Receptor-Ligand Interaction and Molecular Modelling

V Horňák[1], R. Dvorský[1] and E. Šturdík[2]

1 *Institute of Molecular Biology, Slovak Academy of Sciences, 84251 Bratislava, Slovakia*
2 *Department of Biochemical Technology, Slovak Technical University, 81237 Bratislava, Slovakia*

Abstract. A number of computational methods for the description of the ligand-receptor interaction that were developed in the past decade are reviewed in this paper. The central two sections introduce the methods that are already established as useful tools for the qualitative and quantitative description of ligand-receptor complexes, either when the detailed atomic structure of the receptor is known or not The following section gives two examples of the application of the described methodology on two limiting cases.

Key words: Receptor-ligand interactions — Drug design — Molecular modeling

Introduction

Receptor-ligand interactions are central to numerous biological processes such as signal transduction, physiological regulation, gene transcription, and enzymatic reactions. As many proteins regulate key biological functions via interactions with small molecules, these receptor proteins are often prime targets for therapeutic agents. A detailed understanding of interactions between small molecules and proteins may therefore form the basis for a rational drug-design strategy. Rational drug design is attractive as a drug development paradigm for two reasons: it offers some hope for reduction of the enormous costs and time required in traditional random screening protocols for drug discovery, and may facilitate the development of more selective therapeutic agents with fewer undesirable side effects.

Developments in molecular biology over the past 15 years make it possible to obtain experimentally useful quantities of numerous receptor proteins of potential therapeutic importance. Technical advances in structural characterisation methods (X-ray crystallography, NMR spectroscopy) have made it easier to obtain high-resolution structural data for many important ligand-protein complexes. Still, there is at present no structural information for the vast majority of therapeuti-

Correspondence to Viktor Horňák, Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, 842 51 Bratislava, Slovakia
E-mail umbivhor@savba.savba.sk

cally interesting receptor proteins. Over the past years, there have been significant advances and improvements (Kollman 1994) in computational tools used for the quantitative description of the ligand-protein interaction and rational drug-design applications which are the focus of this review.

Receptor-ligand interaction

The term receptor is usually used as a synonym for any biological target that binds specifically with a small molecule, i.e. ligand and, as a result of this interaction, some biological response is observed (Dean 1987). Ligand (or drug if we emphasise its pharmacological effect) is most often an inhibitor or substrate and has low molecular weight. On the other hand, receptor is represented by the macromolecule, be it an enzyme (Seshadri et al. 1995), receptor protein (Blaney et al. 1982), nucleic acid (Pearson and Prescott 1997) or even membrane (Hianik et al. 1996). In a narrower sense, receptors are soluble, membrane-anchored or membrane-embedded proteins that are able to produce a certain biological response. The observed biological response (or biological activity) can be a result of a series of mostly unknown events that follow the specific interaction of ligand and receptor. In the following, we concentrate on the qualitative and quantitative description of receptor-ligand interaction.

The affinity of a ligand L to its binding site at the receptor R is determined by the free energy difference ΔG between the free states of interacting molecules and their complex LR:

$$\mathrm{L} + \mathrm{R} \underset{}{\overset{K_a}{\rightleftharpoons}} \mathrm{LR} \tag{1}$$

This equilibrium is characterised by association constant $K_a = [\mathrm{LR}]/[\mathrm{L}][\mathrm{R}]$ which is related to the free energy ΔG by the well known equation $\Delta G = -RT \ln K_a$. Hence K_a (or the corresponding ΔG) is the measure of the affinity of the ligand to the receptor, and its use and prediction represents the main goal of theoretical models described below. It is important to emphasise that most ligands form non-covalent complexes so that weak (Van der Waals) interactions play a prominent role. Only alkylating agents (e.g. antitumor drugs like cyclophosphamide) as well as active site directed irreversible enzyme inhibitors (e.g. the penicillins and cephalosporins as bacterial cell wall synthesis inhibitors) form covalent bonds.

The theoretical approaches to the description of the receptor-ligand interaction can be divided into two categories. If the detailed molecular structure (either from X-ray or NMR analysis) of the receptor is known, we try to calculate binding affinities from the "first principles" , using the methods of quantum mechanics, molecular mechanics, simulation techniques etc. (Grant and Richards 1995; Leach 1996). On the other hand, if we do not know the structure of the receptor, we use the semiempirical techniques that were developed in QSAR (quantitative structure-activity relationships) (Martin 1978; Kubinyi 1993). The first approach provides the detailed picture and physical understanding of the interaction, and strives to get the observed biological activity with no further experimental input. The second

approach, which is empirical in nature, requires that the biological activities of a series of ligands (mostly structural analogs) are determined experimentally. Using these data, which reflect the influence of structural variations on the binding affinity, we try to deduce the qualitative relationship between structure and properties of ligands and the activity they elicit on interaction. Though the description of the interaction is not so detailed, the biological activity and its potential prediction is made based on *in vivo* systems, henceforth being more closely related to what we are interested in real applications.

Known structure of the receptor

Molecular docking

The aim of molecular docking (Bamborough and Cohen 1996) is to predict the structure of the intermolecular complex formed by two or more molecules. Most methods concentrate on docking of ligands to macromolecular binding sites. The simplest docking algorithms treat the two molecules as rigid bodies and explore only the degrees of their translational and rotational freedom striving to achieve a high degree of shape complementarity. A well-known example of such an approach is the DOCK program (Kuntz et al. 1982). Further development of docking schemes focused on methods considering the conformational flexibility of ligands (Oshiro et al. 1995). As the number of degrees of freedom increases enormously, Monte Carlo methods, simulated annealing and genetic algorithms are often used. All the methods use molecular mechanics for the computation of the interaction energy of ligands within the binding site thereby providing the scoring for the proposed binding modes.

