

**A New Method for Ligand-supported Homology
Modelling of Protein Binding Sites:
Development and Application to the neurokinin-1
receptor**

Dissertation

zur

Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem

Fachbereich Pharmazie
der PHILIPPS-UNIVERSITÄT MARBURG
vorgelegt von

Andreas Evers

aus Köln

Marburg/Lahn 2003

Vom Fachbereich Pharmazie der Philipps-Universität Marburg
als Dissertation angenommen am:

08. Oktober 2003

Erstgutachter:

Prof. Dr. G. KLEBE

Zweitgutachter:

Prof. Dr. O. ESSEN

Tag der mündlichen Prüfung:

12. November 2003

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. G. KLEBE am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Juni 2000 bis Juli 2003 durchgeführt.

Contents

1	Introduction	1
1.1	Structure-based drug design	1
1.2	Scope of this thesis	5
2	Literature survey	7
2.1	Prediction of protein structure	7
2.1.1	Basic principles of protein structure	8
2.1.2	Approaches for predicting protein structure	9
2.1.3	Evaluation of protein structures	14
2.2	Prediction of protein-ligand interactions	15
2.2.1	Ligand docking	15
2.2.2	Affinity prediction	20
2.2.3	Approaches combining protein and ligand-based procedures . . .	25
2.3	Homology modelling and drug design	28
3	Development of an approach for Ligand-supported Homology Modelling of Protein Binding Sites using Knowledge-based potentials	30
3.1	Strategy and computational realisation	30
3.1.1	General overview	30
3.1.2	Step 1: Generation of preliminary protein models	30
3.1.3	Step 2: Placing the ligand(s) into the homology models	31
3.1.4	Step 3a: Incorporating ligand information into the homology modelling process	32
3.1.5	Step 3b: Scoring the generated models	33
3.1.6	Step 4: Optimising and refining the homology models	33
3.2	Results and Discussion	35
3.2.1	Analysis of the generated binding-site models	35
3.2.2	Modelling case studies	46
3.3	Summary and Outlook	54
3.4	Materials and Methods	58
3.4.1	Test data set	58
3.4.2	Generation of binding-site models of the test data set	58

3.4.3	Scaling the DrugScore potentials with respect to the MOD- ELLER force-field	60
4	Generation and validation of a ligand-supported homology model of the neurokinin-1 receptor by virtual screening for a submicromolar inhibitor	61
4.1	Biological target system	61
4.1.1	G-protein-coupled receptors	61
4.1.2	The Neurokinin Receptors	71
4.1.3	NK1 antagonists	72
4.2	Modelling the neurokinin-1 receptor	79
4.2.1	Generation of protein-ligand complexes using the <i>MOBILE</i> ap- proach	79
4.2.2	Analysis of the model	83
4.2.3	Generation of a protein- and ligand-based pharmacophore . . .	87
4.2.4	Virtual Screening	89
4.2.5	Testing for binding	93
4.2.6	Discussion	95
4.3	Modelling the neurokinin-2 receptor	99
4.3.1	Generation of protein-ligand complexes using the <i>MOBILE</i> ap- proach	99
4.3.2	Analysis of the model	104
4.3.3	Discussion	107
5	Summary	110
5.1	Summary	110
5.2	Zusammenfassung	114
	Bibliography	119

1 Introduction

1.1 Structure-based drug design

The genome sequencing projects provide us with an increasing number of fully assigned genomes, including those of humans and vertebrates (e.g. mouse) and important microbial pathogens [Marcotte et al., 1999; Broder & Venter, 2000; Lander et al., 2001; Rubin et al., 2000]. The total number of protein sequences in the Swiss-Prot database [Boeckmann et al., 2003] is about 122600 as of February 2003. In contrast, the number of protein structures deposited in the PDB is 21248 for the same date. Although the rate of experimental structure determination will continue to increase, the number of newly discovered sequences grows much faster than the number of structures solved (Fig. 1.1).

As a matter of fact, the number of experimentally determined novel protein folds has steadily decreased over the last years (see Fig. 1.2). Actually, 90 % of the protein structures solved today correspond to already known folds (Fig. 1.2).

Structural genomics is expected to yield a large number of experimentally determined protein structures, in the long run hopefully resulting in a complete coverage of protein folding space [Sanchez et al., 2000; Brenner, 2000; Brenner & Levitt, 2000; Holm & Sander, 1996]. Thus, referring to suitable reference structures, well spread in sequence and folding space, it will become increasingly possible to generate realistic models for any given protein sequence using comparative modelling techniques [Marti-Renom et al., 2000]. This technique can be considered sufficiently mature, given that there has been only a marginal improvement in the comparative modelling results while going from CASP3 to CASP4 [Tramontano et al., 2001]. Its rate of success depends on the degree of sequence identity with the template structure (as a rule of thumb at least 30 %) and the reliability of the underlying sequence alignment [Moult et al., 1999; Venclovas et al., 2001]. As an advantage, the regions predicted best are often the biologically important ones, [Wei et al., 1999], because they are structurally most conserved by evolution. This provides the perspective that structural genomics will support biology and medicine through the annotation of protein function [Thornton et al., 1999, 2000; Skolnick et al., 2000; Andrade et al., 1999]. Furthermore, the generated protein struc-

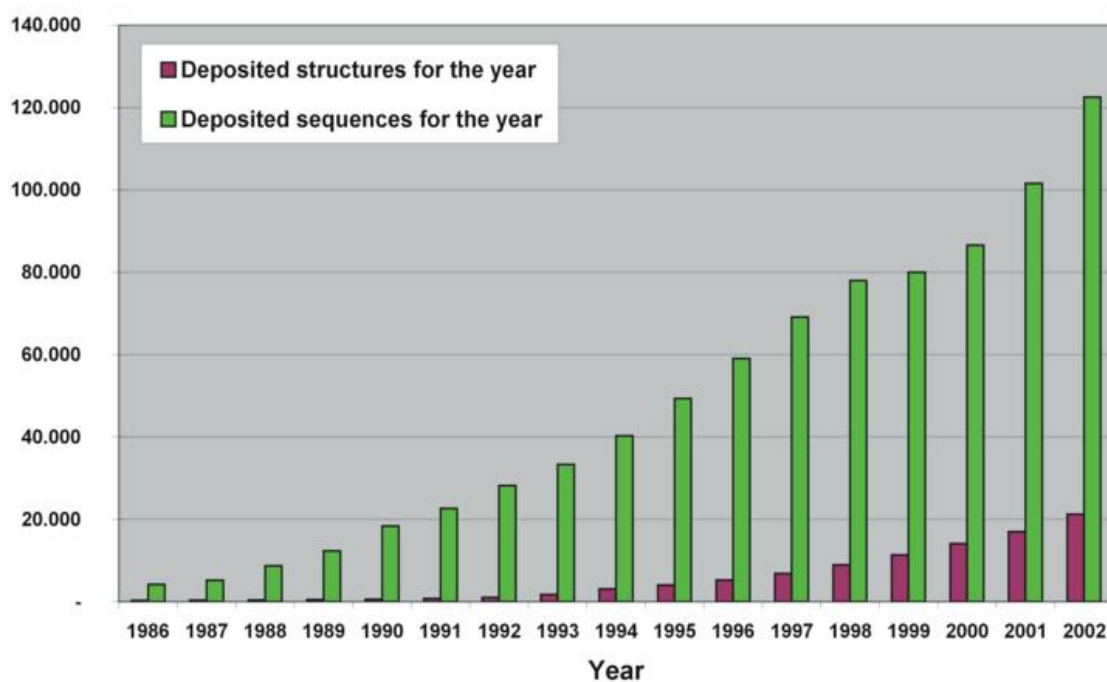


Figure 1.1: Sequence - Structure gap. The number of protein sequences in SWISS-PROT and protein structures in the PDB are plotted as annual distribution. Data from <http://us.expasy.org/sprot/relnotes/index.html> and http://pdb.protein.osaka-u.ac.jp/pdb/holdings_table.html.

tural models can be used for virtual screening to discover potential new lead structures for drug therapy [Russell & Eggleston, 2000].

For the rational structure-based design of drugs, knowledge of the three-dimensional structure of the target protein is indeed inevitably required. The approach followed here starts from a known or hypothetical binding mechanism. A lead structure is rationally designed and subsequently tested experimentally. The obtained results are fed back into a design cycle as new information.

However, if the structure of a target protein is not available, currently two distinct concepts are followed to discover novel drugs. On the one hand side, it is possible to establish a 3D QSAR (quantitative structure-activity relationship) model on the basis of a set of known ligands to extract 3D features primarily explaining observed trends in binding affinity. On the other hand, a homology model of the target protein can be constructed and subsequently used for the search of novel ligands, e.g. by virtual screening. While the QSAR approach is solely based on features derived from the ligands, the second approach only considers information available from the related proteins. The latter procedure is in fact rather approximate, especially if in the query

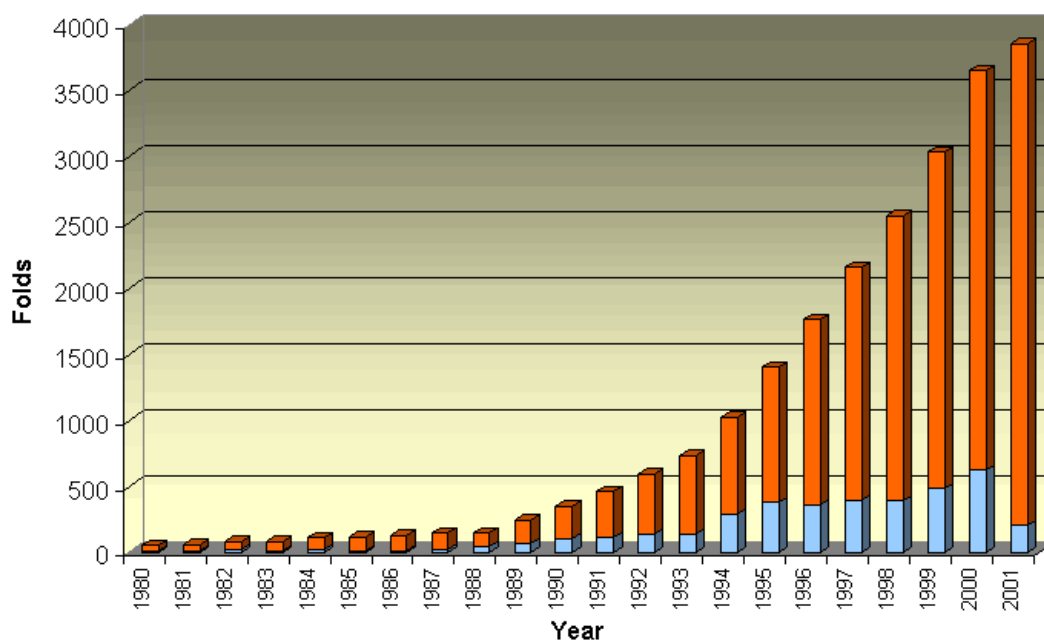


Figure 1.2: The number of protein structures solved versus the identification of new folds. In blue the number of novel folds is given, in red the number of newly solved structures. From <http://www.rcsb.org/pdb/holdings.html>.

protein several amino acids in the active site are replaced with respect to the references. Although ligands binding to a particular protein provide implicit information about the complementary features required at the protein active site, to the best of our knowledge, there is so far no approach which considers ligand information explicitly already during the comparative modelling step.

