystallographic Data File) and Brookhaven (Brookhaven Protein Data Bank) play a very important role. They provide a vast number of experimentally determined chemical structures which can be used for modeling directly, or in the form of

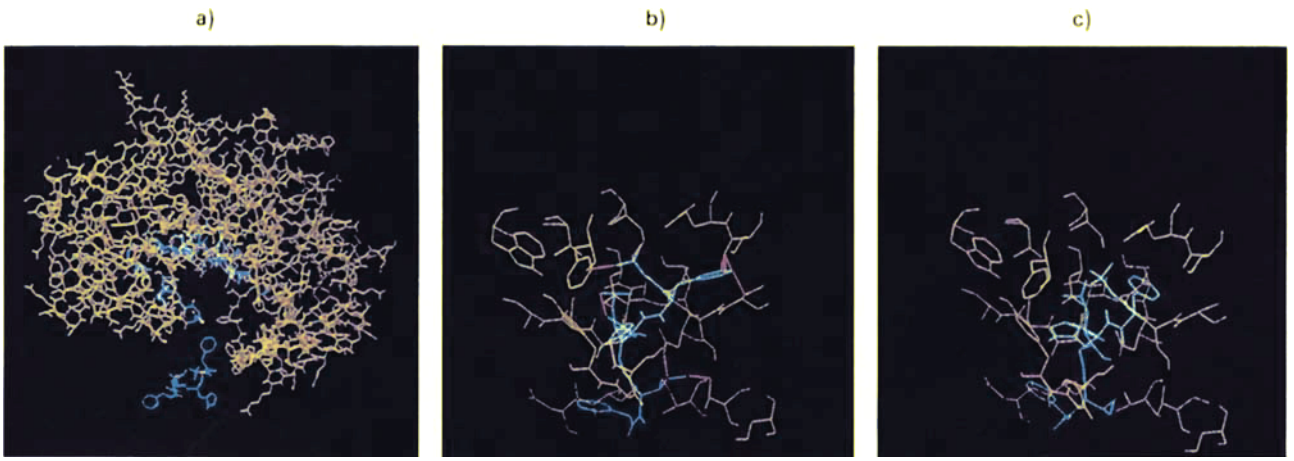

Fig. 1. a) Skeleton model of the enzyme renin (yellow) with a view into the active site (blue) and a model inhibitor (blue) in the lower part of the picture. b) Cavity of the active site (yellow) with poorly docked inhibitor (blue). The red vectors indicate 'hot spots'. The complex has an energy of 14220.53 kcal/mol. c) Optimally docked inhibitor. The energy is markedly less than in b) (180.61 kcal/mol).

templates. It is for this reason that virtually all CAMD program systems include interfaces to these data banks. A description how to exploit the Cambridge Data Bank in a skilful fashion is found in Ref. [19], for example, and in the literature cited therein.

3. Analytical Studies on Models—the Deductive Route

3.1. Basic Principles

To estimate the stability of molecules a number of quite different criteria are considered in chemistry. One may, for example, look at the lifetime of a molecule, i.e. the probability and conditions for decomposition or transformation. The behavior against various reagents is another aspect of stability. Frequently, molecular models are used to interpret the findings in terms of steric hindrance, ring strain, bond strengths, and polarities.[20] For computer models it is very useful that all those aspects of stability can be correlated with the total energies or energy changes of a system. Energy is thus the most important property of a computer model.

Quantum mechanics enables the calculation of the energy of atoms and molecules and therefore forms the basis of most computer models of chemical systems. The characteristic feature of a quantum mechanical model is the wave function, which, as a solution of the Schrödinger equation, contains all the information for a particular state of the system. An important simplification in calculating molecular wave functions is the Born-Oppenheimer approximation.[21] It allows the motions of nuclei and electrons to be considered separately. One obtains the electronic Schrödinger equation,[22] which describes the motion of electrons in the field of stationary nuclei. The electronic wave function is generally represented by orbitals, i.e. solutions for one-electron atoms. In the case of molecules, orbitals are combined according to their symmetry in the molecular orbital (MO) formalism, and according to the bonds between atoms in the valence bond (VB) formalism.[23]

Among the quantum mechanical methods used today, the ab initio methods are those which are closest to solving the electronic Schrödinger equation. The Hartree-Fock self-consistent field (SCF) method[24] is the most important method in practice, particularly in conjunction with basis set expansions, like the Roothaan SCF scheme.[26,27] It represents the lowest level at which the interaction between electrons can be described completely. The computational effort is proportional to the 4th power of the number of electrons, so that meaningful application is restricted to systems of, say 30 second-row atoms, even when supercomputers are used.

Approximations at a somewhat higher level are the method of configuration interaction (CI),[28] the multiconfiguration SCF (MCSCF),[29] and methods based on perturbation theory, such as the Møller-Plesset formalism (MP).[30] Usually they are based on the SCF scheme and provide an improved description of the electron interaction (electron correlation). When used correctly, these methods give substantially more reliable results than simple SCF calculations. Unfortunately, the computational effort for these methods increases with the 5th power of the number of electrons, so that possible applications are even further restricted.

An almost complete parametrization of electron interaction is the key feature of the so-called semiempirical methods.[31] As a consequence, the computational effort is at most proportional to the 3rd power of the number of electrons. The proper choice of parameters is crucial for the usefulness of the results. Hence, these methods can only be used to investigate phenomena, which the parameters chosen are able to describe. A considerable advantage, however, is that one can study quantum effects on systems which are not just trivial fragments of the systems one actually is interested in.

In the Born-Oppenheimer approximation, the energy E of the electronic Schrödinger equation is the potential for the motion of the N nuclei. The gradient of the energy hypersurface (energy as a function of the nuclear coordinates, $E=E(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)$) can be used to determine the forces which act on the nuclei. Minima in the hypersurface correspond to force-free arrangements of the atoms and are referred to as equilibrium geometries or conformations. For determining conformations, it is advantageous if the hypersurface is given as an analytical function (see Section 5.1). This can be achieved, at least locally, by approximation with suitably simple functions. One starts off from a representation in terms of interatomic distances, valence angles, and torsion angles, which correspond to the internal degrees of freedom of molecules.[32] Potential functions with adjustable parameters are used to represent the energy of the internal degrees of freedom.[33-35] In the harmonic approximation, harmonic oscillator potentials are used for bonds and valence angles, while trigonometric functions of proper periodicity are employed for the torsion angles (Fig. 2). In addition, long-range interactions

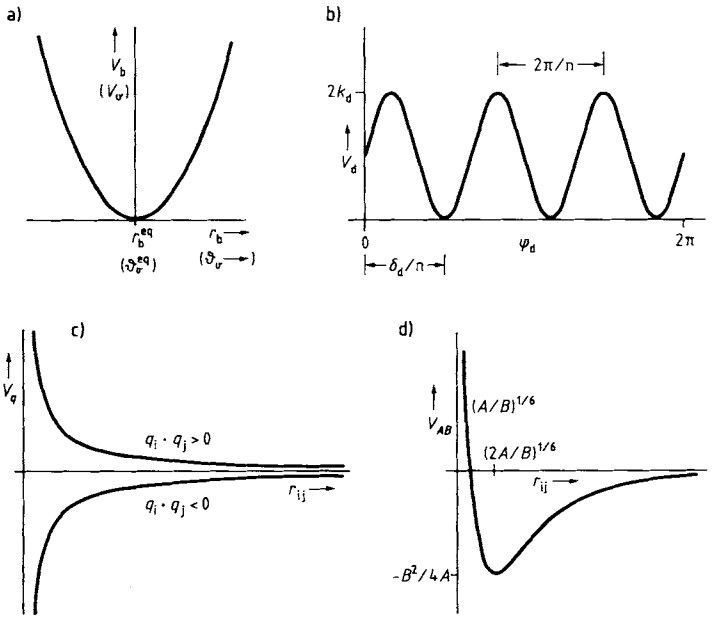

Fig. 2. Potential functions for valence force fields. a) Potential of the harmonic oscillator. b) Potential for torsion angle. c) Potential for Coulomb interaction. d) 12/6 Lennard-Jones Potential for van der Waals interaction.

are represented by Coulomb potentials for the effective charges and by Lennard-Jones potentials for the van der Waals interactions. In the valence force field, all coupling terms between the internal degrees of freedom are neglected, so that

$$
\begin{aligned}
E \approx V = & \sum_{b} k_b (r_b - r_b^{eq})^2 && \text{bond lengths} \\
& + \sum_{\nu} k_\nu (\vartheta_\nu - \vartheta_\nu^{eq})^2 && \text{valence angles} \\
& + \sum_{d} k_d (1 - \cos(n_d \varphi_d - \delta_d)) && \text{torsion angles} \\
& + \sum_{i,j} q_i \cdot q_j / (4\pi\varepsilon r_{ij}) && \text{effective charges} \\
& + \sum_{i,j} A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6} && \text{van der Waals interactions}
\end{aligned} \qquad \text{(a)}
$$

'eq' denotes equilibrium values which fix the position of the particular potential minimum, and ε is the dielectric constant. The parameters k_b, k_ν, k_d, r_b^{eq}, ϑ_ν^{eq}, n_d, δ_d, q_i, A_{ij} and B_{ij} permit fitting to the energy hypersurface. In molecular models they serve to distinguish between different types of atoms and bonds, and thus determine the properties of the model. The term 'force field' usually refers to a set of such parameters in conjunction with the corresponding potential functions. The term is based on the relationship (b).

$$\vec{F} = -\text{grad}(V). \qquad \text{(b)}$$

The type of force field described above (valence force field) has been implemented in various program systems,[33,36,37] which are also used to study biomolecules. It constitutes the simplest form in which the most important interactions can be described and is also suitable as a basis for computer simulation (see Section 5). All contributions to the energy can be readily formulated in computer programs. The computational effort is proportional only to the square of the number of atoms.

Concerning the potential functions, a number of variants exist: for example, bonds are described by some authors in terms of anharmonic potentials;[38,39] linear combinations of trigonometric functions of different periodicities (Fourier expansions)[40] are used for torsion angles; and modified potentials for the van der Waals interaction are employed to describe hydrogen bonds.[37,40]

The extended valence force fields[39,41] contain a few of the coupling terms between the internal degrees of freedom, which is particularly advantageous for strained molecules[42] but makes calculation of the energy and determination of the parameters more complicated (see below). Models of large systems require force fields which are optimized with respect to the balance between computational effort and 'physical content'. This is most readily achieved with the valence force fields.

A common drawback of force fields is the treatment of Coulomb interaction in terms of effective charges. Due to the polarizability of the atoms, the charges have a mutual effect on one another and change with the conformation of the molecule. Distance-dependent as well as scaled dielectric constants have been used as a remedy.[40] The situation is critical because Coulomb interaction often accounts for the major part of the total energy of a conformation.

Given a set of potential functions, the number of force field parameters is determined by the number of atom types g to be distinguished. For example, in a valence force field of the form (a), $g(g+1)/2$ equilibrium distances and force constants are required for the bonds. For the valence angles, as many as $g^2(g+1)/2$ equilibrium angles and force constants are needed, and in principle $(g(g+1)/2)^2$ force constants, phase angles and periodicity factors are necessary to distinguish between all possible types of torsion angles. The nonbonding interactions can be described by g effective atomic charges and $g(g+1)/2$ Lennard-Jones coefficients. In order to keep their number as small as possible, atom types and parameters should be chosen such that there is maximum transferability. However, this imparts an artificial, internal symmetry to the force field, which becomes a property of the model, as well.