Another strategy, often referred to as de novo drug design, builds molecules within the rigid binding site. Fragments of molecules are docked separately into the site and then linked together. This methodology was implemented in programs LUDI (Böhm 1992a,b) and CAVEAT (Lauri and Bartlett 1994). Alternatively, molecules are grown in a stepwise fashion in the receptor cavity from library templates. This was implemented in the computer program GROW (Moon and Howe 1991) and applied by the authors to the automatic design of the peptides inhibiting protease renin.

GRID

Program GRID (Goodford 1985) is used for finding energetically favourable regions in protein binding sites. The interaction of a probe group with a protein of a known structure is computed at sample positions throughout and around the macromolecule, giving an array of energy values. This is usually realised by setting a three-dimensional grid around the molecule followed by a calculation of the non-bonded interaction energy at each grid point. The energy is calculated as the sum of the Lennard-Jones, electrostatic and hydrogen bond contributions. The typical probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl. Contour surfaces at appropriate energy levels are calculated for each probe

and displayed by computer graphics together with the protein structure. Contours at negative energy levels delineate regions of attraction between probe and protein and are found at known ligand binding clefts in particular. The contours also enable other regions of attraction to be identified and facilitate the interpretation of protein-ligand energetics. The contours from different probes can be combined to suggest the regions for ligand docking.

Multiple copy minimisation

A second method (Miranker and Karplus 1991), which produces a similar end result, is based on energy minimisation rather then on systematic searches on a grid. Again, typical functional groups are represented by small probe molecules. First, the active site of the target protein, or its complete surface, is surrounded by many hundreds of copies of the probe molecule. It is essential that these are randomly placed to eliminate bias due to the starting conditions. The next stage is to minimise all of these molecules simultaneously with respect to the protein structure. In the minimisation procedure the probe molecules cannot see each other, in the sense that no forces are calculated between them; only protein-probe interactions are considered. In this way very many trial orientations or starting positions can be tried at very little extra computational expense. If the minimisation procedure is successful the resulting probe positions should show some clustering around favourable binding sites and each cluster can be reduced to a single representative molecule. As in the previous case the most energetically favourable positions can be contoured for graphical display and the maps from different probe molecules combined. One very important extension which has been made to this method is to allow the inclusion of protein flexibility and so cooperative movements between a ligand and its receptors are simulated.

Electrostatic potential and free energy – Poisson-Boltzmann equation

Mobile ions or charges in the solvent produce an electrostatic potential which must fulfil the Poisson equation of electrostatics. On the other hand, this potential influences their own positions thereby establishing the mutual dependence of the electrostatic potential and ion distribution. To describe the ion distribution, we take a single ion as the reference point and treat the other ions in an average way as forming the ion atmosphere around this central ion. The mean distribution of charge around the central ion in such a mean potential $\varphi(\boldsymbol{r})$ is given by the Boltzmann distribution. The mean potential and ion distributions are thus obtained from a simultaneous solution to both the Boltzmann and Poisson equations, in other words, solving the Poisson-Boltzmann (PB) equation in the form (Sharp and Honig 1990):

$$\nabla \cdot \varepsilon(\boldsymbol{r}) \nabla \varphi(\boldsymbol{r}) = f(\varphi) \tag{2}$$

where $f(\varphi)$ is a rather complex function of the mean potential φ, and $\varepsilon(\boldsymbol{r})$ is the dielectric constant which can change in space. Knowing the electrostatic potential from this equation, it is possible to derive the expression for calculating the total electrostatic free energy.

Debye and Hückel first used the linearized form of this equation developing a theory of ion activities. The numerical solution of the full form of PB equation is now available at high accuracy which enables direct calculation of the solvation free energy. Different values of dielectric constants are used to sample the different kinds of solvent whereas the local (position dependent) dielectric constant in the PB equation can be used for treating the solute polarizability in solvation processes. The Poisson-Boltzmann method has proven an efficient and accurate way to model the effects of water and ion screening on the potentials on and around soluble proteins, nucleic acids and membranes.

Free energy difference calculation

Using the so-called coupling parameter approach, the difference in free energy between two states of a system can be determined (Van Gunsteren et al. 1994). The free energy of a molecular system is dependent on the extent of phase or configuration space that is accessible to the system at the thermodynamic state point of interest. Complete sampling of configuration space is not possible except for the simplest model systems. Free-energy calculations of biomolecular systems have thus concentrated on the determination of the relative free energy between two closely related states. In this way only differences between the two states need to be considered. Irrelevant regions of configuration space can be ignored. Although this greatly simplifies the problem, it does not eliminate the necessity to sample relevant regions of configuration space, especially if entropic contributions to the free energy are to be correctly estimated.

The difference in free energy between two states A and B of a system, of which the interaction functions are denoted by V_A and V_B, can be calculated from the expression:

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \langle \partial V / \partial \lambda \rangle_\lambda \mathrm{d}\lambda \tag{3}$$

The potential energy function V is made a function of the coupling parameter λ, such that $V(\lambda_A) = V_A$ and $V(\lambda_B) = V_B$. Averaging over configurations generated with the interaction function $V(\lambda)$ is denoted by $\langle \ldots \rangle_\lambda$. The coupling parameter can be made a function of time, $\lambda(t)$, such that it slowly changes from λ_A to λ_B over the time course of an MD simulation.