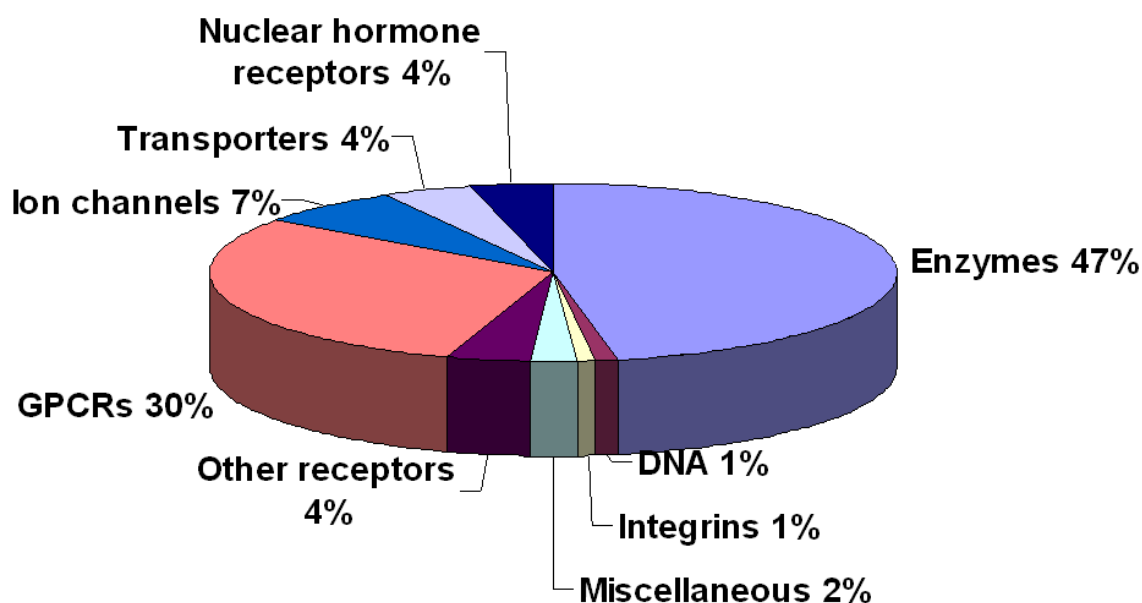


Figure 1.3: Targets of currently marketed small-molecule drugs, subdivided by biochemical class. Data from [Hopkins & Groon, 2002].

Fig. 1.3 gives an overview over currently launched drugs with respect to the addressed targets grouped into several biochemical classes. Enzymes represent nearly half of the total drug targets (47 %), whereas G-protein coupled receptors account for 30 %. Considering that for none of the G-protein coupled receptors (except for bovine rhodopsin) the 3D-structure is experimentally determined and that for most ion channels and a considerable amount of the enzymes the 3D-structure is not available, it becomes obvious that drug design is indeed frequently faced with the situation that a ligand has to be discovered for a target protein for which no experimentally determined structure is yet available. A recent estimate about the number of expected new molecular targets

in the post-genomic era indicates that this overall distribution will virtually remain [Hopkins & Groon, 2002].

Thus, to take advantage of the principles of structure-based drug design, methods are required which generate (more) realistic models of protein binding sites.

1.2 Scope of this thesis

The present work describes in detail how data about related proteins can be complemented with information about the binding modes of bioactive ligands to generate more realistic homology models of protein binding-sites. As a prerequisite for this approach, it is assumed that (1) information about ligands binding to the target protein is available and (2) that the 3D structures of related proteins with significant sequence identity are known.

Figure 1.4 provides a schematic overview of the strategy. It was initiated by the development of the DragHome concept [Schafferhans & Klebe, 2001] and is followed by the *MOBILE* approach (Modelling Binding Sites Including Ligand Information Explicitly).

Starting with the (crystal) structure of one or more template proteins, several preliminary homology models of the target protein are generated (step 1). After placing one or more ligands, known to bind to the target protein, into an averaged binding-site representation of the generated binding-site models (step 2), new protein models are generated, now considering explicitly the docked ligand(s) (step 3). After scoring the thus generated complexes with DrugScore, a final model is obtained by selecting the model which explains best the observed ligand binding (affinities). The modelled complexes can be further refined considering a composite picture of the best side-chain conformers taken from different models and minimising the side-chain-to-ligand interactions using a common force-field (step 4).

In chapter 4 of this work, the presented *MOBILE* approach is applied to a relevant real-life test example. Based on the crystal structure of bovine rhodopsin a homology model of the neurokinin-1 receptor, a G-protein coupled receptor, is generated, under explicit consideration of known antagonists. This model is then successfully used for the search of new neurokinin-1 antagonists.

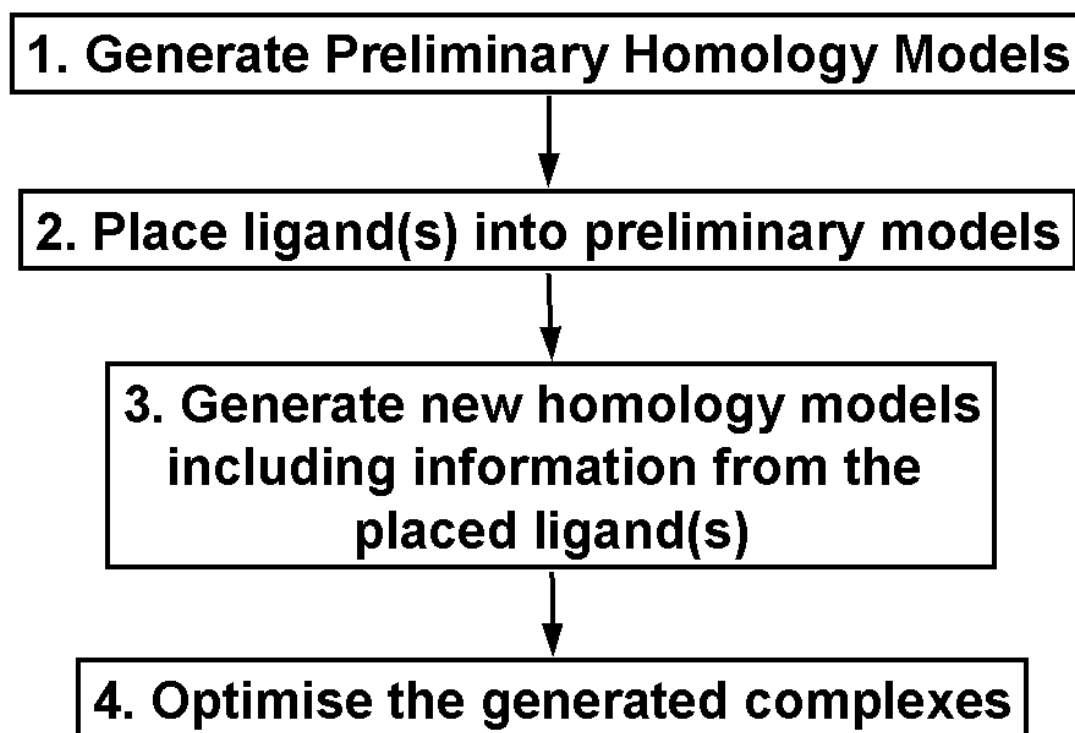


Figure 1.4: Schematic overview over the approach presented in this thesis: After generating preliminary models of the target protein (1), the ligand is docked into the superimposed ensemble of homology models (2). In the next step (3), new homology models of the target protein are generated under explicit consideration of the ligand in its docked orientation. Finally (4), the modelled complexes are further optimised by combining fragments from different models and subjecting the entire complex to an energy minimisation

2 Literature survey

As outlined in the introduction, the approach presented in this Thesis consists of methods related to the prediction and evaluation of protein structures and protein-ligand complexes. These issues are covered by homology modelling, docking and scoring. Further relevant issues essential for the understanding of the presented approach are the problem of protein flexibility (especially in the active site), strategies for ligand-based drug design (i.e. 3D-QSAR approaches) and methods which combine protein- and ligand-based approaches. This chapter gives a brief summary of currently available approaches and applications. A detailed description will be given for those methods which have actually been used in the context of this work.

2.1 Prediction of protein structure

A prerequisite for the understanding of protein-ligand interactions at the atomic level is the knowledge of the 3D structure of the target protein. Experimental 3D structures of proteins at atomic resolution are obtained by X-ray crystallography [Glusker et al., 1994; Drenth, 1999] and NMR [Wüthrich, 1986; Siegal et al., 1999; Clore & Gronenborn, 1991]. In these techniques, a model of the protein is constructed such that consistency with the experimental diffraction or NOE data is achieved. However, the rate at which protein sequences are determined exceeds far the rate at which protein structures are solved experimentally. Furthermore, many pharmacologically important targets are membrane-bound proteins, such as G-protein-coupled receptors (GPCRs), ion channels, or transporter proteins, for which the experimental determination of the 3D structure is, due to technical difficulties, either still impossible or can only be realised with immense experimental effort and complexity.

If the structure of a protein is not given, a theoretical approach for prediction of the structure based on the proteins sequence is required. This section provides a brief overview over the principles of protein structures and outlines strategies for their modelling and validation. A comprehensive introduction will be given for the program MODELLER, which has been used in this study for the generation of homology models.

2.1.1 Basic principles of protein structure

Proteins are polymers composed by the sequences based on twenty different amino acids which are connected by amide bonds. Although adopting unregular *tertiary structures*, common *secondary* and *supersecondary structure* motifs frequently occur in proteins. The most common motifs are the α -helix and the β -strand. These secondary structural elements are connected by loops which adopt less regular structures, such as β -turns [Wilmot & Thornton, 1988].

It has been demonstrated that the native tertiary structure of a given protein is determined solely by the protein's amino acid sequence [Anfinsen, 1973]. The native structure corresponds to the global minimum of free energy in the protein-solvent system (*thermodynamic hypothesis*). Levinthal showed that folding cannot be simulated or screened via a systematical search of all possible conformations because it would take "longer than the lifetime of a universe" [Levinthal, 1968]. The protein folding process has been investigated by Monte Carlo (MC) simulations on a 27-bead self-avoiding chain on a cubic lattice [Sali et al., 1994a,b; Dinner et al., 1996]. From these studies a three-stage process was suggested: Folding starts by a rapid ($\sim 10^4$ MC steps) collapse from a random-coil state to a random semi-compact globule. It then proceeds by a slow, rate-determining search ($\sim 10^7$ MC steps) through the ($\sim 10^{10}$) semi-compact states to find one of about 1000 similar transition states from which the chain rapidly ($\sim 10^5$ MC steps) folds to the native state.

The surface of water-soluble globular proteins preferentially consists of polar and charged amino acids while the interior is composed almost entirely of hydrophobic amino acids. The packing of these residues, which is almost perfect [Klapper, 1971; Tsai et al., 1999], is a consequence of the hydrophobic effect [Kauzmann, 1959], which is generally assumed to be of entropic origin. Water molecules around a non-polar solute form a cage-like ("iceberg") structure, which locally reduces the entropy [Frank & Evans, 1945]. When the non-polar residues of the unfolded protein associate, the water molecules are liberated, increasing the entropy, which is the dominating part of the overall free energy of folding [Tsai & Levitt, 1997].