As indicated above, force field parameters can be determined by fitting the potential functions to quantum mechanical energy hypersurfaces. Furthermore, quantum mechanical calculations are the most important source of the effective charges of atoms (see Section 3.2).

In general, force field parameters are determined from experimental data.[35,42,43] A very widely used method is the so-called consistent force field procedure,[33] in which force field parameters are varied iteratively until given experimental data are reproduced. The most important sources of data are IR and Raman spectroscopy for force constants, and X-ray crystallography and microwave spectroscopy for equilibrium geometries. The parameters of the van der Waals interaction are frequently derived from thermodynamical data, such as heats of fusion and vaporization, as well as volume/temperature diagrams.

3.2. Calculation of Molecular Properties

Quantum mechanical models are necessary for the investigation of electronic properties. A population analysis, i.e. the summation of the moduli of orbital, atomic, or bond contributions to the wave function, permits the electron and charge distribution to be characterized in terms of effective charges and hybridizations.[44] The moments of the charge distribution (dipole moment, quadrupole moment, etc.)[26] and, using test charges, the entire electric field of the system can be determined.

Preferred positions for electrophilic or nucleophilic attack can be identified as regions of enhanced or reduced electron density.[45] The corresponding structuring of the electric field is very important for the initial phase (molecular recognition) of reactions.[3]

Optical properties of molecules are generally related to transitions between different states.[46] The chrominance, i.e. the ability to absorb light of certain wavelengths, is attributable to the energy difference between two states having a sufficiently high transition probability, the latter being a measure of intensity.[46]

Energy differences, regardless of whether they relate to conformational changes, rearrangements or reactions, can often be read off directly from the energy hypersurface as level differences. In some cases, however, structuring of the hypersurface is an indication of nonadiabatic phenomena, for which the Born-Oppenheimer approximation no longer holds.[47] This applies very frequently to transition states of reactions, if there is a competition between homolytic and ionic dissociation of a bond. In such critical cases, it is appropriate to use the more elaborate methods, such as CI or MCSCF, for determining both energy barriers and geometry.

Like the electric field, the geometry of a molecule is one of the important factors determining affinity and activity of molecules. Optimizing the geometry is equivalent to locating the deepest minimum of the energy hypersurface. If the initial geometry is not too far away from the minimum, it is sufficient to relax the structure, i.e. to vary the coordinates of the atoms so that the energy cannot be lowered further. Several methods[48,49] are available for carrying out optimizations of this type, which will not be discussed here in detail.

One aspect of the geometry is spacefilling, which brings into play the concept of molecular surface. From the quantum mechanical point of view, the surface of a molecule is a fractal, i.e. a geometric object with a dimensionality between two and three,[50] which can neither be readily represented nor easily handled. However, the concept of fractals makes it possible to quantify roughness and structuring, which is certain to become important with regard to the understanding of processes on solid surfaces and membranes.

A somewhat more pragmatic concept of spacefilling and surface can be derived from the van der Waals radii of the atoms. A molecule is taken as an arrangement of overlapping spheres (CPK spacefilling model). If the overlap of the spheres is taken into account, the volume and surface area of a molecule can be calculated[51-54] (also see Section 2).

Spacefilling by molecules is generally used to assess steric similarity of molecules (see Section 4). The surface area of a molecule provides information about the behavior towards solvents.[54,55] In addition to the effective charges of the atoms, the curvature of the surface is regarded as a measure for the ease of solvation, i.e. the accessibility of the particular atoms to solvent molecules. For this purpose, an imaginary sphere is allowed to roll on the surface of the spacefilling model, and the appropriate parameters are read from its path.

4. Comparative Methods

With the aid of so-called comparative methods, an attempt is made to extrapolate experimental data, i.e. to transfer the data which are available for a number of compounds to other molecules taking into account structural similarities and differences. In general, these data are reactivities or biological activities which, if they could be obtained prior to synthesis, would help to select the most promising compounds. A study of the relationships of structure and reactivity/activity is based on the experience that molecules which have similar structure also exhibit similar chemical and biological behavior.

Unlike the approaches described in Section 5 the comparative methods do not necessarily require experimental information about the site where the molecule finally acts or reacts. Thus, e.g. quantitative structure-activity relationships can be set up for a number of enzyme inhibitors without knowing anything about the enzyme itself.

4.1. Superposition Methods

In medicinal chemistry, activity of a low molecular weight substance is often equivalent to a specific receptor interaction. This correlation is attributed to a molecular recognition mechanism.[56] Consequently, all compounds which form a stable complex with the same receptor should exhibit a common structure-dependent feature which is referred to as a pharmacophore. In this context, a pharmacophore is not restricted to a certain substitution pattern as is generally done, since it has been found that a recognition pattern can, for example, also be realized by the charge distribution in a molecule.[57] The comparison of molecules by superposition with the aid of one of the methods discussed below is used to set up, confirm or discard a pharmacophore model for a particular biological action and to design new molecular structures which also carry the pharmacophoric pattern. A three-dimensional superposition is necessary because not only the presence of obvious common features like functional groups are responsible for the recognition by the receptor but the three-dimensional orientation of the pharmacophore as well.

In the simplest case of a structural comparison, the centers of functional groups in a set of structurally analogous molecules are considered. Using a least squares fit procedure, the sum of the distances between corresponding centers is minimized by displacing and rotating the rigid molecular skeletons. However, in general the initial conformations of the molecules under investigation do not correspond to those that allow a truly optimal fit. Therefore, it may also be necessary to perform rotations about single bonds or even to allow all atoms to move freely in space. In order to exclude unreasonable conformations or geometries the energy required for the geometry change or another suitable parameter (e.g. bump check) is calculated simultaneously. The remaining differences—possibly together with the energy required for the fit—provide qualitative information about the closeness of the fit of the basic pharmacophore model and the structural properties of the molecules investigated.

The van der Waals volumes too are suitable for testing steric correspondence.[51,52,58] For example, the degrees of freedom are varied so as to achieve a minimum of the united volume of all molecules. A pharmacophore model restricted to these space-filling properties is shown in Figure 3. The fungicides shown there inhibit the enzyme cytochrome P450 monooxidase which is vital to fungi.

The discovery of suitable evaluation functions for the matching of charge patterns by varying the degrees of freedom of the molecules is not a completely trivial matter.

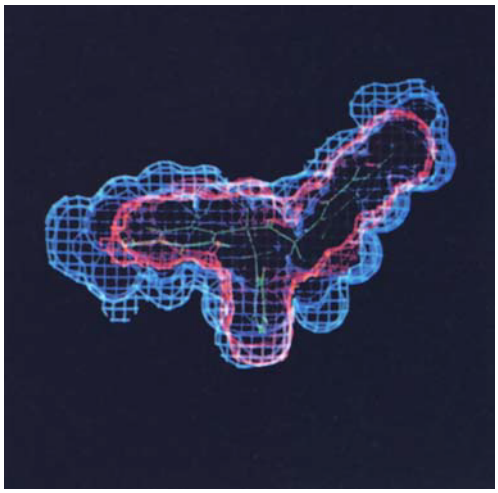

Fig. 3. Chickenwire representation of the van der Waals volumes of the natural substrate (blue) of cytochrome P450 and miconazole (red). The molecular skeleton of miconazole is shown in green.

Namasivayam and *Dean*[59] have described a statistical method for matching charge patterns on molecular surfaces. A least squares fit procedure can also be used to fit the electrostatic potentials of several molecules on an arbitrary surface which is resolved into grid points.[60] In Figure 4, the potential differences are projected on a section

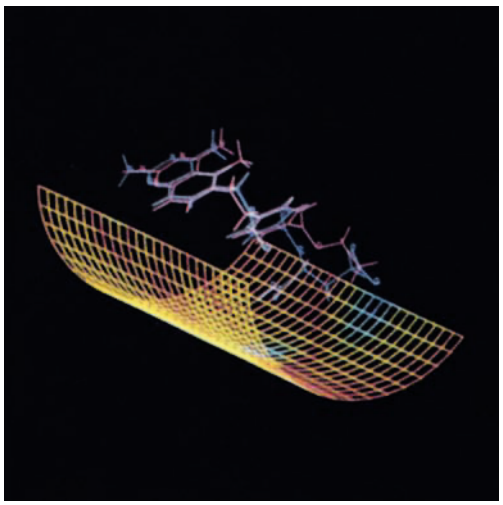

Fig. 4. Projection of the electrostatic potential difference of two quinazoline derivatives (dihydrofolate reductase (DHFR) inhibitors) onto a surface (red: small differences, blue: large differences).

of the curved surface of a cylinder which surrounds the molecules to be compared. By appropriately coloring the vectors between the grid points the differences in the charge patterns before and after the fit can conveniently be analyzed visually. *Broto* and *Moreau*[61] use a so-called intercorrelation function to superimpose structures with respect to the effective atomic charges (as well as other atomic properties). Here, they translated their concept of the autocorrelation function[62] for describing the "property distribution" within a molecule into the problem of making similarities in this distribution quantifiable. *G. Naray-Szabo*[63] uses an electrostatic lock-and-key model to investigate common features of trypsin inhibitors.

A very elegant method which makes it possible to confirm or reject a pharmacophore model is called the active analog approach developed by *Marshall* et al.[64] A pharmacophore hypothesis is first proposed by defining several functional groups or hetero atoms to be essential for activity. A systematic conformational search (torsion or rotatable bonds) in all molecules is supposed to give that three-dimensional arrangement of the pharmacophore which at least can be realized by all active compounds. If there is no such arrangement the initial hypothesis is rejected. If the search has been successful the three-dimensional pharmacophore found this way serves as an initial criterion for the classification of novel compounds. In a second step the volume provided by the receptor—the so-called excluded volume—is determined as the united volume of all active molecules in their active conformations. In spite of having the correctly oriented pharmacophore, inactive molecules should not be capable of being fitted into this volume. Otherwise, there will once again be a reason for doubting the correctness of the proposed pharmacophore.

The so-called distance geometry method developed by *Crippen*[65] for describing molecular geometry can also be applied to the problem of extracting common steric features from a set of molecules.[66,67] The flexibility of a molecule is represented by a minimum and a maximum distance matrix, i.e. by a range of validity of the interactomic distances. For the distance between two atoms (or representative points) of the pharmacophore, the only values allowed are those which lie within the range of validity for all active molecules. This considerably reduces the number of conformations which can be calculated from the distance matrices.

Frequently, the structure of the transition state of the natural substrate rather than a number of active compounds is used as a template for the design of novel drugs (transition state analogs)[68] (also see Section 3.2). Since the properties of transition states are in general difficult to determine experimentally, as a rule, extensive quantum-mechanical calculations (ab initio, MNDO, MINDO etc., also see Section 3.2) of the reaction path from the substrate to the product have to be carried out.

A pharmacophore model which has been derived by one or more of the methods described above provides, with certain restrictions, information about the structure and properties of the receptor. In the literature, the procedure for obtaining such a qualitative image of the receptor is often referred to as receptor (site) mapping.[58,67,69-71] Thereby, it is possible to design novel drugs which do not necessarily exhibit the same structural features as the active molecules compared or the lead compound, and yet to estimate the activity of the novel compound, even if purely qualitatively.