The next step when using the thermodynamic integration technique to calculate relative free energies or binding constants of receptor-ligand complexes, is to formulate a so-called thermodynamic cycle. The basis on which the thermodynamic cycle approach rests is the fact that the free energy G is a thermodynamic state function. This means that as long as a system in equilibrium is changed in a reversible way, the change in free energy ΔG will be independent of the path of change, and along a closed path or cycle one has $\Delta G = 0$. The power of this thermodynamic cycle technique lies in the fact that on a computer also non-chemical processes such as the conversion of one type of atom into another type may be performed. In order to demonstrate the method we consider the relative binding of

two inhibitors I_A and I_B to an enzyme E. The appropriate thermodynamic cycle for obtaining the relative binding constant (K_2/K_1) is

$$\begin{array}{ccc} E + I_A & \xrightarrow{1} & E : I_A \\ {\scriptstyle 3}\downarrow & & \downarrow{\scriptstyle 4} \\ E + I_B & \xrightarrow[2]{} & E : I_B \end{array} \tag{4}$$

The relative binding constant equals:

$$\frac{K_2}{K_1} = \exp\left(-\frac{\Delta G_2 - \Delta G_1}{RT}\right) \tag{5}$$

However, simulation of processes 1 and 2 is virtually impossible, since it would involve the removal of many solvent molecules from the binding site of the inhibitor on the enzyme to be substituted by the inhibitor in a reversible manner. But since scheme 4 is a cycle, we have:

$$\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \tag{6}$$

and if the composition of inhibitor I_B is not too different from that of I_A, the desired result can be obtained by simulating the non-chemical processes 3 and 4 (scheme 4) in a reversible manner. Thus, if we know the binding constant K_1, we can calculate the value of constant K_2.

Empirical free energy calculation

Inadequacies in potential energy functions and conformational sampling restrict the power of the above described method in modelling ligand-receptor complexes. This method is also computationally quite expensive, which limits its practical utility in many cases.

In some recent ligand-binding studies, attempts have been made to overcome certain limitations relating to the use of empirical free energy functions or free energy estimations. These approaches often estimate the free energy of ligand-receptor interactions as a function of hydrophobic contact surface area, number of hydrogen bonds, buried polar surface area, and similar terms (Böhm 1994, Head et al. 1996). As a result, these methods tend to be much less computationally demanding than methods involving potential energy functions and, unlike free energy perturbation methods, they are not dependent on traditional potential energy function or molecular dynamics configurational sampling. These empirical free energy function approaches may offer some promise for improved ligand-receptor binding free energy estimates, though they too are subject to inaccuracies in calibration of the functions.

Role of solvent in receptor-ligand interaction

The importance of solvent for the receptor-ligand interactions is obvious. As most biological processes take place in water, much effort is focused primarily on the understanding of interactions in this particular solvent. Special properties of water gave rise to the hydrophobic effect. Phenomenologically, it denotes the tendency of relatively apolar molecules to stick together in aqueous solution. The traditional explanation is that it results from the lowering of entropy due to the organisation of water around nonpolar solutes. Despite extensive research efforts, the qualitative as well as quantitative description and the origin of the driving force for hydrophobic interaction are still poorly understood. More detailed discussion of this topic is however far beyond the scope of this paper (Blokzijl and Engberts 1993).

Many authors prefer to speak of hydrophobicity as the consequence of the high cohesion of water. In this sense, all aqueous interfaces are "hydrophobic" in that there will always be some force acting to minimise the interfacial area. This is also the concept underlying the quantitative description of the hydrophobic interaction occurring e.g. between ligand and receptor. Ignoring the effects of translational, rotational, and configurational entropy, the process can be viewed as the association of two completely nonpolar surfaces. Then its free energy might be written in the form (Honig et al. 1993):

$$\Delta G^{np} = \gamma \Delta A \tag{7}$$

where ΔA is the accessible surface area buried in the reaction, and γ is the interfacial tension (surface free energy). γ is extracted from experiments in which nonpolar solutes are transferred from some organic phase to water (Sharp et al. 1991a). The differences between these measurements and macroscopic measurements of surface tension suggest that its value may depend on several other factors such as solute volume and the curvature of molecular surfaces (Sharp et al. 1991b).

An alternative approach of taking the effects of aqueous solvation into account is to include water molecules explicitly (Jorgensen et al. 1983) into the system that is consequently subjected to MD simulation, so that standard characteristics can be determined in the presence of water. This approach is inevitable when water does not only play the role as an environment but is directly involved in the particular interactions and basically forms the integral part of the solute. Considerable effort continues to be focused on developing polarizable models for water via its force field parametrization (Halgren 1995).

Yet another view on incorporating solvent effect into the model concentrates on hydrodynamic motions and properties of solvated macromolecules (Northrup 1994). Most studies in this area exploit the molecular dynamics simulations with the modified equations of motion known as Brownian dynamics (Ermak and McCammon 1978) that has been applied to a great variety of problems (Sharp et al. 1987; Veresov 1996).

The structure of the receptor unknown

If we do not have structural information about the receptor, we are basically looking for its hypothetical representation, which is designated by the term pharmacophore.

This term refers to a set of features (pharmacophoric groups) that are common to a series of active molecules. A three-dimensional pharmacophore specifies the spatial relationships between the groups. It is important to realise that most of the approaches to finding 3D pharmacophores assume that all of the molecules bind in a common manner to the macromolecule.