2.1.2 Approaches for predicting protein structure

Protein structure prediction is divided into two basic areas, depending on the existence and similarity of homologous template proteins to the given target protein sequence:

- *ab initio* methods
- homology/comparative modelling techniques

However, the dividing line between these two approaches is diminishing and they are, to some extent, complementary to each other. Principally, threading methods [Bowie et al., 1991; Jones et al., 1992; Jones & Thornton, 1993], which cooperatively fit sequences onto known three-dimensional folds, could be considered as a combination of the two because threading samples sequences to known protein conformations in the PDB and evaluates them in terms of physical energies.

Progress in protein structure prediction methods is assessed by the *Critical Assessment of Protein Structure Prediction* (CASP) meetings [Moult et al., 1995, 1997, 1999, 2001]. In CASP, sequences of proteins whose experimental structures are soon to be released are made publically available. Computational research groups are then invited to predict 3D structures starting with the target sequence and any other publicly available information.

ab initio protein folding

Based on Anfinsen's afore-mentioned *thermodynamic hypothesis* it should be theoretically possible to calculate the 3D structure of a protein from its amino acid sequence by exact description of the physical environment within the cell and computing the molecular dynamics based on the underlying physical laws [van Gunsteren, 1998; Duan & Kollman, 1998].

Ab initio prediction methods require three elements: a representation of the protein geometry, a force field, and an energy surface searching technique [Osguthorpe, 2000]. Almost all methods use some kind of simplified geometry model, in which single virtual atoms represent a number of atoms in the all-atom model. The phase space to be searched is either continuous (*off-lattice models*, e.g. [Liwo et al., 1997a,b, 1998; Osguthorpe, 1997, 1999]) or discrete (*lattice models*, see [Godzik et al., 1993; Skolnick

et al., 1997; Ortiz et al., 1998; Hinds & Levitt, 1992]). The potential functions are usually statistical potentials which are either derived from knowledge-based approaches [Hinds & Levitt, 1994; Samudrala & Moulton, 1998] or use an underlying physical model parameterised at experimental data [Liwo et al., 1997a,b, 1998; Pillardy et al., 2001; Hao & Scheraga, 1999; Osguthorpe, 1997, 1999]. Standard search techniques such as genetic algorithms, Monte Carlo and simulated annealing are usually applied to explore the conformational space of proteins.

Compared to results from earlier CASP experiments, in CASP4 longer fragments (of up to 124 residues) of proteins were predicted within 6 Å deviation from the crystal structure.

The resolution of current *ab initio* structure prediction techniques may be sufficient for genome annotation; however, it is clearly not yet precise enough for detailed studies such as docking and drug design [Hardin et al., 2002].

Homology modelling

The most reliable technique for predicting protein structures is homology modelling, provided the geometry of one or more template proteins with sufficient sequence identity are given. If the sequence identity between template and target protein is high enough, the resulting model may even be sufficiently accurate to perform structure-based drug design.

In homology modelling, one or more template proteins with high sequence identity to the target sequence are identified. The target and template sequences are aligned, and a three-dimensional structure of the target protein is generated starting with the coordinates of the aligned residues of the template protein, combined with models for loop regions and other unaligned segments. The assembled 3D-model is then refined to bring it closer to the structure of the target protein.

Major difference between the various homology modelling techniques is how the 3D model is calculated from the sequence alignment [Sali, 1995b,a]. The original homology method is based on *rigid-body assembly* [Browne et al., 1969; Greer, 1981; Blundell et al., 1987]. The model is constructed from several core regions and from loops and side-chains, which are taken from related structures. The assembly involves fitting rigid bodies onto the framework, which is defined as the average of the C α atoms in the

conserved folding regions. A further group of methods, *modelling by segment matching*, uses approximate positions of conserved atoms from the template protein structures to calculate the coordinates of the remaining atoms [Jones & Thirup, 1986; Unger et al., 1989; Claessens et al., 1989; Levitt, 1992]. This is achieved using a database of short segments of protein structure, energy or geometry rules, or a combination of these criteria. The third group of methods, *modelling by satisfaction of spatial restraints*, uses either distance geometry [Havel & Snow, 1991; Srinivasan et al., 1993] or optimisation techniques [Sali & Blundell, 1993] to satisfy spatial restraints defined by the alignment of the target sequence to homologous template protein structures. In addition to the methods for modelling the entire fold, numerous other techniques for predicting loops [Fideslis et al., 1994; Lessel & Schomburg, 1999] and side-chains [Vasquez, 1996] with respect to a given backbone have also been described. These methods are often used in combination with each other.

The performance of the top eight comparative modelling groups at CASP4 was roughly similar [Tramontano et al., 2001; Marti-Renom et al., 2002]. The key element for successful model building is the quality of the sequence alignment, in particular in regions of low sequence identity. There is a limit for the alignment accuracy that can be achieved for distantly related proteins because their residual sequence similarity that guides the alignment becomes weaker with greater evolutionary distance [Schonbrun et al., 2002].

The accuracy of a comparative model is related to the percentage sequence identity on which it is based, it is thus correlated with the relationship between the structural and sequence similarity of two proteins (Fig. 2.1) [Marti-Renom et al., 2000; Sanchez & Sali, 1998; Koehl & Levitt, 1999]. Models based on more than 50 % sequence identity with their templates tend to have not more than 1 Å rms deviation with respect to the main-chain atoms. Medium-accuracy comparative models are based on 30 to 50 % sequence identity. They tend to have about 90 % of the main-chain modelled with 1.5 Å rms error.[Baker & Sali, 2001] As an advantage, the regions predicted best are often the biologically important ones [Bates et al., 2001], because they are structurally most conserved by evolution. This also allows for the application of these modelled regions for drug design purposes.

Homology modelling with MODELLER MODELLER belongs to the group of methods that model a proteins structure by *satisfaction of spatial restraints* [Sali &

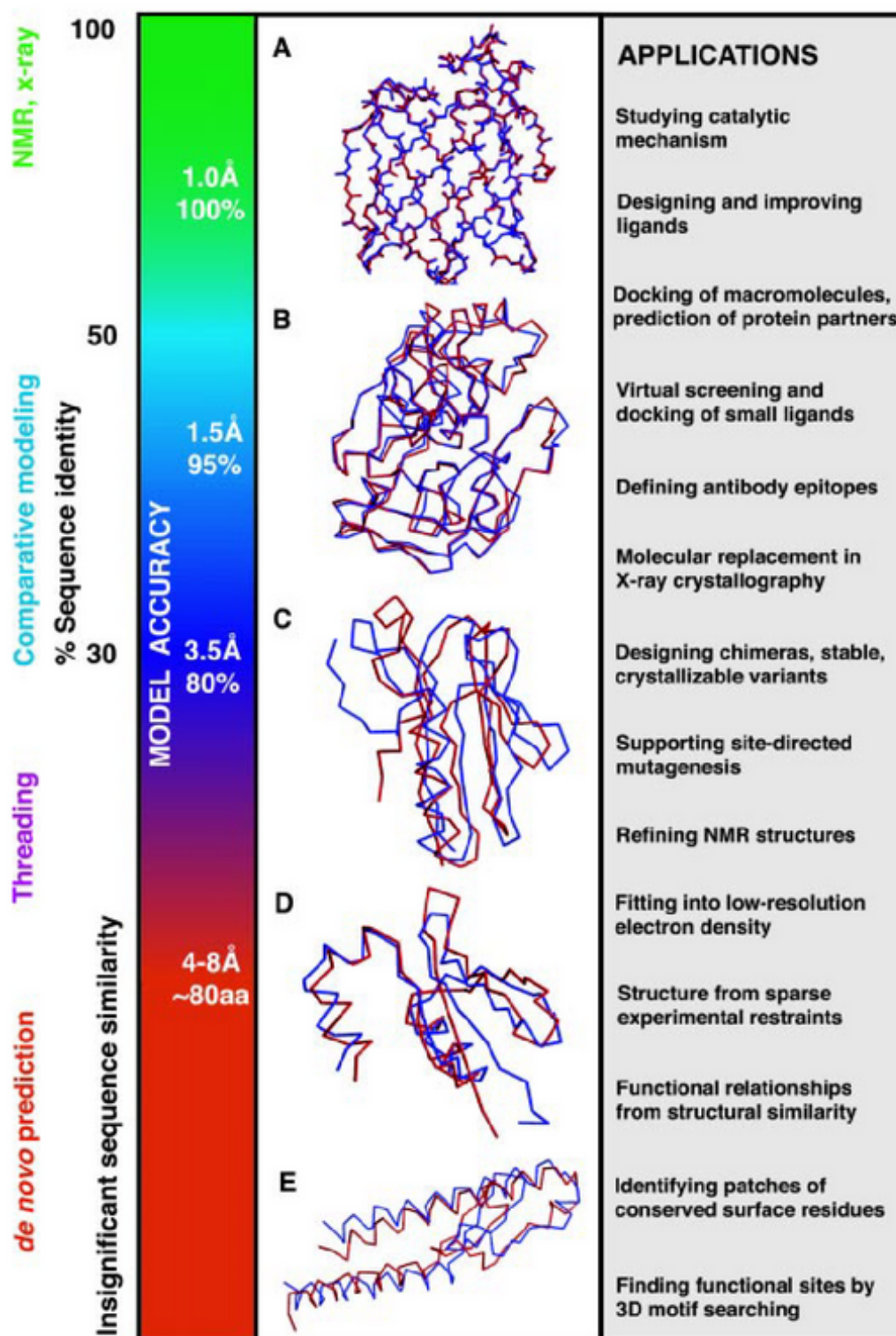


Figure 2.1: **Accuracy and application of protein structure models.** Shown are the different ranges of applicability of comparative protein structure modelling, threading, and de novo prediction; the corresponding accuracy of protein models; and their sample applications. From [Baker & Sali, 2001].

Blundell, 1993]. Protein modelling with MODELLER starts with a sequence alignment of the target protein with related known 3D structures. In the initial phase of the modelling process, restraints on the target protein are derived based on the sequence alignment with the template structures. Further restraints are derived from statistical analysis of the relationships between various general features of the protein structure, i.e. the distributions of distances between C α atoms, residue solvent accessibilities or side-chain torsion angles. These features are associated between the target and the homologous template protein(s). Thus, the conformation of a particular residue can be restrained according to the residue type, the conformation of an equivalent residue in a related protein, and the local similarity between the proteins. These restraints are expressed in terms of *conditional probability density functions (pdf)*. Each of these *pdfs* is a smooth function which describes the distribution of the considered feature as a function of the related variables. To enforce proper stereochemistry, these individual *pdfs* are combined with terms from the CHARMM force field [Brooks et al., 1983] (also expressed as *pdf*) to a molecular probability function:

$$F = F(R) = -\ln \left[\prod_i P_i(f_i, I_i) \right] = \sum_i E_i(f_i, a_i) \quad (2.1)$$

where $-\ln P_i = E_i$. R are the Cartesian coordinates of all atoms, P_i is a conditional probability density function (*pdf*) for a geometric feature f_i that depends on information I_i , which is derived from the template protein structure and associated restraints from analysis of various features of protein structures. E_i is an energy term, and a_i are general parameters related to I_i .

The model is obtained by optimising the objective function F by use of the variable target function method [Braun & Go, 1985], employing methods of conjugate gradients and molecular dynamics with simulated annealing [Clore et al., 1986]. Local restraints are considered first, global restraints subsequently.

A consequence of the *probability density functions* is that side-chains adopt similar conformations as in the template structure in regions which are conserved in the sequence alignment. By entering the optimisation with slightly different structures, models with differing structure can be obtained, which can reflect the conformational flexibility of the target protein in those regions which are different in the sequence alignment.