4.2. Qualitative Structure-Activity Relationships (SAR)

The methods described so far are suitable only for a relatively small number of compounds to be compared. Furthermore, they are based on the assumption that there is

essentially only one mechanism responsible for the biological activity, namely the formation of a receptor-ligand complex. The use of so-called pattern recognition (PR) methods[72] does not depend on any of these conditions. PR is based on structure-dependent molecular properties, so-called descriptors (e.g. number of oxygen atoms in the molecule, molecular diameter, etc.). Depending on the individual values of the descriptors each molecule has a particular descriptor pattern. Using artificial intelligence methods and mathematical procedures such as cluster analysis, principal component analysis and discriminant analysis these patterns serve to assign a compound to activity classes. PR is used in particular where a large number of structural inhomogeneous compounds is encountered, for example in studies on carcinogenicity and genotoxicity.[73]

In the approach of *Broto* et al.[62] a so-called autocorrelation vector is calculated for each molecule from atom-related parameters or properties (e.g. heteroatom/non-heteroatom, atomic log P increment etc.). These vectors can be considered to represent the property distribution within the molecules and are used to classify the compounds according to their activities.

The so-called CASE algorithm[74] (computer automated structure evaluation) breaks up the molecules under investigation into fragments and searches for those substructures which contribute to activity or those which lead to inactivity. Thus, this method immediately provides a pharmacophore model. A similar approach, which also permits the derivation of quantitative structure-activity relationships, has been reported by *Streich* and *Franke*.[75]

4.3. Quantitative Structure-Activity Relationships (QSAR)

In a study of quantitative structure-activity relationships an attempt is made to describe the activity or reactivity within a set of compounds by means of a mathematical formalism which incorporates structure-dependent parameters (descriptors).[76] Although the classical QSAR approaches were introduced purely empirically they can be derived in terms of an extrathermodynamic approximation[77] (additivity of substituent effects, separability of different effects etc.).

The first quantitative structure-activity relationship was developed as early as 1937 by *Hammett*.[78] He correlated the hydrolysis rate of *meta*-substituted and *para*-substituted benzoates with the so-called Hammett σ which is calculated from the dissociation constants of the corresponding benzoic acids. In order to obtain a relationship which could be applied to *ortho*-substitutions as well, *Taft*[79] added a steric parameter E_s to the Hammett equation.

In addition to electronic substituent effects, which were to be described by the Hammett σ, hydrophilic/hydrophobic properties were also considered responsible for biological activity. Therefore, *Hansch* et al.[80] introduced a further parameter, π, [eg. (c)] which is based on studies by *Meyer* et al.[81] and *Overton*[82] and is calculated from the octanol-water partition coefficient (log P value).

$$\log(1/C) = a \cdot \sigma + b \cdot \pi + c \qquad (c)$$

C is the (theoretical) activity; the coefficients a, b and c are determined by a least squares procedure so that equation (c) optimally reproduces the measured biological data of a selected set of molecules (regression analysis). Many systems, however, cannot adequately be described by this approach. *Hansch*[83] attributed these deficiencies to transport mechanisms which also manifested themselves in the in vivo activity data and therefore assumed a parabolic dependence of the activity on π [eq. (d)].

$$\log(1/C) = a \cdot \sigma + b \cdot \pi^2 + c \cdot \pi + d \qquad (d)$$

Probably the most frequently used form of the classical QSAR approach, based on a linear combination of descriptors, also takes into account steric effects with the aid of *Taft*'s E_s [eq. (e)].

$$\log(1/C) = a \cdot \sigma + b \cdot \pi^2 + c \cdot \pi + d \cdot E_s + e \qquad (e)$$

Extensive efforts are still being made to improve the substituent parameters, replace them by others or add new ones. *Hammett*[84] himself differentiated between σ_m for *meta*-substituents and σ_p for *para*-substituents. The knowledge that electronic effects comprise an inductive part and a resonant part led to a splitting of σ into σ_i and σ_r.[85]

In addition to the Hansch π, other substituent parameters which can be derived from measurable molecular properties, have been proposed for modeling hydrophobic interactions: the parachor which is related to the surface tension,[86] the Hildebrand-Scott solubility factor[87] and parameters obtained from chromatography.[88] An empirical hydrophobicity parameter can be calculated by the HIBIS method.[89] The classical QSAR approach employs only scalar descriptors. However, steric properties of substituents are based on three-dimensional orientation. It is therefore difficult to model steric effects by means of parameters such as *Taft*'s E_s or *Charton*'s U,[90] which is calculated from van der Waals radii. By separating the steric descriptor into components *Verloop* et al.[91,92] was nevertheless successful in introducing relative orientation into the so-called STERIMOL parameter. Further steric parameters, all of which are based on the superposition of molecular volumes, are the MSD,[93] MTD,[94] SIBIS,[70,95] and 3D-MSD[52,96] descriptors.

The typical PR descriptors and the components of autocorrelation vectors in *Broto*'s approach[97] are QSAR descriptors of a very special type. There have also been many attempts to correlate the biological activity with other computable parameters which describe the molecule or a substituent effect. For example, parameters derived from semi-empirical quantum mechanical calculations[98] and electron densities from Hückel and CNDO/2 calculations[99] have been used as QSAR descriptors.

Apart from (multiple) regression analysis[100] as the standard method, principal component analysis,[102] discriminant analysis[103] and cluster analysis[104] are also used for the statistical exploitation[101] of a Hansch or a similar QSAR approach. Several statistical parameters, such as the RMS deviation and the correlation coefficient, provide means to check the reliability of a regression analysis.[100]

The significance of a descriptor can be determined approximately by performing two QSAR procedures with and without the descriptor and by comparing the respective correlation coefficients. However, even satisfactory statistics cannot exclude random correlations.[105] The probability of chance correlations clearly grows when the number of descriptors becomes large compared with the number of structures used.[106]

In the QSAR approach by *Free* and *Wilson*,[107] structure-activity relationships are set up without physicochemical parameters. It is assumed that the substituents contribute additively to the activity. The activity value of a compound is taken to be the sum of a basic activity and the contributions of the substituents. The contribution of a substituent depends both on its type and its position in the molecule. If the equation is set up for all tested compounds a system of linear equations is obtained. By solving the equations using a least squares fit procedure the individual substituent contributions can be determined. The effect of substituents other than those present in the compounds tested cannot be calculated; only the effect of different combinations of substituents can be estimated by this approach.

It is possible to combine the Hansch and Free-Wilson methods.[76] In general however, they are used exclusively, since the application of the Hansch method is limited by the availability of physicochemical parameters whereas the Free-Wilson method needs large variations in the substitution pattern. Thus, the two approaches complement each other.

An approach similar to that of the Free-Wilson method has been described by *Crippen*,[108] who used the distance geometry algorithm to set up quantitative structure-activity relationships. This method too is based on substituent constants, which additively contribute to the activity and depend on the position within the receptor binding site. In general, receptor binding energies are used to quantify the activities. In contrast to the classical Free-Wilson method several (generally two) three-dimensional orientations of the pharmacophore are determined. The compounds investigated are assumed to bind to the receptor with different pharmacophore orientations. By taking these different "binding modes" into account the poor correlation between substituent contributions and activity when using only one orientation should be improved.

Some recent publications have been concerned with the combination of QSAR and superposition or receptor mapping methods in order to obtain a model of the receptor which also provides quantitative information.[92, 109, 110] A procedure,[111] which allows generation of a receptor model by means of pseudo atoms, will be described in Section 5.6. The method predicts activities without being based upon an extrathermodynamic approximation as are the classical QSAR approaches.

5. Inductive Modeling: Computer Simulation

The optimization of molecular geometries, as discussed in Section 3, generally gives only information about a single conformation. For molecules with many internal degrees of freedom, however, there may be a large number of minima in the energy hypersurface. Therefore, the most stable conformation can only be identified if all the minima are known. There are several methods of searching systematically for the most stable conformation. Unfortunately, these methods reach their limits very quickly. A molecule with only 50 atoms, for example an open-chain hydrocarbon with the formula $C_{16}H_{34}$, can already have several million local minima. Such a molecule does not remain in a single conformation but fluctuates between several conformations. In view of this, it seems more reasonable to depart from the concept of the most stable conformation and instead to model the dynamic properties of the molecule. All methods for this purpose, which are presented here, are based on force field energies.

5.1. Conformation Search

The systematic scanning of a conformation space in order to find the conformation with the lowest energy (global minimum of the hypersurface) is referred to as conformation search. Usually, advantage is taken of the fact that conformations can virtually always be distinguished by torsion angles. In principle, therefore, all conformations can be found by varying the torsion angles in all possible combinations. One selects a sufficiently large number of initial geometries[112] and relaxes them (see Section 3.2). The conformations thus determined are of course 'close' to the initial geometries. To become more independent of the critical choice of initial geometries, a grid can be placed over the space of torsion angles (grid search).[113]

The energy calculations necessary for the conformation search can be carried out using quantum mechanical methods, whenever high accuracy is required.[114] In general, however, force fields are used. The complexity of this type of conformation search increases exponentially with the number of bonds, so that it is reserved for small molecules.

In the case of large molecules, the search has to be simplified by means of additional restrictions. Frequently, it is possible to reduce the number of variable torsion angles by making use of local symmetry.[113, 115] The most favorable conformations of small subunits are determined and included as fixed blocks in the total molecule, or are used to define grid points for the global search. An important example is the prediction of secondary structure elements in polypeptides and proteins.[116] Introducing pleated sheets, turns, and helices constitutes a preselection of possible values for the main chain torsion angles.[117]

5.2. Molecular Dynamics

The huge number of internal degrees of freedom of biomolecules and polymers, and the coupling of degrees of freedom in cyclic molecules, suggest that use should be made of the dynamic properties of models in the conformation search. Conformation search is performed as a computer experiment in which the behavior of the model is observed, when subjected to external perturbations. This corresponds to the idea that a real molecule in thermal equilibrium finds its most favorable conformation in a se-

quence of reversible geometry changes. To transfer this to a mechanical model, one has to observe the time evolution of geometry ($\vec{x}=\vec{x}(t)$, trajectory). From the expansions for $t+\mathrm{d}t$ and $t-\mathrm{d}t$ one obtains equation (f).[118,119]

$$\vec{x}(t+\mathrm{d}t)=2\vec{x}(t)-\vec{x}(t-\mathrm{d}t)+\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2}(\mathrm{d}t)^2+0((\mathrm{d}t)^4) \qquad \text{(f)}$$

According to *Newton*, for each atom i equation (g) holds and the force field [see Section 3.1, eq. (b)] provides the relationship (h).

$$\frac{\mathrm{d}^2\vec{x}_i}{\mathrm{d}t^2}=\vec{a}_i=\frac{\vec{F}_i}{m_i} \qquad \text{(g)}$$

$$\vec{F}_i=-\mathrm{grad}_i(V) \qquad \text{(h)}$$

Newtons equations of motion are solved by discrete integration over a very large number of finite time steps ($\mathrm{d}t\rightarrow\Delta t$). The size of the first change in geometry (perturbation of the system) defines the level of kinetic energy, which, according to statistical mechanics, corresponds to a temperature T [eq. (i)].

$$\frac{3}{2}NkT=\frac{1}{2}\sum_i^N m_i v_i^2. \qquad \text{(i)}$$

The time evolution can also be formulated in terms of the velocities of the atoms [eq. (j)].