The biological activity is usually represented by an equipotent concentration (dose) after a fixed time interval of testing such as ID_{50}, LD_{50} and EC_{50} or by rate constants. Their logarithms are free energy related quantities.

Linear combination of parameters

The first successful attempt to quantitatively describe the relation between the structure and activity was the pioneering work of Hansch (Hansch and Fujita 1964) which is used in many modifications. The biological activity of i-th compound (ligand) E_i is expressed as a linear combination of compound's hydrophobic x_h, electronic x_e and steric x_s properties:

$$E_i = f_h(x_h) + f_e(x_e) + f_s(x_s) + const \quad i = 1, \ldots, n \tag{8}$$

We write n such equations for each compound in a series and knowing their biological activities and physicochemical parameters, we calculate constants comprised in this linear relationship with the use of the multiparameter linear regression analysis.

To overcome the differences between the behaviour of compounds in a biological system and in a model physicochemical system, we use the extrathermodynamic approach which justifies the use of physicochemical parameters obtained in a model system in the real biological system. Most physicochemical empirical parameters are actually substituent parameters as they are derived from the change of physicochemical properties caused by the substitution replacement of the basic structure.

In the Hansch approach, the parameter for the phase distribution relationship was standardised by $\log P$, the free energy related parameter for the hydrophobicity of molecules, P being the partition coefficient measured in the 1-octanol/water system. Hence, $\log P$ is used as the standard parameter for the hydrophobicity of drug molecules. The effect of the hydrophobicity of a series of drugs on biological activity is in general expressed by the parabolic function of $\log P$. This satisfactorily explains the tendency of the drug to remain in the first aqueous phase if its hydrophobicity is very low. On the other hand, if its hydrophobicity is very high, the drug is trapped in the first lipid phase in the transport process.

De novo approach

In the same time as the seminal papers of Hansch and coworkers have appeared, Free and Wilson (1964) published the mathematical method which allowed to find structure-activity relationships based only on the known biological activities. It assumed the additivity concept of biological activity contributions within congeneric series. Every substituent contributes to the overall activity by a certain constant

value. Biological activity is then expressed as a sum of contributions of individual substituents and the global constant which represents the contribution of the basic (usually nonsubstituted) part of the molecule. The fact that we do not have to know the physicochemical parameters of the substituents represents the main advantage of the Free-Wilson approach, though its applicability, in comparison to the Hansch approach, is more limited.

Minimal steric (topological) difference

This method (Simon et al. 1984) is based on the assumption that ligand-receptor interaction is a linearly decreasing function of the steric misfit of the ligand and the site receptor cavity, i.e. the activity is a function of the sum of the non-overlapping volumes of the ligand and the cavity. The term minimal steric difference originates from the consideration that if the ligand has several low energy conformations it will adopt the one which fits best into the receptor site cavity. A reliable approximation for the shape of the cavity is the most active structure in the set of compounds under study. This molecule is termed the standard. Then, the other molecules are superimposed on the standard. The number of non-superimposable atoms (hydrogen atoms are neglected) gives the MSD value for the considered structure.

The minimal topological difference method (MTD) modifies the previous technique. After defining the orientation of the standard molecule, all compounds are superimposed one upon another. The resulting hypermolecule is used as the standard to calculate the MTD values. Both MSD and MTD parameters can be used in place of steric parameter in the Hansch type equations (Eq. 8).

Molecular Shape Analysis

In this approach (Hopfinger 1983) the congeners in a dataset are first examined using molecular mechanics to determine the most stable conformers. A reference compound is selected against which the shape of all other congeners can be compared. The total common overlap volume of the reference structure and each of the congeners in the dataset is calculated. This value gives the estimate of the cavity volume occupied by the considered ligand. The related parameters such as the contact surface area of the ligand with the receptor can be calculated as well.

Three-dimensional molecular shape descriptors introduced by Motoc (Motoc et al. 1985) represent a very similar approach. The most active molecule in a series is again used as a template for the calculation of the overlapping and non-overlapping volumes for each compound. All these volumes can be again used as a steric measure parameter in Hansch type equations (Eq. 8).

Active Analog Approach

If one assumes that all molecules that have activity at the same target receptor must present their pharmacophoric groups in the same configuration in distance space, then one can search the conformational space of all the molecules to identify a convergent distance map which defines the pharmacophore geometry common to all molecules. This approach (Marshall and Motoc 1986) is particularly useful when

the molecules to be analysed represent a diverse set of structures rather than some homologous series varying only in their substituent groups. The active conformations of flexible compounds are determined via systematic conformational searches using geometrical constraints of a rigid template analog. By logically intersecting the set of allowed conformations, one can determine which patterns are common to all molecules. Here again, it must be assumed that neither the receptor structure nor the binding mode varies for the different molecules that are examined.

ALADDIN, DISCO

The qualitative 'receptor mapping' method developed by Martin et al. (1993) is aimed at the determination of the 3D requirements for a small molecule to exhibit a particular bioactivity (basically 3D pharmacophore). The strategy implemented in computer-automated procedures identifies both the bioactive conformation and the superposition rule for every active compound. First, all low-energy conformations are generated and optimized by any suitable method. Then the program ALADDIN calculates the location of points which may be considered for the superposition of the molecules for all low-energy conformers of a series of compounds. Such points are, for example, atoms, ring centres, and projections from the molecule to hydrogen bond donor, acceptor, and charged groups at the binding site. The molecule with the smallest number of possible conformations forms the template. These positions and the relative energy of each conformation are the input to the program DISCO. It uses a clique-detection method to find superpositions of the molecules that contain at least one conformation of each compound in the user-defined three-dimensional arrangement of site points. DISCO is typically run several times to compare alternative pharmacophore maps.