One major strength of MODELLER is that constraints or restraints derived from a

number of different sources can be added to the homology-derived restraints. For example, restraints could be provided by rules in terms of secondary structure packing, analyses of hydrophobicity, NMR experiments, etc. In this way, a homology model, especially in difficult cases, can be improved by making it consistent with available experimental data and with general knowledge about protein structures. Accordingly, it should in principle be possible to include ligand-based information into the homology modelling process. Within MODELLER, there are several possibilities to accomplish this task: for example, upper and lower boundaries for bond distances can be constrained between protein and ligand atoms or interactions can be approximated by van-der-Waals, Coulomb or H-bond potentials. Another possibility, which was followed in this study, is to approximate knowledge-based atom-pair potentials between protein and ligand atoms in terms of cubic splines to describe and regard their mutual influence in the homology modelling process.

2.1.3 Evaluation of protein structures

Once a protein model is constructed, it is important to assess its validity for the following modelling steps. Several approaches exist which evaluate the quality of a protein model. For the detection of correct protein models, an energy function is required, which efficiently identifies near-native structures. The principles of currently applied energy functions can be divided into physical and knowledge-based approaches [Wallner & Elofsson, 2003].

In *physical energy functions*, energy functions are used which were originally developed for 3D structure refinement and molecular dynamics simulations, such as CHARMM [Brooks et al., 1983] or Amber [Weiner et al., 1986]. Additionally, terms accounting for effects not considered implicitly in the energy function, are generally included into the scoring function. For example, Dominy *et al.* recently used the generalised Born solvation model [Still et al., 1990] to describe solvent effects [Dominy & Brooks, 2002].

Knowledge-based energy functions are calculated as the difference between features of a random protein model and observations from real protein structures. Most frequently, preferences for interacting protein residues are considered (e.g. [Jones et al., 1992; Sippl, 1993; Huang et al., 1995]), measuring the distances or simply counting the number of contacts. Some approaches are based on atom-type preferences [Colovos & Yeates, 1993; Samudrala & Moult, 1998; Melo & Feytmans, 1997, 1998; McConkey et al., 2003].

Another popular knowledge-based approach is Eisenberg's *3D profiles* method [Bowie et al., 1991; Luthy et al., 1992; Eisenberg et al., 1992]. It calculates for each amino acid in the proposed structure (1) the total surface area of a residue that becomes buried in the protein, (2) the fraction of the side-chain area that is covered by polar atoms and (3) the local secondary structure. These three parameters are then used to allocate the residue to one of eighteen environment classes. Each amino acid is assigned a score that reflects its compatibility with the amino acids in the local environment.

Finally, analysis of protein models is often performed by examining the structure and analysing any significant deviation from the norm, e.g. by calculating a Ramachandran plot. Another basic test is to use Procheck [Laskowski et al., 1993] or What.Check [Hooft et al., 1996] which assess the stereochemical quality of a protein structure. They analyze how common, or, conversely, how unusual the geometry of the residues in a given protein structure is, as compared with stereochemical parameters derived from well-refined, high-resolution crystal structures. The program PROVE (PROtein Volume Evaluation) [Pontius et al., 1996] measures deviations from standard atomic volumes as a figure-of-merit for protein structures.

2.2 Prediction of protein-ligand interactions

Two aspects are of utmost importance for successful computer-aided structure-based drug design: generating near-native protein-ligand configurations (*docking*), the identification of those binding modes that agree best to the experimentally given situation, and a computational translation of the obtained protein-ligand geometries into approximate estimates of the binding affinity (*scoring*). These issues will be addressed in the following, discussing in some more detail those methods which were applied in this study (FlexX, AutoDock, and DrugScore), and considering in particular the problem of ligand-induced protein flexibility.

2.2.1 Ligand docking

The goal of docking is the computing of non-covalent protein-ligand complexes. Given the structures of a protein and a ligand, the task is to predict the structure of the formed complex. This is the so-called "docking problem". Assuming that the native geometry

of the complex corresponds to the global minimum of the binding free energy, docking can actually be regarded as an energy optimisation problem [Totrov & Abagyan, 2001], concerned with the search of the lowest-free energy binding mode of a ligand within a given protein binding-site. The macromolecular nature of the protein and the fact that binding occurs in aqueous solution complicate the problem significantly because of the high dimensionality of the configuration space and the considerable complexity of the energetics governing the interactions. Accordingly, heuristic approximations are frequently applied to render the problem tractable within a reasonable time frame. The development of docking methods is therefore also concerned with making appropriate assumptions and finding acceptable simplifications that still provide a sufficiently accurate and predictive model for protein-ligand interactions [Sotriffer et al., 2002a].

Comprehensive overviews on docking algorithms and programs have been given by Sotriffer *et al.* [Sotriffer et al., 2002a] and Halperin *et al.* [Halperin et al., 2002]. The following paragraphs give an introduction into two docking programs (FlexX and AutoDock), which have been used in the context of this work.

Approaches for docking into rigid protein structures

FlexX In FlexX [Rarey et al., 1996a, 1997, 1996b; Kramer et al., 1999], interaction types and geometries according to Böhm [Böhm, 1992a,b] and Klebe [Klebe, 1994] describe the protein ligand interactions. Each interacting group of the molecule to be docked is assigned an interaction type and a corresponding compatibility. Possible interaction types are geometrically restricted hydrogen bonds, interactions between metals and metal acceptors, and hydrophobic interactions, for example those between phenyl rings and methyl groups. For each group capable to form an interaction, a special contact geometry is defined by placing an interaction surface around the centre, usually as part of a sphere. Two groups form an interaction if the interaction centre of one group coincides with the interaction surface of a counter group.

Conformational ligand flexibility is modelled discretely [Klebe & Mietzner, 1994] using a set of preferred torsion angles about acyclic single bonds taken from a library which was compiled from torsional fragments extracted from the Cambridge Structural Database [Allen et al., 1979]. Multiple conformations for ring systems are computed with the program CORINA [Gasteiger et al., 1990; Sadowski & Gasteiger, 1993; Sadowski et al., 1994].

The docking algorithm in FlexX is based on an incremental construction strategy consisting of three phases: In the first phase (*base selection*), the base fragment of the ligand is selected which is then placed into the active site of the protein (*base fragment placement*). Finally, the ligand is reconstructed in an incremental fashion, starting from different placements of the base fragment (*complex construction*). Upon connecting additional fragments, new interactions are screened and the best partial solutions based on the ranking of a scoring function are hooked up until the ligand is completely constructed.

The docking algorithm is relatively sensitive to the selection and placement of the base fragment. If the geometry of a fragment of a molecule to be docked is known (e.g., from a similar ligand crystallised in the target protein), a useful option is to place the referring fragment manually into the binding pocket via the *mapref* command. This reduces the run time of the docking procedure and increases the probability of predicting the correct binding mode of the ligand.

Another possibility to include knowledge about protein-ligand interactions a priori, as user-defined constraint, into the docking process can be realised by FlexX-Pharm [Hindle et al., 2002], an extended version of FlexX. The constraints are determined by selected FlexX interactions and inclusion volumes. They guide the docking process to produce a set of docking solutions with particular properties. By applying a series of look-ahead checks during the flexible construction of ligand fragments within the active site, FlexX-Pharm determines which partially built docking solutions can potentially obey the constraints. Solutions that will not obey the constraints are discarded as early as possible, thus decreasing the calculation time and enabling new docking solutions to emerge.

AutoDock Instead of explicit interaction types and geometries, as realised in FlexX, AutoDock [Goodsell & Olson, 1990; Morris et al., 1996, 1998] uses grid representations of the protein structure.

The binding pocket of the protein is represented by an affinity grid which is calculated for each type of atom occurring in the ligand, typically carbon, oxygen, nitrogen and hydrogen, as well as a grid of the electrostatic potential.

These maps then serve as look-up tables for the calculation of the interaction energy or scoring value during the flexible docking of the ligand. The search can be

performed using one out of three possible search strategies: Monte Carlo simulated annealing, a traditional genetic algorithm, and a Lamarckian genetic algorithm. The latter Lamarckian genetic algorithm is the combination of a traditional genetic algorithm with a local search method to perform energy minimisation. At each generation, a user-defined fraction of the population is subjected to such a local minimisation. The Lamarckian genetic algorithm was observed to be the most efficient and reliable of the three methods [Morris et al., 1998].

Approaches for docking considering protein flexibility

Most docking approaches treat the protein rigid during the docking process, which is a reasonable simplification if the protein binding pocket is sufficiently rigid and does not exhibit significant side-chain rearrangements upon ligand binding. However, in some cases it cannot be justified to neglect protein flexibility [Teague, 2003; Davis & Teague, 1999; Bursavich & Rich, 2002; Najmanovich et al., 2000; Carlson & McGammon, 2000; Verkhivker et al., 2000]: if the system under consideration is known to be flexible or if the available protein structure is not well resolved, as, for example, in the case of a homology model. A study, recently performed by McGovern *et al.* [McGovern & Shoichet, 2003], clearly demonstrated that the success of molecular docking depends significantly on the particular representation of the receptor used in a screen. In their experiment, the best enrichment was produced by docking into the ligand-bound receptor structure instead of docking into the uncomplexed structure or a homology model. The authors note, however, that a receptor determined in one particular ligand-bound conformation could possibly bias the docking screen, preventing the discovery of ligands much different from the ligand already captured in the complex. The same conclusion was drawn by Murray *et al.*, who successfully identified the correct ligand conformation (for three test enzymes) in 79 % of the cases as the lowest energy configuration when the enzyme structure was provided in terms of the crystal structure actually complexed by the tested ligand. Their methodology only docks however in 49 % of the cases successfully when the ligand is screened against an enzyme crystal structure extracted from a complex with another ligand [Murray et al., 1999]. As a consequence, possible adaptations of the protein binding-site upon ligand binding have to be considered in a docking run, in order to avoid false-negative results. As an alternative or pragmatic compromise the scoring functions used to subsequently rank the generated binding modes have to be tailored in a way to tolerate such structural deficiencies.

A possible approach for considering structural variations of proteins is to dock into an ensemble of protein structures showing deviating active site conformations. As shown previously, by averaging structural details and, hence, smoothing the energy landscape, it is possible to circumvent local minima of an otherwise rugged energy surface. This results in a faster convergence of the docking problem [Trosset & Scheraga, 1998; Vakser, 1996]. Furthermore, the treatment of protein structures as ensembles has the advantage that this compensates for some of the structural deficiencies and inequality using only one single protein model. In addition, potential protein flexibility induced upon ligand binding [Ma et al., 1999] can be accounted for.