$$\vec{v}(t+\mathrm{d}t/2)=\vec{v}(t-\mathrm{d}t/2)+\frac{\mathrm{d}^2\vec{x}(t)}{\mathrm{d}t^2}(\mathrm{d}t)^2+0(\mathrm{d}t)^4 \qquad \text{(j)}$$

whereby the model temperature can be kept constant by scaling the velocities. This corresponds to coupling the system to a heat bath.[119,120]

Molecular dynamics (MD) simulations can be carried out either with individual molecules or with entire ensembles of molecules. The molecules are arranged with the desired density in a cell of appropriate volume. In the case of liquids, a cube is generally used; for crystal structures, the form used corresponds to that of the unit cell. By means of so-called periodic boundary conditions, surface artefacts are eliminated and a quasi-infinite number of particles is generated by surrounding the cell with its own images in all directions of space[121] (Fig. 5). By scaling the cell volume, the pressure can be kept constant.[119,120]

The optimum time step Δt for the simulation of molecules with many internal degrees of freedom is 10^{-15} to 10^{-14} seconds. This roughly corresponds to the time scale of molecular vibrations. With supercomputers, it is possible to carry out simulations over 10^6 time steps[122] so that processes such as internal rotations and aggregations can be studied. The dynamics of molecules containing 500 atoms can be investigated very thoroughly, even including a few hundred solvent molecules. Rapid, periodic movements and slow, nonperiodic movements can be read off from the trajectory of the system. The latter are particularly interesting since they lead to conformational changes. In this way, the stability of a hypothetical conformation of a large molecule can be assessed directly. Transitions of the system to other conformations can be accelerated by means of high temperatures. However, the degree of aggregation may change, so that a completely different area of conformation space is entered. It is therefore possible to study phase transitions.

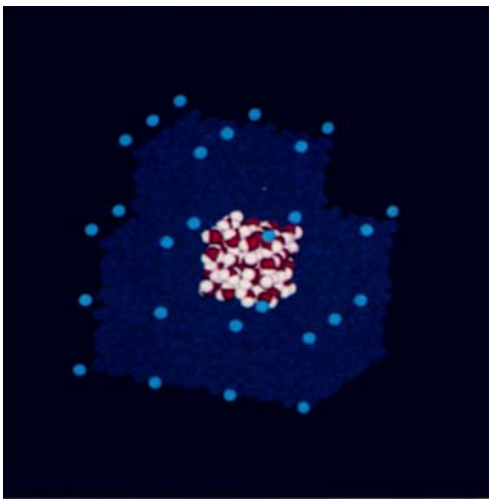

Fig. 5. Periodic boundary conditions for MD and Monte Carlo (MC) simulations (see text).

Correlation and averages play an important role in the evaluation of MD simulations. One considers correlations in time and in space which are represented by so-called correlation functions.[123,124] The different types of motions[125] (periodic, nonperiodic) and in particular coupling between different parts of the molecule can be read off from the time correlation functions. In the case of time-independent correlations, time averages are considered. Preferred local and global conformations can be identified, thereby. This procedure is often used to analyze the structure of solvation shells in the simulation of dissolved molecules to estimate hydrophilic or hydrophobic behavior.[126]

The above statements on the possibilities of MD simulations apply in principle to all types of systems, provided the necessary force field parameters are available. Most applications are found for fluid systems,[127] but solids and surfaces[128] as well as polymers[129] have also been simulated. Biomolecules such as nucleic acids,[130] peptides,[125,131,132] poteins[133] and membranes[134] have been studied. However, all types of systems are subject to the restriction that, with the computers available today,[122,135] it is impossible to consider any process which takes substantially longer than 10^{-9} seconds. Thus, for example, it is impossible to model the folding of proteins solely by means of MD simulation. The same applies to most biological transport processes. Nonetheless, the transport of ions through membrane channels [136] just constitutes the limit of meaningful investigations.

5.3. Monte Carlo Methods

Another type of computer experiments is based on statistical variations of the model. The Monte Carlo proce-

dure is a method for calculating properties of fluctuating systems as so-called ensemble averages.[123, 137] Integrals of the form (k) are made discrete in accordance with eq. (l).

$$\langle A \rangle = \frac{\int f(\vec{x}) \cdot A(\vec{x}) \cdot d\vec{x}_1 d\vec{x}_2 \ldots d\vec{x}_N}{\int f(\vec{x}) \cdot d\vec{x}_1 d\vec{x}_2 \ldots d\vec{x}_N} \quad (k)$$

$$\langle A \rangle \approx \frac{\sum_k^M f(\vec{x}^{(k)}) \cdot A(\vec{x}^{(k)})}{\sum_k^M f(\vec{x}^{(k)})} \quad (l)$$

Random numbers are used to generate M local variations of the system (displacement of an atom, change in a torsion angle), which lead to new states. If the energy of a new state is lower than that of the previous one, the new state is accepted, i.e. the next variation will start from this new state. On the other hand, if the energy is higher, the Boltzmann factor f [eq. (m)] is calculated, and compared

$$f = \exp(-E^{new} - E^{old})/kT \quad (m)$$

with another random number. If the random number is smaller than f, the new state is after all accepted; otherwise, the previous state is retained. f is thus the transition probability for a state of higher energy, and depends only on the directly preceding state. This type of random walk through the conformation space of the system is called a Markov chain.[138] In this way, the system is allowed to fluctuate (Fig. 6) according to a Boltzmann distribution for

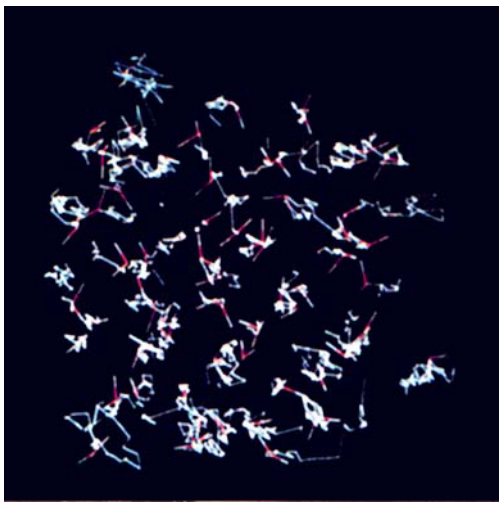

Fig. 6. Brownian motion of H_2O molecules simulated by a MC method.

temperature T. Using the Boltzmann factor as a transition probability is the essential feature of the metropolis sampling method.[139] As in molecular dynamics, it is customary to provide periodic boundary conditions for many-particle systems (see Fig. 5).

Calculation of averages is of course affected by statistical errors, showing up in the root mean square deviation of the values [eq. (n)]. Since s(A) is proportional to $1/\sqrt{M}$, the number of sample points M must be increased by a factor of a hundred to obtain a result which is more precise by one decimal place. In practice, M is of the order of magnitude of 10^6.

$$s(A) = \sqrt{\sum_k^M (A^{(k)} - \bar{A})^2 / M} \quad (n)$$

The Monte Carlo method (MC) too, has so far mainly been used for studying the properties of liquids.[137] The solvation of biomolecules has also been investigated by the MC method.[140] In addition, there are a number of studies on polymers which are primarily concerned with determining conformations.[141] Because of the statistical choice of variations, the MC method is very suitable for conformation search. A very important aspect of MC simulations is the calculation of thermodynamic entities, especially the entropy, which will be discussed in the next section.

5.4. Calculation of Entropy Contributions

The energy balance of a process is not determined solely by the differences in internal energies or enthalpies of the subsystems involved. The 'usable' free energy ΔG [eq. (o)]

$$\Delta G = \Delta H - T\Delta S \quad (o)$$

also has a contribution from the change in entropy, which is temperature-dependent. This may be taken as an indication that the dynamics of molecules play a role here. According to the third law of thermodynamics entropy is related to the probability of a state or a conformation of a system [eq. (p)]. This forms the basis for the calculation of entropy by means of computer models.

$$S = k \cdot \ln(W) \quad (p)$$

There are various contributions to the entropy. The external ones arise from the translational motion of particles, the rotational motion of particles, and the number of particles (solvation, aggregation), while the internal contributions are due to molecular vitrations and internal rotations.

The individual contributions can be calculated from the partition functions Z_{trans}, Z_{vib}, Z_{rot} of statistical mechanics [eq. (q)].[142]

$$S = Nk\left[T\frac{\partial \ln(Z)}{\partial T} + \ln(Z)\right] \quad (q)$$

Thus, in the approximation of the harmonic oscillator and the rigid rotator even quantum effects are taken into account. The contributions from the number of particles primarily manifest themselves as 'stoichiometric' factors but also have an indirect effect on the other contributions. They represent, for example, the main contribution to the entropic part of the hydrophobic interaction.[143]

For very large systems, the direct calculation of entropy via partition functions is too inexact owing to the approxi-

mations involved. Instead, MC[144] or MD simulations[145] can be used to calculate the entropy from the fluctuations of the internal coordinates. This is equivalent to determining the probability W. Unfortunately, only periodic motions can be taken into account. However, this is a viable method for determining the entropy contributions to the energy differences between different conformations of a molecule.

5.5. Docking

The interaction between a receptor molecule (R) and possible ligands decides whether a particular type of ligand is preferred, and can thus repress others. An example of this is the competitive inhibition of enzymes, the natural substrate (S) being displaced by an inhibitor (I). If entropy contributions are neglected, the relevant energy balances (r) apply.

$$\Delta E(\mathrm{S}) = E(\mathrm{RS}) - E(\mathrm{R}) - E(\mathrm{S})$$
$$\Delta E(\mathrm{I}) = E(\mathrm{RI}) - E(\mathrm{R}) - E(\mathrm{I}) \qquad \text{(r)}$$

For a complex to be formed at all, ΔE must be negative. $|\Delta E(\mathrm{I})| > |\Delta E(\mathrm{S})|$ is then a necessary condition for I to have an inhibitory effect. In order to calculate ΔE, one has to know the optimum conformations of the R, S, and I molecules, and the conformations of the complexes RS and RI. For the free R molecule, the X-ray structure is generally used, while the methods described in Sections 3.1 and 5.1. can be employed for substrate and inhibitor molecules. Determining the conformations of the complexes, however, is a special problem since two molecules are involved. The search for the conformations of the complexes, generally referred to as docking, has three aspects:[146]

a) relative orientation of S/I with respect to R
b) optimum conformation of S/I in contact with R
c) optimum conformation of R in contact with S/I

In the docking of rigid ligands with rigid receptors, only (a) is important. In such cases, interactive docking (see Section 2) is the method of choice. However, if the ligands are not rigid, the problem of multiple minima is encountered immediately (b). Interactive docking can very easily give misleading results if the global minimum is not found. In such cases, a special form of MC simulation has proved useful.[147] This method uses an evolution algorithm[49] to search for several conformations simultaneously, and optimizes and accommodates them. Conformation changes in the receptor (c) can be taken into account by subsequent optimization of the entire receptor-ligand complex by MD or MC simulation, provided the changes in geometry of the receptor are small. At the same time, entropy contributions to complex formation can be determined.