CoMFA

One of the most popular methods nowadays is CoMFA-Comparative Molecular Field Analysis (Cramer et al. 1988, 1993). First a group of compounds having a common pharmacophore is selected and three dimensional structures of the compounds are generated. The energy minimised structures are stored in a database and fitted to each other according to their chemical similarity by using a pharmacophore hypothesis and postulating orientation rules. Once the molecules are aligned, a grid or lattice is established which surrounds the set of analogs in potential receptor space. The fields which a certain probe atom would experience at every grid point are calculated for each molecule, leading to thousands of columns that are correlated with a biological activity. For the steric field, Lennard-Jones potential is used, the electrostatic field is calculated from Coulomb potential. Normally, the steric and electrostatic fields are kept separate for ease of interpretation of the results. Grid points without variance (e.g. inside the volume shared by all molecules) or with small variance (e.g. in the corners of the box, far away from the Van der Waals spheres of the molecules) are eliminated. The last step in a CoMFA study is a partial least squares (PLS) analysis to determine the minimal set of grid points which is necessary to explain the biological activities of the compounds.

Most often the results of a CoMFA study are presented in graphical form, with contours for favourable and unfavourable regions of the different fields. The use of additional fields to the default steric and electronic fields (e.g. GRID) is quite common as a valuable extension of the CoMFA program.

Distance geometry

Perhaps the most detailed and versatile procedure for mapping receptor sites is the distance geometry method developed by Crippen and Ghose (Crippen and Havel 1988; Ghose and Crippen 1990). The computerised treatment consists of a sophisticated procedure which combines information concerning the conformational possibilities of the ligand with the corresponding observed receptor binding affinities.

In distance geometry the structure of a ligand is represented by a distance matrix. Flexibility of a molecule is expressed by a distance range matrix showing the upper and lower bounds on the distance between each atom pair. Distance geometry modelling of the structure of the active site cavity is based on a very simple idea. Suppose we have two flexible ligand molecules m and n, and the atoms m_i and m_j of molecule m and atoms n_i and n_j of molecule n occupy the same respective regions of the active site. The distance between the ith and jth atoms in the two molecules must be very close in their active conformations (the conformations in which they bind with the receptor). Since in the distance geometry representation of the flexible molecules atomic distances have ranges, the active conformations should be represented by a common distance range. If we have several molecules, such comparisons will gradually decrease the range, thereby contracting the possible conformational region. Ultimately, embedding these distances will give the three-dimensional structure of the site pockets accommodating the ligand atoms. The problem is we do not know how to determine 'equivalent' ligand atoms at the receptor site, as the explicit structure of the receptor is not known. Therefore we have to make a hypothesis regarding the binding mode of the ligand at the receptor site. A computerised search can be used to make a good hypothesis at this stage.

The resulting set of geometrically feasible binding modes is then tested according to the experimental binding affinities. The values of receptor-ligand binding affinities are modelled via physicochemical parameters of atoms (or molecular fragments). Most often, these three properties are used: octanol-water partition coefficient, atomic refractivity and formal charge density. Each physicochemical property represents a particular type of force active in the ligand receptor interaction. When an atom of the ligand enters a site pocket, it interacts with the site and the interaction energy is a function of one or more physicochemical properties of the atom, where the nature of the function is determined by the characteristics of the active site. The binding energy of a particular binding mode is given by (Ghose and Crippen 1985):

$$E^{calc} = -CE^c + \sum_{i=1}^{n_s} \sum_{j=1}^{n_p} \left(C_{i',j} \sum_{k=1}^{n_o} P_j(t_k) \right) \tag{9}$$

where E^c is the energy of the conformation under consideration; $C_{i',j}$ are coefficients characterising the site type and the physicochemical property (this is what we want to determine by some optimisation technique); i' is the type of the site i; n_s represents the number of site pockets; n_p represents the number of physicochemical properties to be correlated with the site; n_o represents the number of atoms occupying the site pocket; and $P_j(t_k)$ represents the jth physicochemical parameter of the occupied atom of type t_k.

We usually get more than just one geometrically feasible binding mode, each representing different expression for the calculated binding energy (Eq. 9). Only one of these binding modes having the most favourable binding energy is selected. Simultaneously, we require that the calculated and experimentally observed binding energies are as close as possible. Thus, coefficients $C_{i',j}$ are evaluated using constrained least squares technique (quadratic programming) which is mathematically formulated as follows:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{m} (E_p^{calc} - E^{obsd})^2 \\ &\text{such that } E_p^{calc} > E_q^{calc} \end{aligned} \tag{10}$$

where m is the number of molecules, and indices p, q go through all geometrically feasible binding modes for a given ligand.

Given the interaction coefficients ($C_{i',j}$ in Eq. 9), geometry of the site cavity and the structure of a molecule, one can easily calculate its binding energy from the geometrically feasible binding modes. The actual binding energy will correspond to the binding mode yielding highest binding energy.