Several approaches have been developed that can dock ligands into ensembles of protein structures:

- Knegtel *et al.* used (1) simple and (2) energy-weighted averaging for the description of interactions between a ligand and each receptor structure from the ensemble by generating composite grids. These were subsequently used for scoring within DOCK [Knegtel et al., 1997]. Both averaging methods performed equally well in their test cases. Österberg *et al.* extended this approach by testing four methods for merging multiple ensemble entries into one single grid-based lookup table of interaction energies using AutoDock [Österberg et al., 2002]. In their test set, mean and minimum averaging methods perform poorly, but two weighted averaging methods yield consistent and accurate ligand docking.
- Another approach to handle protein flexibility is realised in FlexE, [Claussen et al., 2001] a variant of the FlexX program. FlexE is based on a united protein description generated from an ensemble of superimposed structures of an ensemble. For the structurally deviating parts of the protein, discrete alternative conformations are explicitly taken into account during the incremental construction of the ligand in the binding-site. These geometric alternatives are then combinatorially joined to create new valid protein structures. Thus, conformations of the protein are not limited to those explicitly present in the ensemble.
- DragHome was especially developed for the purpose of ligand docking into approximate homology-modelled proteins [Schafferhans & Klebe, 2001]. The binding-site models are analysed in terms of putative interaction sites, which are predicted by LUDI [Böhm, 1998; Böhm & Klebe, 1996; Böhm, 1994b,a, 1993, 1992a,b]. They are then translated via Gaussian functions into arithmetically or

geometrically averaged binding-site descriptions representing physico-chemical properties. The use of "soft" Gaussian functions to describe protein-ligand interactions smoothes the potential energy surface and, thus, takes into account the limited accuracy of the modelled structures for the purpose of docking. The ligands are similarly expressed by property densities based on Gaussian functions. The docking is performed by maximising the overlap between the functional descriptions for ligand and binding-site representations using the ligand alignment program SEAL [Kearsley & Smith, 1990; Klebe et al., 1994b, 1999].

- In SLIDE, ligand and receptor flexibility is introduced after the initial ligand placement. Collisions are resolved by using mean field theory [Jackson et al., 1998; Koehl & Delarue, 1994, 1996] to select rotations about single bonds in the ligand and the protein side-chains, reducing a maximal number of collisions by minimal conformational changes of both binding partners [Schnecke et al., 1998; Schnecke & Kuhn, 2000].

Another approach for ligand docking, accommodating receptor flexibility, was recently described by Lin *et al.* [Lin et al., 2002, 2003]: Initially, a long molecular dynamics (MD) simulation of the uncomplexed receptor is performed to sample sets of protein conformations. The second phase of their "relaxed complex" scheme involves rapid docking of candidate inhibitors into a large ensemble of MD snapshots, followed by a more accurate scoring using the MM/PBSA (Molecular Mechanics/Poisson Boltzmann Surface Area) approach [Kollman et al., 2000; Massova & Kollman, 2000; Srinivasan et al., 1998] to find the best ligand-receptor complexes.

It should be noted that (to our knowledge) all above-mentioned approaches still await successful applications for finding a novel ligand for a flexible protein system. On the one hand, this is probably due to the fact that some methods are computationally too demanding to perform a virtual screening since the dimensionality of the configuration space is significantly increased; on the other hand, a reliable scoring scheme to identify the native protein-ligand configuration may still be missing.

2.2.2 Affinity prediction

In the following section, a brief overview over general aspects concerning affinity prediction will be given. A detailed description of the scoring function DrugScore is provided,

which is used in the context of this work. For a more comprehensive overview over current approaches for affinity prediction, the reader is referred to refs [Gohlke & Klebe, 2002a; Sotriffer et al., 2002a].

Affinity prediction in case the protein structure is given

General overview The strategy followed in rational drug design depends on whether the three-dimensional structure of the biological target molecule is known or not. If the structure of a target receptor is available, information about the binding-site and principles of protein-ligand interactions can be used to estimate the binding affinity of a given protein-ligand orientation obtained by crystal structure analysis or by computer docking (see section 2.2.1). Accurate and fast scoring is important both for the determination of the correct binding modes from a sample of protein-ligand configurations and for the ranking of a large sample of different ligands with respect to their affinity.

Three main classes of scoring functions can be distinguished: force-field based methods, empirical scoring functions, and knowledge-based methods. *Force-field based methods* [Ewing et al., 2001; Morris et al., 1998; Jones et al., 1997] calculate binding affinity using energy functions developed for 3D structure refinement and molecular dynamics calculations, usually employed in free energy perturbation (FEP) methods [Kollman, 1993, 1996]. Although these methods are relatively accurate, they are computationally demanding and limited to structurally similar ligands. *Empirical scoring functions* approximate the free energy of binding in terms of a "master equation" as a weighted sum of several factors corresponding to arbitrary enthalpic and entropic contributions. The coefficients are optimised by fitting the derived functional form to observed binding data of a training set of protein-ligand complexes with known structure [Eldridge et al., 1997; Murray et al., 1998; Böhm, 1994a, 1998; Wang et al., 1998, 2002; Jain, 1996; Head et al., 1996]. The most recently developed approaches are the so-called *knowledge-based scoring functions* [Muegge & Martin, 1999; Muegge, 2000, 2001; DeWitte & Shakhnovich, 1996; Mitchell et al., 1999a,b; Gohlke et al., 2000a,b; Wallquist et al., 1995; Verkhivker et al., 1995]. They are based on the idea that a sufficiently large data sample can serve to derive rules and general principles inherently stored in this knowledge base. Accordingly, the development of a knowledge-based scoring function at an atomic level is based upon observed frequency distributions of typical interactions in experimentally determined structures: in any system, only those interactions

that are close to the frequency maxima of the interactions in the knowledge base are considered as favourable. Using the concept of the "inverse Boltzmann law" [Sippl, 1995], the frequency distributions of interatomic interactions, derived from protein crystal structures, are converted into "potentials of mean force" or "knowledge-based potentials".

A pragmatic strategy to enhance the hit rate from a docking screening is to reevaluate the best docked binding poses with multiple scoring functions, a procedure called *consensus scoring* [Charifson et al., 1999; Clark et al., 2002]. Only the compounds scored at the top commonly by all scoring functions are further considered. Therefore, in a statistical sense, consensus scoring is more robust and accurate than any single scoring procedure [Wang & Wang, 2001]. It was indeed shown by Wang *et al.* [Wang et al., 2003] and Bissantz *et al.* [Bissantz et al., 2000] that combining any two or three scoring functions clearly enhances the success rate to retrieve active hits. However, when applying consensus scoring, one should keep in mind that, even so the number of false positives can be reduced, the number of true positives might also decline.

DrugScore In the following, the knowledge-based scoring function *DrugScore*, developed in our lab [Gohlke et al., 2000a] will be briefly described. It was used in the present work to predict and score the binding modes of protein-ligand complexes for the development of our approach for ligand-supported homology modelling of protein binding-sites (*MOBILE*) (chapter 3). Furthermore, we applied DrugScore to rank the putative hits in our database screen for the neurokinin-1 receptor (chapter 4).

For the development of DrugScore, the structural information of 1376 crystallographically determined protein-ligand complexes was retrieved from the database ReliBase [Hendlich et al., 2003; Günther et al., 2003; Bergner et al., 2001]. Subsequently, this information was converted into statistical preferences based on 17 atom types [Hendlich, 1998]. The distance-dependent atom-pair distributions $g_{i,j}(r)/g(r)$ for the individual atom types ij are divided by a common reference state which is taken as the average over the distance distributions of all atom pairs $g(r) = \sum \sum g_{i,j}(r)/i * j$. To consider only direct protein-ligand contacts, the upper sample radius has been set to 6 Å (see Fig. 2.2).

At this distance, no further atoms, e.g. of a water molecule, can mediate a protein-

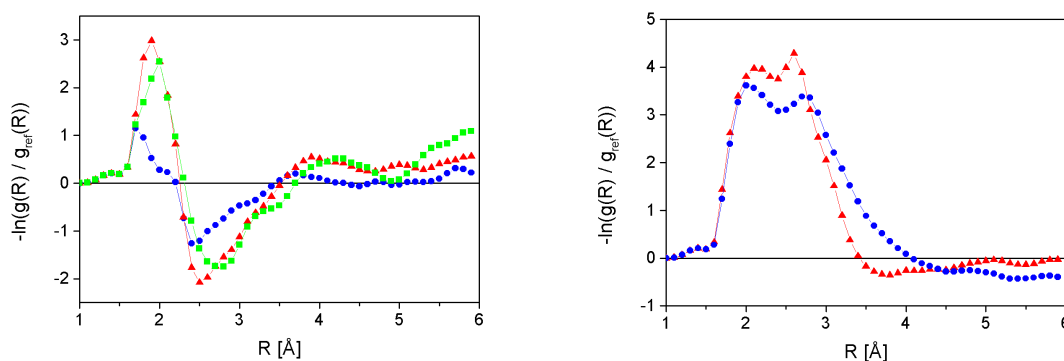


Figure 2.2: **Knowledge-based pair potentials** between polar/ charged (O.co2-N.pl3 (green), O.3-O.co2 (red), O.3-O.3 (blue)) ligand and protein atoms (left) and non-polar/aromatic (C.ar-C.ar (red), C.3-C.3 (blue)) atom pairs (right) as a function of distance R , calculated according to equation 2.2. The first atom-type symbol refers to the ligand, the second to the protein. From [Gohlke et al., 2000a].

ligand interaction. The individual potentials have the form

$$E_{ij}(r) = -kT \left(\ln \frac{g_{i,j}(r)}{g(r)} \right) - \ln g(r) \quad (2.2)$$

These pair potentials are applied together with potentials depending on single (protein or ligand) atom types that express the propensity of an atom type to be buried within a particular protein environment upon complex formation. Contributions of these surface potentials and the pair potentials are weighted equally in the final scoring function.

Affinity prediction when the protein structure is unknown

If the three-dimensional structure of the target protein is unknown, "quantitative structure activity relationships" (QSAR methods) [Kubinyi, 1993; Kubinyi et al., 1997] can be used to establish a relationship between molecular structure and biological activity within a series of active compounds. These models do not only explain the relative differences among the observed affinities, but also allow for an affinity prediction of novel compounds.

The most widely used application of 3D-QSAR in molecular modelling and drug design is the comparative molecular field analysis (CoMFA), first described by Cramer *et al.* [Cramer et al., 1988].

Given a series of ligand molecules binding within an affinity range of four orders of magnitude to the target protein, the first step is to find a spatial orientation of the ligands which is representative for the differences in the binding geometry at the protein binding-site. If the protein structure is not given, the ligands are mutually superimposed in their assumed active-site conformation, e.g., by applying a strategy such as the "active analog approach" [Marshall et al., 1979]. It compares flexible compounds with rigid analogs in particular by mutually excluding part of the accessible conformational space due to structural restrictions. It finally suggests geometries for all ligands in the data set that could possibly correspond to the active-site conformation. Several approaches have been developed which accomplish the task of superimposing ligand molecules while maximising their spatial overlap with respect to similar physico-chemical properties [Lemmen & Lengauer, 2000].

Assuming that a reasonable ligand alignment has been generated, a sufficiently large lattice box is positioned around the molecules with a predefined grid spacing. Field values representing atomic or physico-chemical properties are calculated at each grid point. Subsequently, field properties are correlated with differences in the dependent target property, for example, the binding affinity. In its original implementation, Lennard-Jones- and Coulomb-interactions energies are calculated for all molecules of the data set at the grid points. The general form of the QSAR equation results to:

$$Affinity_n = k + \alpha_1 S_{n,1} + \dots + \alpha_M S_{n,M} + \beta_1 E_{n,1} + \dots + \beta_M E_{n,M} \quad (2.3)$$

The indices 1, 2, ..., M reflect the respective grid points, and $S_{n,1}, \dots, S_{n,M}$ and $E_{n,1}, \dots, E_{n,M}$ describe steric and electrostatic energies at these points. The coefficients $\alpha_1, \dots, \alpha_M$ and β_1, \dots, β_M are obtained from a system of linear equations by partial least-squares (PLS) analysis [Geladi & Kowalski, 1986; Wold et al., 1993]. Binding affinities of new molecules not included in the training set can be predicted using the derived model.