5.6. Receptor Modeling

As already mentioned in Section 4 the comparative methods may also be used to produce receptor models.[58,67,69-71] Only when combined with classical QSAR methods (Free-Wilson or Hansch) do these models also provide quantitative information about the activity of analogs.[92,109,110]

The method described below makes direct use of the experimental activity data of a number of structurally analogous molecules in order to determine a receptor model[111]. But unlike classical QSAR methods it is not based on physicochemical parameters *(Hansch)* or empirically determined substituent contributions *(Free-Wilson)*. The method gives the active conformations of the active molecules and a model receptor, which makes it possible to calculate activities of untested compounds. Thus, receptor modeling provides information of both an active analog approach (see Section 4.1) and a QSAR procedure, but still is not a hybrid of the two methods. The essential feature is the calculation of the binding energy between each molecule investigated and a pseudomolecule, the model receptor. This pseudo-binding energy comprises the electrostatic and van der Waals interaction between a molecule and the model receptor and the conformational energy of the molecule. The energy terms used correspond exactly to the force field contributions described in Section 3 [eq. (a)]. The receptor consists of pseudoatoms which differ from "normal" atoms in that they are capable of varying their van der Waals radii and their effective charges during the calculation.

By an iterative procedure, the geometry of the molecules and of the model receptor and the effective charges and the van der Waals radii of the pseudoatoms are varied so as to achieve an optimal fit of the pseudo-binding energies and the experimentally determined enthalpies (calculated from the activity data). A further requirement is that the molecule/model receptor complexes are stable, i.e. are located in an energy minimum—as far as possible a global one. The method makes use of two approximations: a) the receptor is identical for all molecules and b) entropy and solvation effects are not taken into account explicitly. Figure 7 shows a model receptor which was calculated from six quinazoline derivatives which inhibit dihydrofolate reductase (DHFR) (also see Section 6.1).

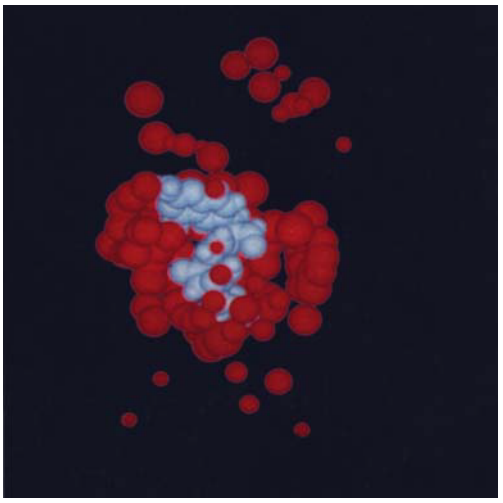

Fig. 7. Space-filling CPK representation of a model receptor (red) for DHFR inhibitors with a docked quinazoline derivative (blue).

In order to check the model receptor, further experimentally investigated molecules (so-called test molecules) are docked into the pseudomolecule. Comparison of the resulting pseudo-binding energies with the experimental binding energies of the test molecules gives information about the predictive power of the model receptor.

Novel untested compounds can now be docked into the model receptor. The calculated pseudo-binding energy allows the activity to be predicted, the reliability of which clearly depends on the predictive power of the model as determined by the test substances.

6. Applications

CAMD is already very widely used today. It will be quite impossible to list all areas of application here. We have therefore chosen three examples, which will illustrate the function and importance of this new instrument. So to speak for historic reasons, our first example is DHFR, which is both of great pharmacological interest and probably one of the enzymes which has been investigated the longest and the most thoroughly by computer-assisted methods. The same applies to renin. In the section on modeling for enzyme engineering, it is intended to demonstrate that this promising area is inconceivable without computer graphics.

6.1. Studies on Dihydrofolate Reductase

Interest in the structure and function of dihydrofolate reductase (DHFR) has grown very rapidly over the past few years and, consequently, so has the volume of relevant literature.

DHFR catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate and thus plays a key role in the metabolism of a large number of organisms. It is the target enzyme of the cytostatic methotrexate and the bactericide trimethoprim. Since the three-dimensional structures of DHFR from chicken liver,[148] *Lactobacillus casei*[149] and *Escherichia coli*[149] and the structure of a complex of methotrexate, cofactor and enzyme were available at about the same time as powerful molecular graphics, dihydrofolate reductase became a preferred research object for molecular design. A number of groups in industrial research[150] worked very intensively on the design of improved, more specifically acting analogs of the bactericide trimethoprim and on the development of new inhibitors with antimalarial activity.[151] At the universities, *Hansch* et al. and *Langridge* et al.[152] attempted to design more effective inhibitors for DHFR by combining QSAR and molecular modeling (also see Section 4.3), while *Höltje* et al. and *Richards* et al. tried to obtain better inhibitors using semiempirical[153] and ab initio quantum mechanical methods[154] (see Section 3). The work by *Cody*,[155] *North* et al.,[156] and *Ghose* and *Crippen*[109] on the modeling of DHFR inhibitors should also be mentioned. Extensive NMR studies of inhibitor-enzyme complexes were carried out as well.[157]

Recently, *Kraut*, *Matthews*[158] and co-workers succeeded in determining the structure of at least ten enzyme-inhibitor complexes of DHFR from chicken and *E. coli*. Analysis of these structures helped to explain why trimethoprim had a greater inhibitory effect on bacterial DHFR than on vertebrate DHFR. Those amino acids which are on opposite sides of the cavity of the active site are 1.5 to 2 Å further apart in chicken DHFR than in DHFR from *E. coli*, leading to different arrangements of the inhibitors in the enzyme-inhibitor complexes of the two enzymes and different levels of inhibition. These results give rise to the expectation that, after isolation and crystallographic analysis of a target protein, it may be possible in the future, to use the methods described in Section 5 to design selective inhibitors which do not intervene in the metabolism of the host organism. Drug development would thus become much more efficient.

6.2. Studies on Renin

Inhibitors of the enzyme renin are potential therapeutics against hypertension. No X-ray structure of the enzyme itself has so far been determined. However, some aspartyl proteinases, such as pepsin, *Rhizopus chinensis* pepsin, penicillopepsin and endothiapepsin are closely related to renin. Thus, when the high-resolution 3D-structures of these related enzymes[159-162] and the sequences of mouse renin[163-166] and subsequently also of human renin[164] became available, it was possible to develop three-dimensional models of mouse renin[165] and human renin[166, 167] with the aid of CAMD. On the basis of these computer models and the structures of the enzyme-substrate complexes of pepstatin in *Rhizopus chinensis* pepsin and penicillopepsin,[159, 160, 162] several research groups attempted to design renin inhibitors.[168] This was done in a number of steps:

1) The cavity in which the active site of the enzyme is located was investigated with regard to the available space and the possible sites of binding between the enzyme and the substrate or inhibitor molecule (structural elements which form hydrogen bonds, hydrophobic pockets, etc.).
2) Further important information was obtained from the arrangement of the inhibitor in the enzyme-inhibitor complex. Potential inhibitors should be capable of assuming the same geometry in certain parts of their structure and occupy the same enzyme-binding sites as the model inhibitor. The proposed structures were therefore subjected to special fitting algorithms (see Section 4).
3) Once a structure fulfilled these conditions at low conformation energy it was then introduced interactively into the cavity of the human renin model. If a subsequent relaxation procedure, carried out by means of a molecular mechanics calculation, led to an energetically favorable arrangement, a novel active structure had fortunately been designed. However, definitive information about the inhibitory activity which may have been achieved can only be obtained by experiment. More careful consideration of the transition state of the enzymatic reaction of renin led to novel, highly active compounds,[169] the so-called transition state analogs, and to a deeper understanding of the enzymatic mechanism.[160]

6.3. Protein Engineering

Protein engineering is a relatively new, promising field whose impact on other areas is difficult to assess. Protein engineering can be used to discover structure–activity relationships in enzymes and, based on this, to produce proteins with selectively altered, i.e. novel or improved, properties.[170, 171] These properties include thermal stability, substrate specificity and kinetic behavior of enzymes, as well as therapeutic properties of proteins. For example, the effect of disulfide bridges on the in vitro stability of T4 lysozyme derivatives has been very thoroughly investigated.[172] Extensive studies led among other things to the determination of the structure–activity relationships for subtilisin[171, 173] and for tyrosyl-tRNA synthetase.[174] This requires that one or more amino acids in a protein can be replaced selectively by a genetic engineering method (site-directed mutagenesis) and that the three-dimensional structures of the proteins are made available by crystallography or NMR. These structures are represented, modeled, and exploited by molecular modeling methods. In the simplest case, modeling is carried out interactively.[175] Starting from the structure of the wild type enzyme, the amino acids of interest are replaced. The backbone conformation of the original enzyme is retained and each of the side chains is oriented so that no forbidden steric interactions occur. This can be carried out completely visually, as described in Section 2, or by means of molecular dynamics calculations. The interaction between the protein and the substrate can be estimated by docking procedures (see Section 2). There are also sophisticated methods which include homologous substructures from the Brookhaven protein data bank.[176] To what extent replacement of certain amino acids is accompanied by a change in the tertiary structure of the protein cannot be readily predicted up to now.

Docking of the substrate into the active site of the mutated enzyme investigated can of course also be carried out using an automatic docking program (see Section 5.5). This will certainly give more objective results than interactive docking.

Other objects of interest in the field of enzyme engineering have been the peptide cleavage by thermolysin[177] and electrostatic effects in enzyme catalysis.[178]

7. Conclusions and Prospects

In this review, very different approaches have been described, all of which have a single goal: the design of molecular structures with very specific properties taking into account all available information. Consequently, CAMD is not restricted to one standard procedure; efficient molecular design, as described for the practical examples in Section 6, is based on the availability of several methods side by side and the awareness of the strengths and weaknesses of these methods.

CAMD has now become an indispensable tool in very different areas of molecular research. Apart from more "classic" applications, such as the drug design described in Section 6, CAMD is certain to be used extensively in protein research in the future. Thus, it may be speculated[179] whether, once the structures of the coat proteins of rhinovirus and poliovirus[180] have been solved, it will be possible to use computer-assisted methods to discover new drugs against these viruses.

In the industrial sector, the search for novel biologically active compounds is characterized by an increasing research expenditure. This trend may be reversed by reinforcing industrial research efforts in molecular biology and the use of computer-assisted molecular design to explain the biochemical mechanisms. CAMD will therefore be applied to an even greater extent to the development of new drugs and plant protection agents.

New perspectives are also expected to open up with regard to chemical reactions on solid surfaces (heterogeneous catalysis) and the elucidation of solid state properties.[181]

Further development of CAMD certainly depends on the extent to which supercomputers will be available in the future for calculations for molecular simulation, and on the progress made in the development of new methods and software, possibly involving artificial intelligence.