An example of applications of computational methods

The first example concerns the development of the modified scheme of classical QSAR method originally introduced by Free and Wilson (1964). Most QSAR methods for quantitative description of ligand binding to receptor implicitly assume that similar ligands bind in the same orientation to an identical portion of the binding site. Multiple binding modes, once being regarded as rare exceptions, have been observed however in quite a few cases (De La Paz et al. 1992; Mattos and Ringe 1993). Obviously, this has to be taken into account if we are building a model for a prediction of ligand binding affinities. The basic contribution of this model consists in considering multiple embedding of the individual ligands in the binding site, which are involved in the final structure-activity equation via summation of nonlinear contributions. The model was tested on a series of polychlorinated biphenyls and resulted in a better explanation of activities than previous approaches (Horňák et al. 1998a,b). The application of this method on the set of polychlorinated dibenzofurans is under way.

The second example demonstrates the use of molecular dynamics techniques. We tried to investigate the importance of selected groups of 3'-GMP and 2'-GMP

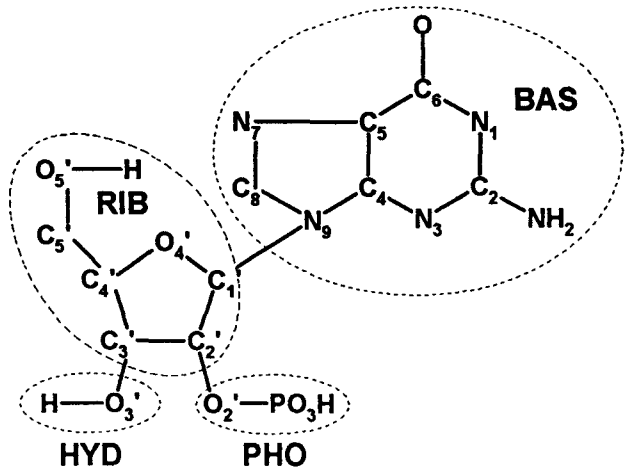

Figure 1. The structural representation of 2'-GMP. The groups selected for interaction energy calculations (HYD, PHO, RIB, BAS) are outlined by dashed lines. The only difference compared to 3'-GMP is the swapped position of HYD and PHO groups.

(Fig. 1) on binding to RNase Sa. Molecular dynamics (Brooks et al. 1983) was used as the modelling tool for this purpose. The known structures of RNase Sa with 3'-GMP (s33s) and 2'-GMP (s22s) were used as a starting point (Sevcik et al. 1991, 1993). Abreviated names of systems under study are given in brackets. There are minor differences in crystal structures of s22s and s33s complexes. Most notable are differing orientations of two aminoacids – Arg40 and Gln32 in the vicinity of the binding site. To take these differences into account, we tested two more systems obtained by changing the ligands in RNase Sa, putting 2'-GMP in place of 3'-GMP from the first complex (giving s32s) and analogously 3'-GMP into the second complex instead of 2'-GMP (giving s32s). All four systems were initially minimised with the protein structure positionally fixed and then a dymamics run was performed with production phase for 100ps. Systems were coupled to an external bath at a constant temperature of 300K (Berendsen et al. 1984).

According to kinetic measurements (Both et al. 1982), both ligands bind specifically to RNase Sa, 2'-GMP having slightly lower binding affinity than 3'-GMP. The simple simulations that we performed cannot quantitatively explain such differences in binding because many more complex phenomena should be involved, such as the influence of pH, solvent and related entropic effects, which represent the problems still under active investigation. We focused on qualitative study of interaction contributions of the following parts of the ligands: nucleotide base (BAS), ribose ring (RIB), phosphate group (PHO) and hydroxyl group (HYD) in positions 2' and 3' (Fig. 1). Interaction energies of these groups fluctuated around equilibrium values, which emphasised the importance of BAS and PHO groups to the overall binding

Table 1. Interaction energies of individual groups from molecular dynamics simulations for all systems.

Group	Energy(kJ mol^{-1})	System			
		s33s	s32s	s23s	s22s
HYD	average	1.97	0.84	−15.51	−4.14
	STD	0.51	0.65	2.91	0.72
PHO	average	−85.87	−82.04	−90.68	−93.55
	STD	2.77	3.60	2.86	3.11
RIB	average	−3.04	−3.38	−15.18	−19.04
	STD	1.11	1.38	4.46	1.29
BAS	average	−64.90	−62.79	−70.24	−70.77
	STD	1.90	2.41	2.26	2.03
total	average	−151.84	−147.37	−191.61	−187.50
	STD	3.19	4.02	6.23	4.00

Averaged values through trajectory along with their standard deviations are given.

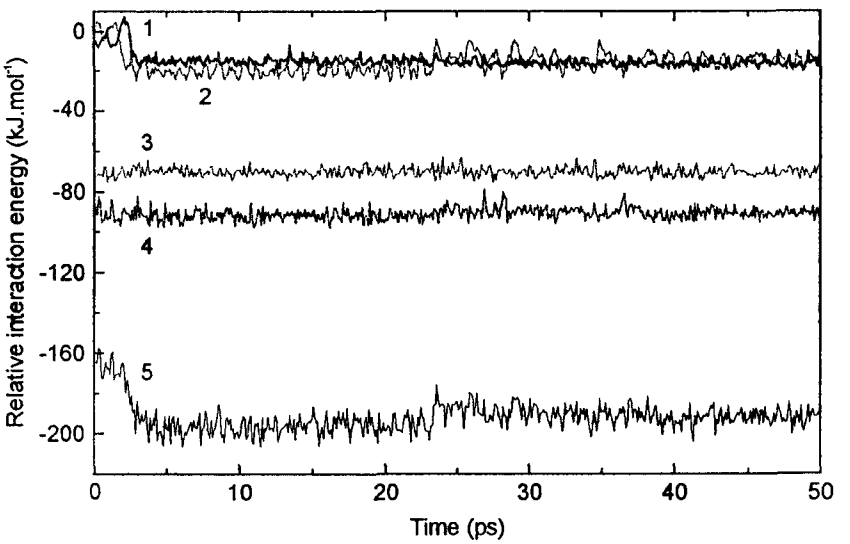

Figure 2. Time dependencies of individual interaction energies ordered from top to bottom: HYD (1, bold), RIB (2), BAS (3), PHO (4), and total interaction energy (5) with RNase Sa. Only the first half of the trajectory is shown in order to make energetical changes at the beginning of the trajectory clearer.