Despite the straightforward definition of CoMFA, there are a number of serious problems and possible pitfalls [Kubinyi, 1993; Thibaut et al., 1994]. One alternative to CoMFA is the comparative molecular similarity indices analysis (CoMSIA) [Klebe et al., 1994a]. Here, Gaussian functions are used to describe steric, electrostatic, and hydrophobic similarities. Similarly, hydrogen-bond donor and acceptor properties are considered [Klebe et al., 1999]. Compared to CoMFA, this approach avoids partic-

ularly steep potentials next to molecular surfaces. Thus, similarity indices are also determined close to and even in the area occupied by the ligands.

2.2.3 Approaches combining protein and ligand-based procedures

When the 3D structure of a target protein is given, drug design usually applies structure-based techniques, while ligand-based information is often neglected. On the other hand, if affinity data for a series of ligands is available, generally ligand-based approaches are applied to discover and optimise novel compounds. If the corresponding structure of the target protein is available, this knowledge remains often unused apart from the assistance to produce a reasonable ligand alignment. In many cases, it could be beneficial to combine protein- and ligand-based approaches, and use the strength of one approach to overcome shortcomings of the other method, or, at least, to use both strategies to mutually confirm results from these complementary approaches. The following sections present examples from literature, where ligand- and protein-based strategies were combined in order to obtain more reliable predictions for protein-ligand interactions.

Combination of protein- and ligand-based information for molecular docking

The docking program FlexX [Rarey et al., 1996a, 1997, 1996b; Kramer et al., 1999] (see section 2.2.1) offers the option to place the base fragment for the incremental construction algorithm of the ligand manually into the binding pocket via the *mapref* command. However, at the moment this option only works for small, relatively rigid fragments. A further option, realised in FlexX-Pharm [Hindle et al., 2002], is the possibility to incorporate more generic information about important characteristics of protein-ligand binding modes into the docking calculation in terms of pharmacophore-type constraints and restraints.

A so-called "similarity-driven approach to flexible ligand docking" was presented by Fradera *et al.* [Fradera et al., 2000]. Given a reference ligand or a pharmacophore positioned in the protein active site, the method allows inclusion of a similarity term during docking. Using the docking program DOCK 4.0 [Ewing & Kuntz, 1997; Ewing

et al., 2001] for the placement of the ligand in the protein binding-site, the similarity program MIMIC [Mestres et al., 1998] is included as a module for the calculation of similarity indices. MIMIC either (1) corrects docking energy scores at certain steps of the ligand incremental construction or (2) applies similarity corrections at the end of the calculation.

Combination of protein- and ligand-based information for affinity prediction

Combination of protein- and ligand-based information is also possible for the prediction of binding affinities. A first step in this direction is the usage of the protein structure, along with a docking protocol, to guide the molecular alignment of ligands in a comparative molecular field analysis (CoMFA). Novel ligands could then be searched by first docking them into the protein structure and then predicting their affinity with respect to the established CoMFA model. This procedure has already been applied by several groups (for example see [Gamper et al., 1996; Bursi & Grootenhuys, 1999; Lozano et al., 2000; Cho et al., 1996; Holloway et al., 1995; Tokarski & Hopfinger, 1997; Vaz et al., 1998; Sippl, 2003]).

This latter approach, which could be termed "protein-supported ligand-based drug design", can also be applied if the structure of the target protein is not available but sequentially related proteins with known 3D-structure have been determined. In such a situation, a homology model can be generated which guides the docking process of the ligand molecules [Jalaie & Erickson, 2000; Schafferhans & Klebe, 2001; Kim, 1998] (see also next section). As an advantage, the significance of the generated CoMFA model can also serve to estimate the quality of the generated homology model.

Wade *et al.* used selected interaction energy components to include information from the protein into a 3D-QSAR model. Their COMBINE approach [Ortiz et al., 1995, 1997; Wang & Wade, 2001] quantifies ligand-receptor interaction energies by molecular mechanics in terms of van der Waals and Coulomb contributions per residue, followed by a PLS analysis in order to derive a 3D-QSAR model. Other similar approaches have also been described [Kurinov & Harrison, 1994; Kulkarni & Kulkarni, 1999; Rognan et al., 1999; Grootenhuys & van Galen, 1995; McCarthy et al., 1997].

Another procedure which considers protein information not only for the suggestion of

a meaningful ligand alignment but also includes information about the surrounding protein environment in the potential field calculations, is the recently developed AFMoC approach [Gohlke & Klebe, 2002b]. AFMoC tailors protein-specifically adapted DrugScore pair-potentials to one particular protein by considering additional ligand-based information in a CoMFA-type approach. As in CoMFA, the results of the analysis can be interpreted in graphical terms by contribution maps, and binding affinities of novel ligands are predicted by evaluating the established 3D QSAR equation. Compared to the original, solely ligand-based CoMFA approach, the AFMoC method has the following striking advantage: Whereas in CoMFA a large data set of structurally diverse training compounds spreading over a sufficient range of binding affinities is necessary, AFMoC is able to generate predictive models also with a fairly limited number of ligands, spanning over a lower range of affinity. As a further advantage, AFMoC allows the user to gradually move from general knowledge-based potentials to protein-specifically adopted ones, depending on the confidence in the protein structure (e.g. in the case of a homology model) and the amount of ligand data available for training. Whereas CoMFA can only interpolate among the data points defined by the ligands of the data set, AFMoC can still extrapolate to some extent into areas not yet experienced by any of the training set compounds since in these areas the original DrugScore potentials generate a reasonable predictive power.

Combination of protein- and ligand-based information for homology modelling

The incorporation of ligand information explicitly into the protein modelling process has not yet been described in literature. Instead, ligand information was used to subsequently refine an already existing homology model by energy minimisation or molecular dynamics. Furthermore, ligands are docked into homology models and the relevance of the docking solution or the significance of a CoMFA model based on docked ligands served as validation criterion to assess the generated protein model. For a more detailed overview, see section 2.3.

2.3 Examples of homology modelling for the purpose of structure-based drug design

Several studies have been described in literature where homology models of proteins have been used to explain putative interactions between a protein and ligands ([Bourdou et al., 1997; Lozano et al., 1997; Garcia-Nieto et al., 2000; Zhang et al., 2001; Marhefka et al., 2001; Le Novere et al., 2002; Gieldon et al., 2001; Escherich et al., 2001; Bathelt et al., 2002; Lopez-Rodriguez et al., 2001; Bissantz et al., 2003; Vaidehi et al., 2002; Gouldson et al., 1997; Moro et al., 1998b,a]). In some cases, the generated models were used for the design of new potent inhibitors ([Tiraboschi et al., 1999; Rong et al., 2002; Kiyama et al., 1999]).

However, in none of these examples ligand information was explicitly considered during the modelling of the protein. Usually, a homology model based only on one or more template structures is initially produced and subsequently the ligand(s) are placed into the modelled binding pocket. This is accomplished either manually or using an automatic docking tool. Another strategy follows the superimposition of template and model followed by the merging of the coordinates of a ligand as adopted in the template structure into the model. In most cases, the resulting complexes are further optimised, e.g. by using a molecular dynamics protocol, which, to some extent, uses the ligand information for the protein modelling, at least in terms of a local optimisation.

In another approach, ligand information was used by Jansen *et al.* for modelling the serotonin 5 – HT_{1A} receptor. The optimisation was performed with the minireceptor modelling program Yak [Vedani et al., 1993] based on an extracted active site of a homology model using three high-affinity ligands [Jansen et al., 1997].

An important task in model-building protein-ligand complexes is the quality assessment of the produced models, in particular if different orientations of active site residues are possible. This step is usually performed by visually analysing the interaction geometry between protein and ligand functional groups. This consecutive procedure appears rather inefficient. Only a few approaches assess the quality of the generated complexes in a more sophisticated way: Johnson *et al.* created a library of protein models that are subsequently screened by rigid ligand docking. The more relevant protein models achieve better scored docking solutions, and the quality of the generated binding modes is used to select the most relevant models [Johnson et al., 2003]. This approach has

been applied to the modelling of F_v antibody fragments. Results were compared to the known crystal structures. However, it requires experimental data about the conformation of the docked ligand. Bissantz *et al.* evaluated their generated homology models (agonist and antagonist bound models of three human G protein coupled receptors) by retrieving known agonists and antagonists via docking from a database assembled by such entries and additional randomly collected "drug-like" compounds [Bissantz et al., 2003]. Jalaie *et al.* developed a homology model of spinach photosystem II. After docking inhibitors, a highly predictive CoMFA model was derived from the resulting alignment. It helped to score the quality of the homology model [Jalaie & Erickson, 2000]. A similar approach was followed in our group by Schafferhans and Klebe. Structurally distinct thrombin inhibitors were docked onto models of thrombin generated from a set of serine proteases with 28 up to 40 % sequence identity. Compared to the crystal structures of actually known thrombin complexes, ligand binding modes were obtained with an average rms deviation of 1.4 Å [Schafferhans & Klebe, 2001]. Based on the generated alignment of 88 thrombin inhibitors, a significant 3D QSAR model could be established.

3 Development of an approach for Ligand-supported Homology Modelling of Protein Binding Sites using Knowledge-based potentials

3.1 Strategy and computational realisation

3.1.1 General overview

Below, we describe in detail how we complement data about related proteins with information about the binding modes of bioactive ligands to generate more realistic homology models of protein binding-sites. A schematic overview of our strategy, which was initiated by the development of the DragHome concept and is followed by *MOBILE* (Modelling binding-sites including ligand information explicitly), is given in Fig. 1.4. Starting with the (crystal) structure of one or more template proteins, several preliminary homology models of the target protein are generated (step 1). After placing one or more ligands, known to bind to the target protein, into an averaged binding-site representation of the generated binding-site models (step 2), the protein models are generated, now considering explicitly the docked ligand(s) (step 3a). After scoring the thus generated complexes with DrugScore, a final model is obtained by selecting the model which explains best the observed ligand binding (affinities) (step 3b). The modelled complexes can be further refined considering the composite picture of the best side-chain conformers taken from different models and minimising the side-chain-to-ligand interactions using a common force-field (step 4).

3.1.2 Step 1: Generation of preliminary protein models

The program MODELLER [Marti-Renom et al., 2000; Sali & Blundell, 1993; Fiser et al., 2000] is used to generate initial homology models in the first step of our approach (Fig. 1.4). MODELLER generates protein 3D structures by satisfying spatial

restraints imposed by the sequence alignment with the template structure and applying the terms of the CHARMM-22 force-field [Brooks et al., 1983]. A 3D protein model is obtained by optimising the molecular probability density function while simultaneously minimising input restraint violations. To guarantee sufficient conformational sampling of each active site residue, several homology models are generated in this step. Preliminary tests showed that a number between 10 and 100 models provides a satisfactory sampling. To optimise the local interactions, all models obtained are subjected to a crude simulated annealing refinement protocol available in MODELLER.