Received: December 29, 1986;
revised: January 19, 1987 [A 617 IE]
German version: *Angew. Chem. 99* (1987) 413

[1] C. Trindle, *Croat. Chem. Acta 57* (1984) 1231; H. Primas, U. Müller-Herold: *Elementare Quantenchemie*, Teubner, Stuttgart 1984.
[2] J. D. Watson, F. H. C. Crick, *Nature (London) 171* (1953) 737.
[3] J. Avery, *Int. J. Quantum Chem. 26* (1984) 843; P. Hobza, R. Zahradnik, J. Ladik, *ibid. 26* (1984) 857.
[4] C. Levinthal, *Sci. Am. 214* (1966) 42.
[5] R. Langridge, *Fed. Proc. Fed. Am. Soc. Exp. Biol. 33* (1974) 2332; R. J. Feldmann, *Annu. Rev. Biophys. Bioeng. 5* (1976) 477; C. Humblet, G. Marshall, *Drug Dev. Res. 1* (1981) 409; P. Gund, J. D. Andose, J. B. Rhodes, G. M. Smith, *Science (Washington DC) 208* (1980) 1425; R. Langridge, T. E. Ferrin, I. D. Kuntz, M. L. Conolly, *ibid. 211* (1981) 661; P. A. Bash, N. Pattabiraman, C. Huang, T. E. Ferrin, R. Langridge, *ibid. 222* (1983) 1325; R. Diamond, *Twelfth Int. Congr. Crystallogr. 1981*, Abstract 18 5-02; D. A. Pensak, *Ind. Res. Dev. 25* (1983) 74; J. G. Vinter, *Chem. Br. 21* (1985) 32; A. J. Hopfinger, *J. Med. Chem. 28* (1985) 1133.
[6] D. N. J. White, J. E. Pearson, *J. Mol. Graphics 4* (1986) 134.
[7] P. Kollman, *Acc. Chem. Res. 18* (1985) 105.
[8] T. A. Jones, *J. Appl. Crystallogr. 11* (1978) 268.
[9] Biograf, developed by *William Goddard* at the University of California, San Diego, CA.
[10] Chem-X, developed by *Keith Davies* at the University of Cambridge.
[11] Discover & Insight, developed by *Arnold Hagler* at the Agouron Institute, University of California, San Diego, CA; FRODO, orginally developed by *Alwyn Jones*, Max-Planck-Institut für Biochemie, Martinsried (FRG); HYDRA, developed at the University of York, England.
[12] Macromodel, developed by *Clark Still*, Columbia University, New York; MOGLI, developed by *A. Dearing* and Evans & Sutherland, Salt Lake City, UT.
[13] SYBYL, developed by *Garland Marshall*, Washington University, St. Louis, MO.
[14] A. J. Morffew, *J. Mol. Graphics 1* (1983) 83.
[15] For example, Leeds liquid crystal 3-D viewer, Millenium, Stevenage, Herfordshire SG1 4QX, England.
[16] W. M. Newman, D. F. Sproul: *Principles of Interactive Computer Graphics*, McGraw-Hill, New York 1979.
[17] N. C. Cohen, P. Colin, G. Lemoine, *Tetrahedron 37* (1981) 1711.
[18] B. Busetta, I. J. Tickle, T. L. Blundell, *J. Appl. Crystallogr. 16* (1983) 432.
[19] R. Taylor, *J. Mol. Graphics 4* (1986) 123, and references cited therein.
[20] L. Pauling: *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, NY 1960.
[21] M. Born, R. Oppenheimer, *Ann. Phys. (Leipzig) 84* (1927) 457; H. Eyring, J. Walter, G. E. Kimball: *Quantum Chemistry*, Wiley, New York 1944.

[22] W. Kutzelnigg: *Einführung in die Theoretische Chemie, Band 1 und 2*, Verlag Chemie, Weinheim 1975.

[23] R. Daudel, R. Lefebvre, C. Moser: *Quantum Chemistry*, Wiley Interscience, New York 1965; R. McWeeny, B. T. Sutcliffe: *Methods of Molecular Quantum Mechanics*, Academic Press, London 1969.

[24] D. R. Hartree: *Calculation of Atomic Structure*, Wiley, New York 1957.

[25] C. C. J. Roothaan, *Rev. Mod. Phys. 23* (1951) 69, *32* (1960) 179.

[26] L. C. Snyder, H. Basch: *Molecular Wave Functions and Properties*, Wiley, New York 1972.

[27] H. F. Schaefer III: *The Electronic Structure of Atoms and Molecules*, Addison-Wesley, Reading, MA, USA 1972.

[28] I. Shavitt in H. F. Schaefer III (Ed.): *Methods of Electronic Structure Theory*, Plenum, New York 1977.

[29] A. C. Wahl, G. Das in H. F. Schaefer III (Ed.): *Methods of Electronic Structure Theory*, Plenum, New York 1977.

[30] J. S. Binkley, M. J. Frisch, D. J. De Frees, K. Raghavachari, R. A. Whiteside, H. B. Schlegel, E. M. Flunder, J. A. Pople, *Gaussian 82*, Carnegie-Mellon University, Pittsburgh, PA, USA 1983.

[31] G. A. Segal (Ed.): *Semiempirical Methods of Electronic Structure Calculations*, Plenum, New York 1977; M. Scholz, H. J. Köhler: *Quantenchemische Näherungsverfahren*, Hüthig, Heidelberg 1981.

[32] E. B. Wilson, J. C. Decius, P. C. Cross: *Molecular Vibrations*, McGraw-Hill, New York 1955.

[33] S. R. Niketic, K. Rasmussen, *Lecture Notes in Chemistry: The Consistent Force Field. Vol. 3*, Springer, Berlin 1977.

[34] O. Ermer: *Aspekte von Kraftfeldrechnungen*, Bauer, München 1981.

[35] U. Burkert, N. L. Allinger, *Molecular Mechanics* (*ACS Monogr. 177* (1982)).

[36] P. K. Weiner, P. A. Kollman, *J. Comput. Chem. 2* (1981) 287.

[37] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem. 4* (1983) 187.

[38] E. M. Engler, J. D. Andose, P. von R. Schleyer, *J. Am. Chem. Soc. 95* (1973) 8005; D. H. Wertz, N. L. Allinger, *Tetrahedron 30* (1974) 1579.

[39] A. T. Hagler, P. S. Starn, J. S. Ariel, *J. Am. Chem. Soc. 101* (1979) 813.

[40] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc. 106* (1984) 765.

[41] J. H. Schachtschneider, R. G. Snyder, *Spectrochim. Acta 19* (1963) 117.

[42] J. E. Williams, P. J. Stang, P. von R. Schleyer, *Annu. Rev. Phys. Chem. 19* (1968) 531.

[43] M. Levitt, *Annu. Rev. Biophys. Bioeng. 11* (1982) 251; A. Warshel in G. A. Segal (Ed.): *Semiempirical Methods of Electronic Structure Calculation*, Plenum, New York 1977.

[44] R. S. Mulliken, *J. Chem. Phys. 23* (1955) 1833; E. R. Davidson, *ibid. 46* (1967) 3320; K. R. Roby, *Mol. Phys. 27* (1974) 81; R. Heinzmann, R. Ahlrichs, *Theor. Chim. Acta 42* (1976) 33.

[45] M. J. S. Dewar: *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill, New York 1969.

[46] J. Griffiths: *Colour and Constitution of Organic Molecules*, Academic Press, London 1976.

[47] H. C. Longuet-Higgins, *Adv. Spectrosc. 2* (1961) 49; E. E. Nikitin, *Adv. Quantum Chem. 5* (1970) 135.

[48] H. R. Schwarz, H. Rutishauser, E. Stiefel: *Numerik symmetrischer Matrizen*, 2nd ed., Teubner, Stuttgart 1972; R. Fletcher, M. J. D. Powell, *Comput. J. 6* (1963) 163; J. A. Nelder, R. Mead, *ibid. 7* (1965) 308; R. Fletcher, C. M. Reeves, *ibid. 7* (1964) 149; H. B. Curry, *Q. Appl. Math. 2* (1944) 258; A. S. Householder: *Principles of Numerical Analysis*, McGraw-Hill, New York 1953.

[49] H. P. Schwefel: *Numerical Optimization of Computer Models*, Wiley, New York 1981.

[50] B. Mandelbrot: *The Fractal Geometry of Nature*, Freeman, New York 1984.

[51] N. C. Cohen, *ACS Symp. Ser. 112* (1979) 371.

[52] I. Motoc, G. R. Marshall, R. A. Dammkoehler, J. Labanowski, *Z. Naturforsch. A 40* (1985) 1108.

[53] B. Lee, F. M. Richards, *J. Mol. Biol. 55* (1971) 379; F. M. Richards, *Annu. Rev. Biophys. Bioeng. 6* (1977) 151.

[54] R. Lavery, A. Pullman, *Biophys. Chem. 19* (1984) 171.

[55] P. K. Ponnuswamy, P. Thiyagarajan, *Biopolymers 17* (1978) 2503; C. J. Alden, S.-H. Kim, *J. Mol. Biol. 132* (1979) 411.

[56] H. Weinstein, R. Osman, J. P. Green, *ACS Symp. Ser. 112* (1979) 161.

[57] D. Hadzi, M. Hodoscec, D. Kocjan, T. Solmajer, F. Avbelj, *Croat. Chem. Acta 57* (1984) 1065; P. M. Dean, *Pharmacochem. Libr. 6* (1983) 137.

[58] Z. Simon, A. Chiriac, I. Motoc, S. Holban, D. Ciubotariu, Z. Szabadai, *Stud. Biophys. 55* (1976) 217.

[59] S. Navasivayam, P. M. Dean, *J. Mol. Graphics 4* (1986) 46.

[60] R. Klein, unpublished.

[61] G. Moreau, P. Broto, *J. Mol. Graphics*, in press.

[62] P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem. Chim. Ther. 19* (1984) 61.

[63] G. Naray-Szabo, *THEOCHEM 27* (1986) 401.

[64] G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler, D. A. Dunn, *ACS Symp. Ser. 112* (1979) 205.

[65] G. J. Crippen, *J. Comput. Phys. 24* (1977) 96.

[66] G. J. Crippen, *J. Med. Chem. 23* (1980) 599.

[67] M. R. Linschoten, T. Bultsma, A. P. Ijzerman, H. Timmerman, *J. Med. Chem. 29* (1986) 278.

[68] P. R. Andrews, *ACS Symp. Ser. 112* (1979) 149.

[69] T. E. Klein, C. Huang, T. E. Ferrin, R. Langridge, C. Hansch, *ACS Symp. Ser. 306* (1986) 147.

[70] I. Motoc, G. R. Marshall, J. Labanowski, *Z. Naturforsch. A 40* (1985) 1121.

[71] G. R. Marshall, *Pharmacochem. Libr. 6* (1983) 129.

[72] P. C. Jurs, J. T. Chou, M. Yuan, *ACS Symp. Ser. 112* (1979) 103; K. H. Ting, R. C. T. Lee, G. W. A. Milne, H. Shapiro, A. M. Gaurino, *Science (Washington DC) 180* (1973) 417; W. J. Dunn III, S. Wold, Y. C. Martin, *J. Med. Chem. 21* (1978) 922.

[73] T. R. Stouch, P. C. Jurs, *EHP Environ. Health Perspect. 61* (1985) 329; W. Sacco, W. P. Ashman, P. H. Broome, J. King, *DHEW Publ. FDA US 78-1046* (1978) 68; S. A. Leavitt, M. J. Mass, *Cancer Res. 45* (1985) 4741; S. L. Rose, P. C. Jurs, *J. Med. Chem. 25* (1982) 769.

[74] G. Klopman, *J. Am. Chem. Soc. 106* (1984) 7315.

[75] W. J. Streich, R. Franke, *Quant. Struct. Act. Relat. Pharmacol. Chem. Biol. 4* (1985) 13.

[76] R. Osman, H. Weinstein, J. P. Green, *ACS Symp. Ser. 112* (1979) 21.

[77] J. E. Leffler, E. Grunwald: *Rates of Equilibria of Organic Reactions*, Wiley, New York 1963.

[78] L. P. Hammett, *J. Am. Chem. Soc. 59* (1937) 96.