(Table 1). As an example, we plotted the time dependencies of interaction energies for the system s23s which showed most significant deviations (Fig. 2). They were accounted for by the closer examination of the trajectory which shows that in the interval 20–30ps the O5' in RIB group and O2' from HYD group formed hydrogen bonds with Arg40 and Tyr86 respectively, thus changing the overall interaction energy. Though both these groups contribute only about 15% to the overall interaction, their conformational reconfiguration caused a notable decrease of total interaction energy at the beginning of simulation, which is visible in Fig. 2.

Based on the reported simulation we can deduce that phosphate group and guanine base play a prominent role in the binding of the substrate in the active site of RNase Sa though concerted contributions of the other parts can influence the overall interaction energy.

Conclusion

The number of theoretical methods reviewed in this paper were developed for the study of receptor-ligand interactions during last years. They indicate the growing significance of theoretical modelling for the understanding of molecular mechanisms of interaction and prediction of biological activity. On the other hand, it is important to realise the capabilities and limitations of the individual methods, which were briefly demonstrated on two sample cases While classical QSAR methods were developed having their practical application in mind, typical simulation techniques mostly provide only qualitative analysis, providing more insight into how things work on the molecular level. On the whole, all these approaches have a great potential for being more and more commonly used in the future.

Acknowledgements. This work was supported by Slovak Grant Agency (grant No 1/4201/97)

References

Bamborough P , Cohen F E (1996) Modeling protein-ligand complexes Curr Opin Struct Biol **6**, 236—241

Berendsen H J C , Postma P M , van Gunsteren W F , Di Nola A , Haak J R (1984) Molecular-dynamics with coupling to an external bath J Chem Phys **81**, 3684—3690

Blaney J M , Jorgensen E C , Connolly M L , Ferrin T E , Langridge R , Oatley S J , Burridge J M , Blake C C F (1982) Computer graphics in drug design Molecular modeling of thyroid hormone-prealbumin interactions J Med Chem **25**, 785—790

Blokzijl W , Engberts J (1993) Hydrophobic effects - opinions and facts Angew Chem Int Ed Engl **32**, 1545—1579

Bohm H J (1992a) The computer program LUDI A new simple method for the de-novo design of enzyme inhibitors J Comput Aid Molec Design **6**, 61—78

Bohm H J (1992b) LUDI Rule based automatic design of new substituents for enzyme inhibitor leads J Comput Aid Molec Design **6**, 593—606

Bohm H J (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure J Comput Aid Molec Design **8**, 243—256

Both V , Witzel H H , Zelinková E (1982) Chemical modifications and kinetic study of ribonuclease Sa active site Gen Physiol Biophys **1**, 261—268

Brooks B R , Bruccoleri R E , Olafson B D , States D J , Swaminathan S , Karplus M (1983) CHARMM a program for macromolecular energy minimization and dynamics calculations J Comput Chem **4**, 187—217

Cramer R D , Patterson D E , Bunce J D (1988) Comparative molecular field analysis (CoMFA) 1 Effect of shape on binding of steroids to carrier proteins J Amer Chem Soc **110**, 5959—5967

Cramer R D , DePriest S A , Patterson D E , Hecht P (1993) The developing practice of comparative molecular field analysis In 3D-QSAR in Drug Design Theory, Methods and Applications (Ed H Kubinyi) pp 443—486, Escom, Leiden

Crippen G M , Havel T F (1988) Distance Geometry and Molecular Conformation pp 361—431, Research Study Press (Wiley), New York

De La Paz P , Burridge J M , Oatley S J , Blake C C F (1992) Multiple modes of binding of thyroid hormones and other iodothyronines to human plasma transthyretin In The Design of Drugs to Macromolecular Targets (Ed C R Beddell) pp 119—172, John Wiley and Sons, Chichester

Dean P M (1987) Molecular Foundations of Drug-Receptor Interaction pp 1—34, Cambridge University Press, Cambridge

Ermak D L , McCammon J A (1978) Brownian dynamics with hydrodynamic interactions J Chem Phys **69**, 1352—1360

Free S M , Wilson J W (1964) A mathematical contribution to structure activity studies J Med Chem **7**, 395—399

Ghose A K , Crippen G M (1985) Use of physicochemical parameters and related three-dimensional quantitative structure-activity relationships A demonstration using *Escherichia coli* dihydrofolate reductase inhibitors J Med Chem **28**, 333—346

Ghose A K , Crippen G M (1990) The distance geometry approach to modelling receptor sites In Comprehensive Medicinal Chemistry, Vol 4 (Ed C Hansch) pp 715—733, Pergamon Press, Oxford

Goodford P J (1985) A computational procedure for determining energetically favourable binding sites on biologically important macromolecules J Med Chem **28**, 849—857

Grant G H , Richards W G (1995) Computational Chemistry Oxford University Press, Oxford

Halgren T A (1995) Potential energy functions Curr Opin Struct Biol **5**, 205—210