3.1.3 Step 2: Placing the ligand(s) into the homology models

As a next step, proper ligand orientations need to be obtained. Three scenarios are described, characterised by a decreasing amount of experimental information available:

1. One or more ligands are known to bind to the target protein, and the complex crystal structures with the related template proteins are available. It can be assumed that the ligand binding modes are similar in the target and the template protein. Accordingly, ligands are then transferred among these structures keeping their orientation as a restraint for the subsequent modelling process.
2. One or more ligands are known to bind to the target, however, no complex crystal structure with the template is available. In this case, the ligand(s) can either be placed into the template protein structure by docking, and the resulting orientation can then be used to restrain the following protein modelling process. Alternatively, the coordinates of a similar ligand, crystallised together with the template protein, serves as a reference to restrain the protein modelling process. The known ligand is then transferred into the modelled proteins as described in the following section.
3. If no structural information about ligands binding to the template protein is available, one or more ligands (known to bind to the target protein) are docked into the homology models of the target protein. Since a homology modelling program generates a set of different models with similar energies, ligand docking is attempted as a placement into ensembles of the modelled protein structures. Two different approaches were combined to place ligands into ensembles of model-built protein structures. Following Sotriffer *et al.* [Sotriffer et al., 2002b], DrugScore

potential grids were calculated in the binding pocket of each homology model by evaluating protein-ligand interactions between a predefined probe atom, placed at each grid point, and the surrounding protein environment. At short interatomic distances, the pair potentials were supplemented by a Gaussian-type repulsive term as described by Gohlke *et al.* [Gohlke & Klebe, 2002b]. Grids of identical size were used for each homology model. Their dimensions were adjusted to fully embed the ligand in its crystallographically determined binding mode with an additional margin of at least 4 Å. The ligands were then docked into the merged binding pockets using AutoDock 3.0 after averaging the grid maps representing the potential energy using the clamped grid method as described by Osterberg *et al.* [Österberg *et al.*, 2002]. The Lamarckian genetic algorithm was applied using the docking protocol as given by Sotriffer *et al.* [Sotriffer *et al.*, 2002b].

3.1.4 Step 3a: Incorporating ligand information into the homology modelling process

Having placed the ligand(s) in a near-native orientation into the consensus binding-site of the modelled protein, new models are generated using MODELLER which additionally incorporate information about these ligand(s). During this modelling step, the ligands are kept fixed in space. The presence of the ligand(s) is included into the homology modelling process in terms of user-defined restraints. Scaled DrugScore pair potentials are added to the MODELLER force-field to provide information about the interactions experienced between fixed ligand(s) and flexible protein atoms. The scaling of DrugScore potentials with respect to the MODELLER force-field is described in detail in the section 3.4.3. No further interactions between protein and ligand atoms are considered. To make the DrugScore potentials suitable for a minimisation procedure, we approximate them by cubic splines (assigning a range from 0 to 6 Å and a bin size of 0.1 Å). This can be realised through the MODELLER interface. To include the repulsive interactions at short distances, the above-mentioned Gaussian repulsion term has been added [Gohlke & Klebe, 2002b].

The protein modelling process is not necessarily restrained to one ligand. If several ligands are known to occupy distinct parts of the binding pocket, a combination to a composite "super-ligand" can be attempted.

3.1.5 Step 3b: Scoring the generated models

Having generated a set of ligand-supported homology models, the next objective is to identify the best one(s). Quality assessment of homology models usually applies fold plausibility criteria or tries to assess local features considering protein atom interactions only [Bowie et al., 1991; Colovos & Yeates, 1993; Dominy & Brooks, 2002; Eisenberg et al., 1992; Laskowski et al., 1993; Luthy et al., 1992; Melo & Feytmans, 1998; Simons et al., 1999; Wang et al., 1995b,a]. The primary interested in the scope of this Thesis is to obtain near-native models of protein binding-sites, accordingly the standard protocols for evaluating protein homology models would be insensitive and non-conclusive. Also, the MODELLER objective function would not provide a proper criterion, as it assesses matching with all requested input restraints. Assuming that the modelled protein-ligand geometry corresponds to a near-native geometry, we require a scoring function suitable to evaluate protein-ligand interactions. As DrugScore shows good performance to identify near-native ligand poses from a set of decoy binding modes in rigid binding pockets, this method was used in turn to identify near-native binding-site geometries with respect to residue side-chain orientations towards the ligand(s).

3.1.6 Step 4: Optimising and refining the homology models

To optimise the modelled binding-sites, a strategy of combining good solutions on a per-residue basis from different homology models is pursued. In the case of identical main-chain orientations, the most appropriate side-chain rotamers are assembled from the different models. As the ligand(s) have been placed in the previous modelling step, the DrugScore rankings between ligand atoms and individual side-chain rotamers are used to select the most appropriate solution from the set of generated protein side-chain orientations. In this context, the number of side-chain conformers for each residue is reduced by performing a complete linkage clustering, merging two conformers within a user-defined threshold (by default 1.0 Å). Then the conformer with the best DrugScore value is selected as cluster representative and the with unfavourable rankings are eliminated. Finally, all combinations between the remaining cluster representatives are generated. Solutions that produce intramolecular clashes are discarded. The total DrugScore scores of the combined pockets are obtained by summing up the individual scores of the considered side-chain conformers. Finally, the model with the best total DrugScore values is chosen.

Since DrugScore pair-potentials implemented into MODELLER consider directionality of interactions only implicitly, a subsequent structural optimisation using the MAB force-field in MOLOC is performed. This force-field handles H-bonds using explicit angular dependencies [Gerber, 1998; Gerber & Müller, 1995]. In addition, this step finally removes strained interactions within the binding-site residues.

3.2 Results and Discussion

The following sections will demonstrate that the *MOBILE* approach produces more reliable homology modelled complexes if ligand information is included in this process. Subsequently, two "real life" applications will be given to assess the scope and demonstrate the power of this method.

3.2.1 Analysis of the generated binding-site models

Comparison of binding-site models generated with and without ligand information

To assess the influence of ligand information on the protein modelling step, root mean-square deviations (rmsd) of the modelled binding-site residues with respect to orientations found in reference crystal structures were evaluated for the test data set (46 protein-ligand complexes, see 3.4.1 and Table 3.5). Models were generated (1) without ligand information and (2) considering ligands in terms of the DrugScore pair potentials. For each of the 46 test set proteins, 10 models were generated with new side-chain and backbone orientations. Of these, the one with the lowest rmsd with respect to the crystal structure was selected, and an average rmsd considering all atoms of all binding-site residues was computed. It amounts to 1.90 Å if ligand information was considered (strategy 2) and increases to 2.08 Å if no ligand information has been used (strategy 1). A paired t-test [Zar, 1999] indicates (with a significance of 0.01) that these mean values are significantly different. With respect to the average rmsds of each of the 46 binding pockets, in 28 cases better models were obtained considering the ligand in terms of DrugScore potentials (see Fig. 3.1). In 17 cases better models were generated neglecting the ligand. In one case a model of equal quality (with respect to rmsd) was obtained. According to the paired t-test, the respective 46 mean rmsd values show a significance level of 0.1 in favour of the ligand-supported models. Remarkably, when the ligand is considered in terms of van-der-Waals potentials only, the resulting binding-site models are even worse (in terms of rmsd values) than those generated neglecting ligand information. The mean rmsd value over all atoms from all binding-sites amounts to 2.34 Å if the ligand is considered in terms of van-der-Waals potentials. A possible explanation for this different performance might be attributed to the significant difference in the steepness of the DrugScore versus van-der-Waals po-

tentials. Soft potentials are more tolerant with respect to slight structural deficiencies that generally occur with model-built structures.

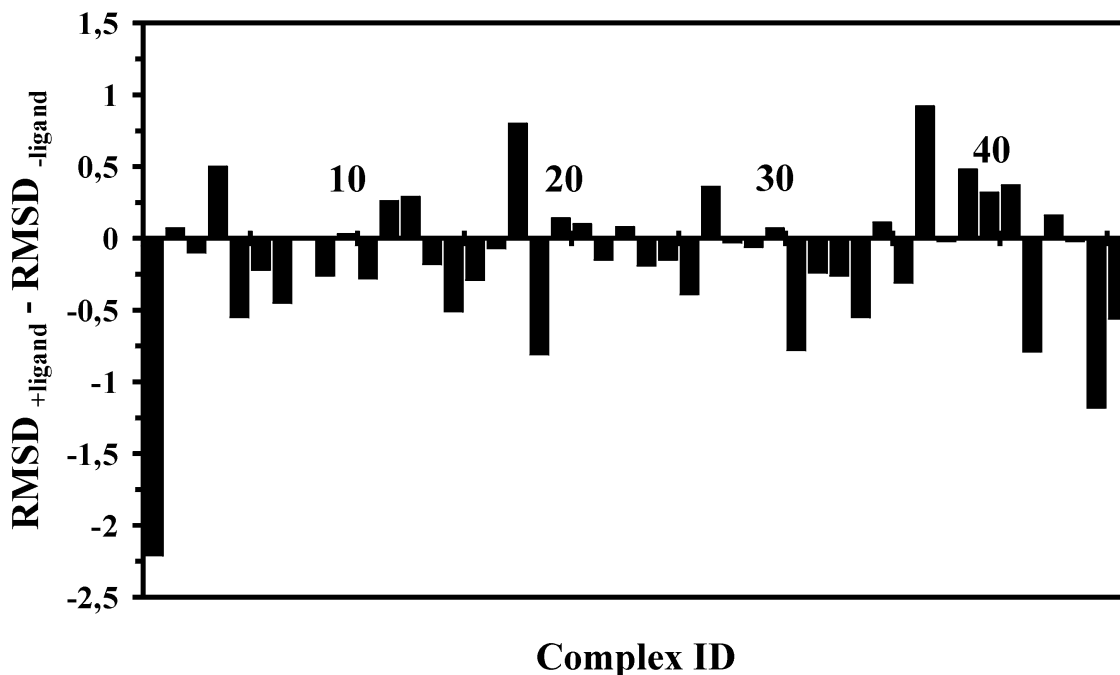


Figure 3.1: The differences between rmsd values exhibited by binding-site models (including side-chain and main-chain atoms) generated with ligand information (+ligand) and without ligand information (-ligand).

Fig. 3.2 illustrates the benefit of including ligand information into the protein modelling process. Here, the orientations of all 11 binding-site residues of glycosidase complexed with adenine (1aha) were predicted either neglecting (Fig. 3.2a,b) or considering (Fig. 3.2c,d) ligand information. Regarding all binding-site residues, the best of the 10 generated models showed an overall rmsd value of 1.25 Å (neglecting ligand information) and 0.8 Å (considering ligand information), respectively.