[79] R. W. Taft, *J. Am. Chem. Soc. 74* (1952) 3126, *75* (1953) 4538.

[80] C. Hansch, P. P. Malony, T. Fujita, R. Muir, *Nature (London) 194* (1962) 178.

[81] K. H. Meyer, H. Hemmi, *Biochem. Z. 277* (1935) 39.

[82] E. Overton: *Studium über die Narkose*, Fischer, Jena 1901.

[83] C. Hansch, *Acc. Chem. Res. 2* (1969) 232.

[84] L. P. Hammett: *Physical Organic Chemistry*, McGraw-Hill, New York 1940.

[85] H. H. Jaffé, *Chem. Rev. 53* (1953) 191; R. W. Taft, I. C. Lewis, *J. Am. Chem. Soc. 80* (1958) 2436; C. G. Swain, E. C. Lupton, *ibid. 90* (1968) 4328.

[86] O. R. Quayle, *Chem. Rev. 53* (1953) 484.

[87] J. H. Hildebrand, R. L. Scott: *The Solubility of Nonelectrolytes*, Reinhold, New York 1950.

[88] R. M. Carlson, R. E. Carlson, H. L. Kopperman, *J. Chromatogr. 107* (1975) 219; E. Tomlinson, *ibid. 113* (1975) 1.

[89] I. Motoc, O. Dragomir-Filimonescu, *MATCH 12* (1981) 127.

[90] M. Charton, *Prog. Phys. Org. Chem. 10* (1973) 81.

[91] A. Verloop, W. Hoogenstraaten, J. Tipker in E. J. Ariens (Ed.): *Drug Design, Vol. 7*, Academic Press, New York 1976.

[92] A. Verloop, *QSAR Strategies Des. Bioact. Compd. Proc. Eur. Symp. Quant. Struct. Act. Relat. 5th 1984* (1985) 98–104.

[93] Z. Simon: *Quantum Biochemistry and Specific Interactions*, Abacus Press, Kent, England 1976.

[94] A. T. Balaban, A. Chiriac, I. Motoc, Z. Simon, *Lecture Notes in Chemistry: Steric Fit in QSAR, Vol. 15*, Chapters 4, 5, 7, Springer, Berlin 1980.

[95] I. Motoc, O. Dragomir-Filimonescu, *MATCH 12* (1981) 117; Z. Simon, *ibid. 13* (1982) 357.

[96] I. Motoc, G. R. Marshall, *Z. Naturforsch. A 40* (1985) 1114.

[97] P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem. Chim. Ther. 19* (1984) 79.

[98] M. J. S. Dewar, P. J. Grisdale, *J. Am. Chem. Soc. 84* (1962) 3539.

[99] P. R. Andrews, *J. Med. Chem. 15* (1972) 1069.

[100] N. R. Draper, H. Smith: *Appl. Regression Analysis*, Wiley, New York 1966.

[101] P. P. Mager, *Med. Res. Rev. 2* (1982) 93.

[102] W. J. Streich, S. Dove, R. Franke, *J. Med. Chem. 23* (1980) 1452.

[103] S. Dove, R. Franke, O. L. Mndshojan, W. A. Schkuljiew, L. W. Chashakjan, *J. Med. Chem. 22* (1979) 90.

[104] C. Hansch, S. H. Unger, A. B. Forsythe, *J. Med. Chem. 16* (1973) 1217; W. J. Dunn III, W. J. Greenberg, S. S. Callejas, *ibid. 19* (1976) 1299.

[105] G. Klopman, A. N. Kalos, *J. Comput. Chem. 6* (1985) 492; Y. C. Martin: *Quantitative Drug Design*, Dekker, New York 1978.

[106] J. G. Topliss, R. P. Edwards, *ACS Symp. Ser. 112* (1979) 131.

[107] S. M. Free, Jr., J. W. Wilson, *J. Med. Chem. 7* (1964) 395.

[108] G. M. Crippen, *J. Med. Chem. 22* (1979) 988.

[109] A. K. Ghose, G. M. Crippen, *J. Med. Chem. 27* (1984) 901.

[110] C. Hansch, J. M. Blaney, *Proc. Rhone-Poulenc Round Table Conf. 3rd Meeting 1982* (1984) 185; M. Recanatini, T. Klein, C. Z. Yang, J. McClarin, R. Langridge, C. Hansch, *Mol. Pharmacol. 29* (1986) 436.

[111] R. Klein, unpublished.

[112] D. N. J. White, C. Morrow, *Comput. Chem. 3* (1979) 33.

[113] D. N. J. White, D. H. Kitson, *J. Mol. Graphics 4* (1986) 112.

[114] B. Pullman in B. Pullman (Ed.): *Quantum Mechanics of Molecular Conformations*, Wiley, New York 1976; L. Pedersen, *EHP Environ. Health Perspect. 61* (1985) 185.

[115] P. De Santis, A. M. Liquori, *Biopolymers 10* (1971) 699; G. Nemethy, H. A. Scheraga, *ibid. 3* (1965) 155; H. A. Scheraga, *ibid. 22* (1983) 1.

[116] P. Y. Chou, G. D. Fasman, *Adv. Enzymol. Relat. Areas Mol. Biol. 47* (1978) 45; G. D. Rose, L. M. Gierasch, J. A. Smith, *Adv. Protein Chem. 37* (1985) 1.

[117] G. N. Ramachandran, V. Sasisekharan, *Adv. Protein Chem. 23* (1968) 283.

[118] L. Verlet, *Phys. Rev. 159* (1967) 98; W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys. 34* (1977) 1311.

[119] D. Fincham, D. M. Heyes, *Adv. Chem. Phys. 63* (1985) 493.

[120] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys. 81* (1984) 3684.

[121] D. N. Theodorou, U. W. Suter, *J. Chem. Phys. 82* (1985) 955.

[122] W. van Gunsteren, H. Berendsen, J. Hollenberg, *Supercomputer 7* (1985) 26.

[123] F. Vesely: *Computerexperimente an Flüssigkeitsmodellen*, Physik Verlag, Weinheim 1978.

[124] T. Tsang, W. D. Jenkins, *Mol. Phys. 41* (1980) 797; J. G. Powles, *ibid. 48* (1983) 1083; G. S. Grest, S. R. Nagel, A. Rahman, T. A. Witten, *J. Chem. Phys. 74* (1981) 3532.

[125] P. J. Rossky, M. Karplus, A. Rahman, *Biopolymers 18* (1979) 825.

[126] K. Remerie, W. F. van Gunsteren, J. B. F. N. Engberts, *Mol. Phys. 56* (1985) 1393.

[127] J. L. Tallon, R. M. J. Cotterill, *Aust. J. Phys. 38* (1985) 209; W. M. Evans, G. J. Evans, *Adv. Chem. Phys. 63* (1985) 377; G. S. Pawley, M. A. Brass, T. M. Dove, K. Refson, *J. Chim. Phys. Phys. Chim. Biol. 82* (1985) 249.

[128] D. C. Wallace, *NATO ASI Ser. Ser. B121* (1985) 521; M. Kimura, F. Yonezawa, *Springer Ser. Solid State Sci. 46* (1983) 80; S. M. Levine, S. H. Garofalini, *Mater. Res. Soc. Symp. Proc. 61* (1986) 29.

[129] T. A. Weber, E. Helfand, *J. Chem. Phys. 71* (1979) 4760; E. Helfand, *Physica A (Amsterdam) 118* (1983) 123; D. Brown, J. H. R. Clarke, *J. Chem. Phys. 84* (1986) 2858.

[130] M. Levitt, *Cold Spring Harbor Symp. Quant. Biol. 47* (1983) 251; B. Tidor, K. K. Irikura, B. R. Brooks, M. Karplus, *J. Biomol. Struct. Dyn. 1* (1983) 231; S. Harvey, M. Prabhakaran, B. Mao, J. A. McCammon, *Science (Washington DC) 223* (1984) 1189.

[131] A. T. Hagler, P. Dauber, D. J. Osguthorpe, J. Hempel, *Science (Washington DC) 227* (1985) 1309.

[132] R. S. Struthers, J. Revier, A. T. Hagler, *Ann. NY Acad. Sci. 439* (1985) 81.

[133] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature (London) 267* (1977) 585; M. Karplus, J. A. McCammon, *ibid. 277* (1979) 578; S. H. Northrup, M. R. Pear, J. A. McCammon, M. Karplus, T. Takano, *ibid. 287* (1980) 659; M. Levitt, *J. Mol. Biol. 168* (1983) 595; W. F. van Gunsteren, H. J. C. Berendsen, *ibid. 176* (1984) 559; W. F. van Gunsteren, M. Karplus, *Biochemistry 21* (1982) 2259; J. A. McCammon, M. Karplus, *Acc. Chem. Res. 16* (1983) 187; W. F. van Gunsteren, H. J. C. Berendsen, J. Hermans, W. G. J. Hol, J. P. M. Postma, *Proc. Natl. Acad. Sci. USA 80* (1983) 4315; P. Krüger, W. Strassburger, A. Wollmer, W. F. van Gunsteren, *Eur. Biophys. J. 13* (1985) 77.

[134] J. P. Ryckaert, A. Bellemans, *Faraday Discuss. Chem. Soc. 66* (1978) 95; P. van der Ploeg, H. J. C. Berendsen, *J. Chem. Phys. 76* (1982) 3271; *Mol. Phys. 49* (1982) 233; R. D. Mountain, R. M. Mazo, J. J. Volwerk, *Chem. Phys. Lipids 40* (1986) 35.

[135] N. S. Ostlund, R. A. Whiteside, *Ann. NY Acad. Sci. 439* (1985) 195.

[136] W. Fischer, J. Brickmann, P. Läuger, *Biophys. Chem. 13* (1981) 105; P. Läuger, H. J. Apell, *ibid. 16* (1982) 209; J. Brickmann, W. Fischer, *ibid. 17* (1983) 245; D. H. J. Mackay, P. H. Berens, K. R. Wilson, A. T. Hagler, *Biophys. J. 46* (1984) 229.

[137] K. Binder (Ed.): *Monte Carlo Methods in Statistical Physics*, Springer, Berlin 1979; B. J. Berne (Ed.): *Statistical Mechanics*, Plenum, New York 1977.

[138] G. G. Lowry (Ed.): *Markov Chains and Monte Carlo Calculations in Polymer Science*, Dekker, New York 1970.

[139] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys. 21* (1953) 1087.

[140] A. T. Hagler, J. Moult, *Nature (London) 272* (1978) 222; A. T. Hagler, D. J. Osguthorpe, B. Robson, *Science (Washington DC) 208* (1980) 601; G. Corongiu, E. Clementi, *Biopolymers 20* (1981) 551, *21* (1982) 763; V. Madison, D. J. Osguthorpe, P. Dauber, A. T. Hagler, *ibid. 22* (1983) 27.

[141] F. L. McCrackin, *J. Chem. Phys. 47* (1967) 1980; M. Lal, *Mol. Phys. 17* (1969) 57; M. Lal, D. Spencer, *ibid. 22* (1971) 649; R. De Santis, H. G. Zachmann, *Prog. Colloid Polym. Sci. 64* (1978) 281; R. Potenzone, D. C. Doherty, *Polym. Mater. Sci. Eng. 52* (1985) 264.