Hansch C , Fujita T (1964) ρ-σ-π Analysis A method for the correlation of biological activity and chemical structure J Amer Chem Soc **86**, 1616—1626

Head R D , Smythe M L , Oprea T I , Waller C L , Green S M , Marshall G R (1996) VALIDATE A new method for the receptor-based prediction of binding affinities of novel ligands J Amer Chem Soc **118**, 3959—3969

Hianik T , Sargent D F , Smriga M , Šikurová L , Nemcová P (1996) Interaction of adrenocorticotropin-(1-24)-tetracosapeptide with lipid bilayers Gen Physiol Biophys **15**, 239—250

Honig B , Sharp K , Yang A (1993) Macroscopic models of aqueous solutions biological and chemical applications J Phys Chem **97**, 1101—1109

Hopfinger A J (1983) Theory and application of molecular potential energy fields in molecular shape analysis J Med Chem **26**, 990—996

Horňák V , Baláž Š , Májeková M (1998a) Binding modes of PCBs to a degrading enzyme A receptor-mapping study Gen Physiol Biophys **17**, 105—116

Horňák V , Baláž Š , Schaper K J , Seydel J K (1998b) Multiple binding modes in 3D-QSAR microbial degradation of polychlorinated biphenyls Quant -Struct Act Relat **17**, 427—436

Jorgensen W L , Chandrasekhar J , Madura J D , Impey R W , Klein M L (1983) Comparison of simple potential functions for simulating liquid water J Chem Phys **79**, 926—935

Kollman P A (1994) Theory of macromolecule-ligand interactions Curr Opin Struct Biol **4**, 240—245

Kubinyi H (1993) QSAR Hansch Analysis and Related Approaches VCH, Weinheim

Kuntz I D , Blaney J M , Oatley S J , Ferr in T E (1982) A geometric approach to macromolecule-ligand interaction J Mol Biol **161**, 269—288

Lauri G , Bartlett P A (1994) CAVEAT - A program to facilitate the design of organic molecules J Comput Aid Molec Design **8**, 51—66

Leach A R (1996) Molecular Modelling Principles and Applications Addison Wesley Longman Limited, London

Marshall G R , Motoc I (1986) Approaches to the conformation of the drug bound to the receptor In Molecular Graphics and Drug Design (Eds A S V Burgen, G C K Roberts, M S Tute) pp 116—152, Elsevier, Amsterdam

Martin Y C (1978) Quantitative Drug Design A Critical Introduction Marcel Dekker Inc , New York

Martin Y C , Bures M G , Danaher E A , DeLazzer J , Lico I , Pavlik P A (1993) A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists J Comput Aid Molec Design **7**, 83—102

Mattos C , Ringe D (1993) Multiple binding modes In 3D-QSAR in Drug Design Theory, Methods and Applications (Ed H Kubinyi) pp 226—254, Escom, Leiden

Miranker A , Karplus M (1991) Functionality maps of binding sites A multiple copy simultaneous search method Protein Struct Funct Genet **11**, 29—34

Moon J B , Howe W J (1991) Computer design of bioactive molecules - A method for receptor-base denovo ligand design Prot Struct Func Genet **11**, 314—328

Motoc I , Marshall G R , Dammkoehler R A , Labanowski J (1985) Molecular shape descriptors three dimensional molecular shape descriptors Z Naturforsch **40**, 1108—1113

Northrup S H (1994) Hydrodynamic motions of large molecules Curr Opin Struct Biol **4**, 269—274

Oshiro C M , Kuntz I D , Dixon J S (1995) Flexible ligand docking using a genetic algorithm J Comput Aid Molec Design **9**, 113—130

Pearson N D , Prescott C D (1997) RNA as a drug target Chem Biol **4**, 409—414

Seshadri K , Rao V S R , Vishveshwara S (1995) Interaction of substrate uridyl 3',5'-adenosine with ribonuclease A A molecular dynamics study Biophys J **69**, 2185—2194

Sevcik J , Dodson E J , Dodson G G (1991) Determination and restrained least-squares refinement of the structures of ribonuclease Sa and its complex with 3'-guanylic acid at 1 8-Åresolution Acta Cryst D **47**, 240—253

Sevcik J , Hill C P , Dauter Z , Wilson K S (1993) Complex of ribonuclease from *Streptomyces aureofaciens* with 2'-GMP at 1 7-angstrom resolution Acta Cryst D **49**, 257—271

Sharp K , Fine R , Honig B (1987) Computer simulations of the diffusion of a substrate to an active site of an enzyme Science **236**, 1460—1463

Sharp K. A., Honig B. (1990): Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. J. Phys. Chem. **94**, 7684—7692

Sharp K. A., Nicholls A., Fine R. F., Honig B. (1991a): Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science **252**, 106—109

Sharp K. A., Nicholls A., Friedman R., Honig B. (1991b): Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. Biochemistry USA **30**, 9686—9697

Simon Z., Chiriac A., Holban S., Ciubotaru D., Mihalas G. I.. (1984): Minimum Steric Difference. Research Studies Press Ltd., Letchworth, England, pp. 1—95

Van Gunsteren W. F., King P. M., Mark A. E. (1994): Fundamentals of drug design from a biophysical viewpoint. Quart. Rev. Biophys. **27**, 435—481

Veresov V. G. (1996): Brownian dynamics simulation of pH-effects on conductance through potassium channels. Gen. Physiol. Biophys. **15**, 389—402

Final version accepted accepted August 15, 1999