[142] B. J. McClelland: *Statistical Thermodynamics*, Chapman, London 1973; D. A. McQuarrie: *Statistical Mechanics*, Harper and Row, New York 1976.

[143] A. Ben-Naim: *Hydrophobic-Interactions*, Plenum, New York 1980.

[144] D. L. Beveridge, M. Mezei, G. Rhavishanker, B. Jayaram, *J. Biosci. 8* (1985) 167; M. Mezei, P. K. Mehrotra, D. L. Beveridge, *J. Am. Chem. Soc. 107* (1985) 2239.

[145] M. Karplus, J. Kushick, *Macromolecules 14* (1981) 325; R. M. Levy, M. Karplus, J. Kushick, D. Perahia, *ibid. 17* (1984) 1370.

[146] R. L. Humphries, *J. Theor. Biol. 91* (1981) 477.

[147] H. Wallmeier, unpublished.

[148] K. W. Volz, D. A. Matthews, R. A. Alden, S. T. Freer, C. Hansch, T. B. Kaufman, J. Kraut, *J. Biol. Chem. 257* (1982) 2528.

[149] J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, J. Kraut, *J. Biol. Chem. 257* (1982) 13650; D. J. Filman, J. T. Bolin, D. A. Matthews, J. Kraut, *ibid.* 13663.

[150] L. F. Kuyper, B. Roth, D. B. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. A. Norrington, *J. Med. Chem. 28* (1985) 303; A. Stuart, T. Paterson, B. Roth, E. Aig, *ibid. 26* (1983) 667; R. M. Hyde, R. A. Paterson, C. R. Beddell, J. N. Champness, D. K. Stammers, D. J. Baker, P. J. Goodford, L. F. Kuyper, R. Ferone in J. A. Blair (Ed.): *Chemistry and Biology of Pteridines*, de Gruyter, Berlin 1983, pp. 505–509; D. K. Stammers, J. N. Champness, J. G. Dann, C. R. Bedell, *ibid.*, pp. 567–571; D. J. Baker, C. R. Bedell, J. N. Champness, P. J. Goodford, F. E. Norrington, B. Roth, D. K. Stammers, *ibid.*, pp. 545–549; M. Mabilia, R. A. Pearlstein, A. J. Hopfinger, *Eur. J. Med. Chem. Chim. Ter. 20* (1985) 163; *QSAR Strategies Des. Bioact. Compd. Proc. Eur. Symp. Quant. Struct. Act. Relat. 5th 1984* (1985) 120; A. J. Hopfinger, *J. Med. Chem. 26* (1983) 990; L. F. Kuyper, B. Roth, D. B. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. A. Norrington, D. J. Baker, P. J. Goodford, *ibid. 25* (1982) 1120; H. B. Schlegel, M. Poe, K. Hoogsteen, *Mol. Pharmacol. 20* (1981) 154.

[151] B. Birdsall, J. Feeney, C. Fascual, G. C. K. Roberts, I. Kompis, R. L. Then, K. Mueller, A. Kroehn, *J. Med. Chem. 27* (1984) 1672; H. Maag, R. Locher, J. J. Daly, I. Kompis, *Helv. Chim. Acta 69* (1986) 887; K. Mueller, *Actual. Chim. Ther. 11* (1984) 113.

[152] C. D. Selassie, Z. X. Fang, R. L. Li, C. Hansch, T. Klein, R. Langridge, T. B. Kaufman, *J. Med. Chem. 29* (1986) 621; C. Hansch, B. A. Hathaway, Z. Guo, C. D. Selassie, S. W. Dietrich, J. M. Blaney, R. Langridge, K. W. Volz, T. B. Kaufman, *ibid. 27* (1984) 129; C. Hansch, *Drug Intell. Clin. Pharm. 16* (1982) 391; R. Li, C. Hansch, D. Matthews, J. M. Blaney, R. Langridge, T. J. Delcamp, S. S. Susten, J. H. Freisheim, *Quant. Struct. Act. Relat. Pharmacol. Chem. Biol. 1* (1982) 1.

[153] H. D. Höltje, P. Zunker, *QSAR Strategies Des. Bioact. Compd. Proc. Eur. Symp. Quant. Struct. Act. Relat. 5th 1984* (1985) 127.

[154] A. F. Cuthbertson, W. G. Richards, *THEOCHEM 25* (1985) 167.

[155] V. Cody, *J. Mol. Graphics 4* (1986) 69; *Prog. Clin. Biol. Res. 172* (1985) 275.

[156] E. A. Potterson, A. J. Geddes, A. C. T. North in J. A. Blair (Ed.): *Chemistry and Biology of Pteridines*, de Gruyter, Berlin 1983, pp. 299–303.

[157] J. Feeney, *NATO ASI Ser. Ser. C164* (1986) 347.

[158] D. A. Matthews, J. T. Bolin, J. M. Burridge, D. J. Filman, D. J. Volz, J. Kraut, *J. Biol. Chem. 260* (1985) 392.

[159] M. N. G. James, A. R. Sielecki, *J. Mol. Biol. 163* (1983) 299.

[160] M. N. G. James, A. R. Sielecki, *Biochemistry 24* (1985) 3701.

[161] E. Subramanian, I. D. A. Swan, M. Liu, D. R. Davies, J. A. Jenkins, I. J. Tickle, T. L. Blundell, *Proc. Natl. Acad. Sci. USA 74* (1977) 556.

[162] R. Bott, E. Subramanian, D. R. Davies, *Biochemistry 21* (1982) 6956.
[163] K. S. Misono, J. J. Chang, T. Inagami, *Proc. Natl. Acad. Sci. USA 79* (1982) 4858; J.-J. Panthier, S. Foote, B. Chambreaud, A. D. Strosberg, P. Corvol, F. Rougeon, *Nature (London) 298* (1982) 90.
[164] F. Soubrier, J.-J. Panthier, P. Corvol, F. Rougon, *Nucleic Acids Res. 11* (1983) 7181; T. Imai, H. Miyazaki, S. Hirose, H. Hori, T. Hayashi, R. Kageyama, H. Ohkubo, S. Nakanishi, K. Murakami, *Proc. Natl. Acad. Sci. USA 80* (1983) 7405.
[165] T. Blundell, B. L. Sibanda, L. Pearl, *Nature (London) 304* (1983) 273.
[166] B. L. Sibanda, T. Blundell, P. M. Hobart, M. Fogliano, J. S. Bindra, B. W. Dominy, J. M. Chirgwin, *FEBS Lett. 174* (1984) 102; B. L. Sibanda, *Biochem. Soc. Trans. 12* (1984) 953; W. Carlson, E. Haber, R. Feldmann, M. Karplus, *Pept. Struct. Funct. Proc. Am. Pept. Symp. 8th 1983*, 821; W. Carlson, M. Handschumacher, M. Karplus, E. Haber, *J. Hypertens. 2 (Suppl. 3)* (1984) 281; K. Murakami, *ibid. 2 (Suppl. 3)* (1984) 285; W. Carlson, M. Karplus, E. Haber, *Hypertension (Dallas) 7* (1985) 13; K. Akahane, H. Umeyama, S. Nakagawa, I. Moriguchi, S. Hirose, K. Kinji, K. Murakami, *ibid. 7* (1985) 3.
[167] K. Murakami, N. Ueno, *Proc. Jpn. Acad. Ser. B61* (1985) 257.
[168] B. L. Sibanda, A. M. Hemmings, T. Blundell, *Proc. FEBS Adv. Course 1984*, 339; de Gruyter, Berlin 1985. P. Raddatz, C. Schnittenhelm, G. Barnickel, *Kontakte 1985*, 3; J. Boger, *Spec. Publ. R. Soc. Chem. 55* (1986) 271; G. Breipohl, H. Frühbeis, J. Knolle, D. Ruppert, B. A. Schoelkens, *J. Hypertens. 3* (1985) 406.
[169] M. Szelke, B. J. Leckie, A. Hallett, D. M. Jones, J. Sueiras-Diaz, B. Atrash, A. F. Lever, *Nature (London) 299* (1982) 555.
[170] G. Winter, P. Carter, H. Bedouelle, D. Lowe, R. J. Leatherbarrow, A. R. Fersht, *Life Sci. Res. Rep. 35* (1986) 55; D. M. Blow, A. R. Fersht, G. Winter (Eds.): *Design, Construction and Properties of Novel Protein Molecules (Proceedings of a Royal Society Discussion Meeting held on June 5 and 6, 1985)*. The Royal Society, London 1986.
[171] D. A. Estell, J. V. Miller, T. P. Graycar, D. B. Powers, J. A. Wells, *World Biotech. Rep. 1984*, 181.
[172] L. J. Perry, H. L. Heyneker, R. Wetzel, *Gene 38* (1985) 259; L. J. Perry, R. Wetzel, *Science (Washington DC) 226* (1984) 555; *Biochemistry 25* (1986) 733.
[173] D. A. Estell, T. P. Graycar, J. V. Miller, D. B. Powers, J. P. Burnier, P. G. Ng, J. A. Wells, *Science (Washington DC) 233* (1986) 659; R. R. Bott, M. H. Ultsch, A. A. Kossiakoff, J. Burnier, J. A. Wells, D. B. Powers, B. A. Katz, B. C. Cunningham, S. S. Power, *World Biotech. Rep. 1985*, 611; J. A. Wells, D. B. Powers, *J. Biol. Chem. 261* (1986) 6564.
[174] A. R. Fersht, J.-P. Shi, A. J. Wilkinson, D. M. Blow, P. Carter, M. M. Y. Waye, G. P. Winter, *Angew. Chem. 96* (1984) 455; *Angew. Chem. Int. Ed. Engl. 23* (1984) 467; A. R. Fersht, R. J. Leatherbarrow, T. N. C. Wells, *Trends Biochem. Sci. 11* (1986) 321; T. N. C. Wells, C. K. Ho, A. R. Fersht, *Biochemistry 25* (1986) 6603; R. J. Leatherbarrow, A. R. Fersht, G. Winter, *Proc. Natl. Acad. Sci. USA 82* (1985) 7840; P. G. Thomas, A. J. Russell, A. R. Fersht, *Nature (London) 318* (1985) 375.
[175] R. R. Bott, *World Biotech. Rep. 1984*, 611.
[176] T. Blundell, M. J. E. Sternberg, *Trends Biotechnol. 3* (1985) 228.
[177] G. M. Smith, D. G. Hangauer, J. D. Andose, B. R. Bush, E. M. Fluder, P. Gund, E. F. McIntyre, *Drug. Inf. J. 18* (1984) 167.
[178] A. Warshel, F. Sussman, *Proc. Natl. Acad. Sci. USA 83* (1986) 3806.
[179] W. G. J. Hol, *Angew. Chem. 98* (1986) 765; *Angew. Chem. Int. Ed. Engl. 25* (1986) 767.
[180] M. G. Rossmann, E. Arnold, J. W. Erickson, E. A. Frankenberger, J. P. Griffith, H.-J. Hecht, J. E. Johnson, G. Kamer, M. Luo, A. G. Mosser, R. R. Rueckert, B. Sherry, G. Vriend, *Nature (London) 317* (1985) 145; J. M. Hogle, M. Chow, D. J. Filman, *Science (Washington DC) 229* (1985) 1358.
[181] C. R. A. Catlow, M. Doherty, G. D. Price, M. J. Sanders, S. C. Parker, *Mater. Sci. Forum 7* (1986) 163.