Even for the model which was generated without regarding the ligand, the overall rmsd value is rather satisfying. With 1.25 Å it is even better than the average value found for all 46 test set complexes including ligand information (1.90 Å). However, three modelled residues (Tyr70, Tyr111, Ile155, see Fig. 3.2) will clash with a bound ligand if it is inserted in its crystallographically determined orientation. Besides visual inspection, the obtained DrugScore rankings potentially indicate the quality of the generated binding-site models. While the complex generated considering ligand

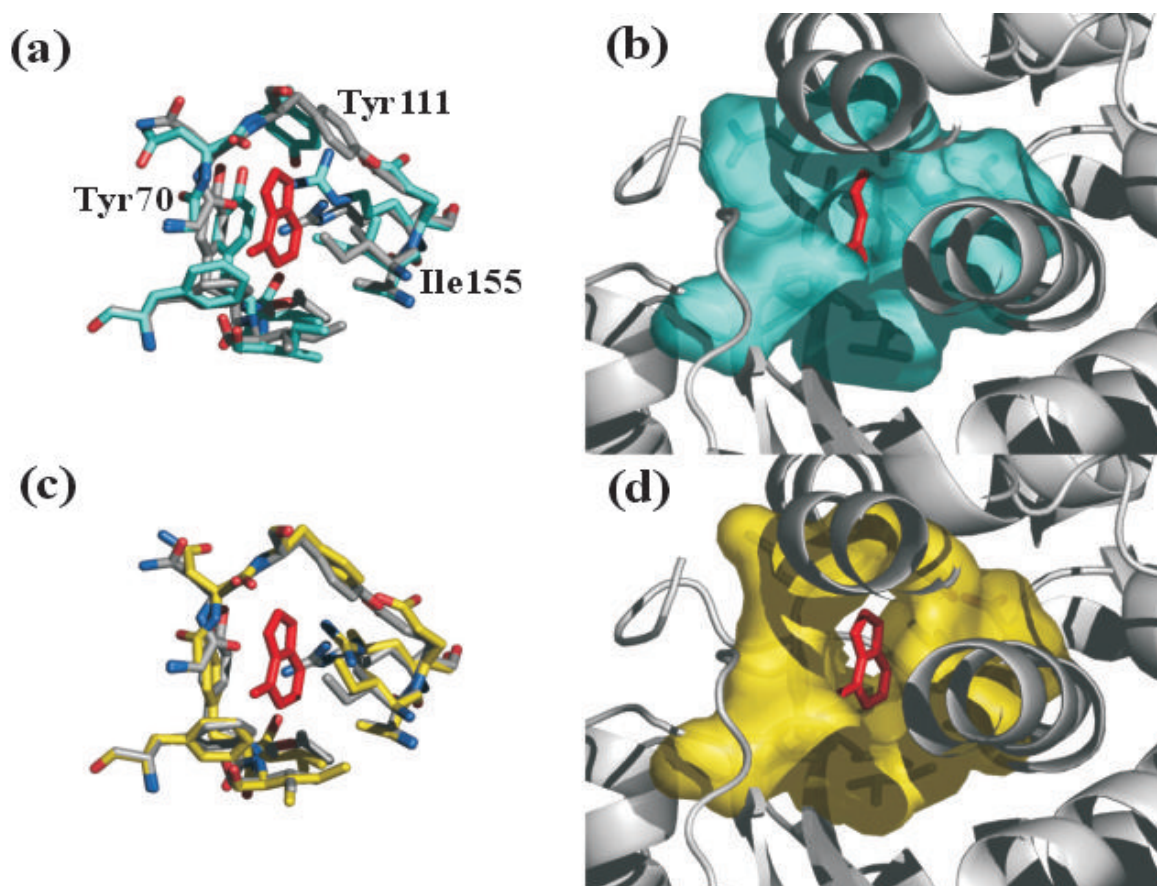


Figure 3.2: Binding site of glycosidase complexed with adenine (1aha). The residue orientation taken from the crystal structure (atom type coded colouring) and those generated by homology modelling (cyan or yellow) are shown (a) neglecting ligand information and (c) including ligand information. Figures (b) and (d) depict the molecular surfaces of the binding-site residues in their modelled orientation. The bound adenine has been considered in its orientation as found in the crystal structure. For clarity, only the binding-site residues are shown in (a) and (c).

information scores only slightly worse than the native complex, the model generated neglecting ligand information exhibits a strongly unfavourable score.

Assessing the side-chain prediction accuracy of MODELLER

To assess MODELLER's power to correctly predict side-chain orientations in protein binding-sites, again, homology models for all members of the 46 test suite were generated. Deviating from the previous test, now 100 models (to sample search space more exhaustively) were generated considering only the side-chain orientations of the binding-site residues (thus keeping their backbone coordinates fixed). The orientations of all binding-site side-chains were generated simultaneously. Ligand information was (1) included in terms of the DrugScore potentials, (2) included in terms of van-der-Waals potentials and (3) fully ignored. The quality of the protein models was validated in two ways: (1) the computed binding-site models were considered in total and (2) for each single residue, only that conformer which had the lowest rmsd compared to the crystal structure was considered.

The results are summarised in Table 3.1. Regarding the modelled binding pockets in total, the best solutions are obtained while including the ligand in terms of DrugScore potentials (1.74 Å rmsd). Paired t-tests [Zar, 1999] indicate that this mean value significantly differs (with a significance level of 0.05) from the mean values which are obtained when neglecting ligand information (1.82 Å rmsd) or including it in terms of van-der-Waals potentials (1.88 Å rmsd). Considering the best conformer of each predicted side-chain, all three approaches (ligand included in terms of (1) DrugScore potentials, (2) van-der-Waals potentials, (3) ignored) seem to generate equally good results (mean rmsd values of 1.04, 1.06, and 1.03 Å). This is probably due to the fact that the conformational space of a residue is exhaustively screened by 100 probe conformers, irrespective whether a ligand is present or not.

Comparing the rms deviations for each of the 46 modelled binding pockets in turn, in 30 cases better models were obtained when considering the ligand in terms of the DrugScore potentials (see Fig. 3.3). In 14 cases models with a lower rmsd were obtained when neglecting the ligand, and in two cases models with equal rmsds were generated. According to the paired t-test, the respective mean rmsd values are significantly different (with a significance level of 0.1).

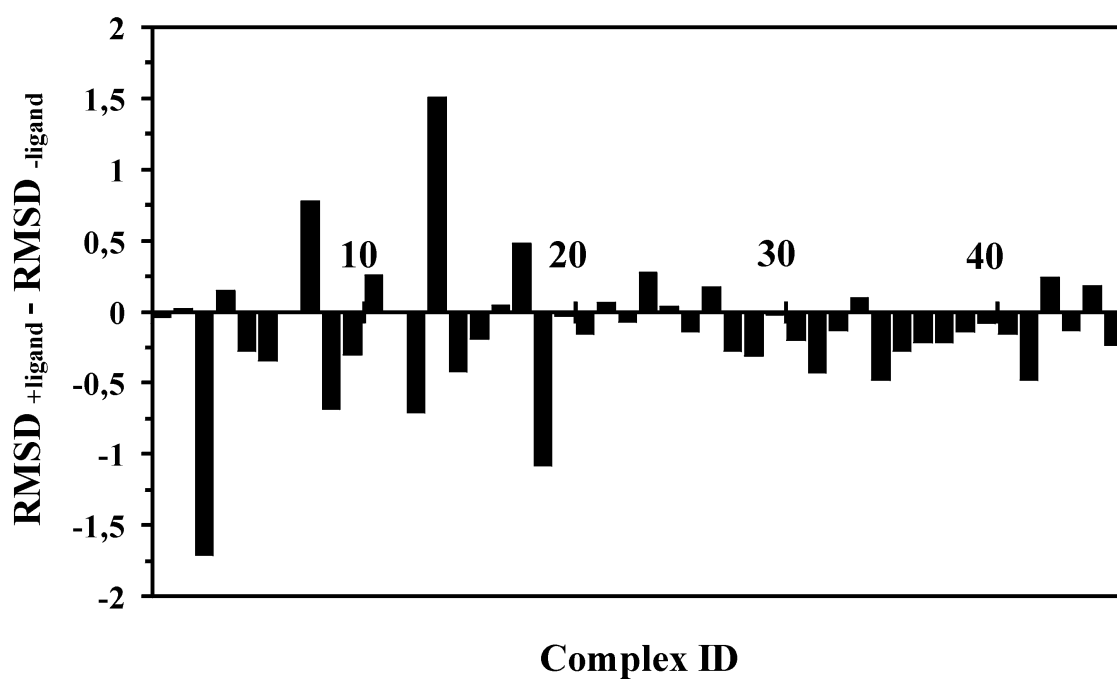


Figure 3.3: The differences between rmsd values observed for binding-site models (including only side-chain atoms) generated with ligand information (+ligand) and without ligand information (-ligand).

Table 3.1: Results for predicting side-chains located in the active sites of the test set proteins

active site	
predictions with DrugScore potentials	
MIN ^{a)}	1.74
BEST ^{b)}	1.04
predictions with VDW potentials	
MIN	1.88
BEST	1.06
predictions without ligand information	
MIN	1.82
BEST	1.03

Side-chain predictions were performed for all binding-site residues of the test data set (Table 3.5) simultaneously, keeping the ligand and the remaining part of the protein fixed. In each case, the rmsd is calculated for all atoms in the given category (i.e. no averaging over residues or structures).

^{a)} The MIN values consider the binding pockets as whole entities. For each protein, the binding pocket with the rmsd value closest to the crystal structure was considered.

^{b)} The BEST values consider the best side-chain conformer for each single residue of the generated binding site models compared to the residue in the crystal structure.

Table 3.2.1 gives a detailed list of the deviations of the multiple binding-site models generated in the presence of ligand information (in terms of DrugScore potentials). MODELLER computed geometries with $\text{rmsd} < 2.0 \text{ \AA}$ in 32 of 46 cases. Obviously, the prediction accuracy does not depend on the number of residues to be modelled but rather on the type of residues for which conformations have to be generated. If a binding pocket contains rigid, space-filling amino acids (in particular Phe, Tyr, Trp or His), remarkably large deviations are modelled compared to the crystal structure: Among the 32 satisfactorily modelled pockets, on the average 2.8 Phe, Tyr, Trp or His residues are present, whereas the 14 cases with $\text{rmsd} < 2.0 \text{ \AA}$ comprise 5.8 residues of this type. This is probably due to the fact that an incorrect geometry of a bulky residue also provokes incorrect geometries of adjacent residues. This influence increases with a growing number of bulky residues in an active site.

Identification of the best binding-site models

The above-described homology models generated to assess the prediction accuracy of MODELLER were subsequently used to evaluate DrugScore's ability to identify near-native complex geometries ("near-native" requires an $\text{rmsd} < 2.0 \text{ \AA}$ over all binding-site residues with respect to the corresponding crystal structure).

For each of the 46 test cases, DrugScore rankings were calculated for the crystal structure and the 100 model-built complexes. The crystal structures, assumed to represent the global optimum, should obtain the best score. In fact, DrugScore was able to retrieve the crystal structures on rank 1 in 32 out of 46 cases (70 %). Data in Tables 3.2.1 and 3.3 indicate good correlation between DrugScore ranks and deviations of model-built vs. crystallographically determined binding-sites. If a near-native geometry ($< 2.0 \text{ \AA}$) was generated by MODELLER (32 cases), DrugScore was also able to identify a pose with $\text{rmsd} < 2.0 \text{ \AA}$ on rank 1 in 21 cases (66 %).

Combination of side-chain conformers in carboxypeptidase A

Even in the overall best model, MODELLER does not necessarily generate the best possible orientation for all binding-site residues. Thus, an improved model can be obtained by combining conformers from different models. Fig. 3.4a shows the crystallographically determined binding-site of carboxypeptidase A complexed with L-

Table 3.2: Results for binding-site models of 46 protein-ligand complexes generated by MOD-ELLER and scored with DrugScore

pdb code	rmsd of ¹ <i>st</i> DS (Å) ^{a)}	best value (Å) ^{b)}	rmsd residues in the binding-site ^{c)}
≤1.0 Å			
1ABF	0.23	0.13	KQE WF CDDMTRMNN
1LAH	0.96	0.96	D YF SSLSRLSTQD
≤1.5 Å			
1LNA	1.02	1.02	NNA FLVHEILRH
1IMB	1.13	0.90	EDIDGTEGSGTAY EID
1HSL	1.15	1.15	YL SSLSRLGTTQD
1F3E	1.17	0.98	D Y DCIQGGLAVMG
121P	1.18	1.18	AGGVGKSA FV DEDPTTAGNKDLSAK
1MLD	1.32	1.32	IRRINLR H GTVSAM
1BLH	1.41	1.28	ASK Y SNEINGQAI
≤2.0 Å			
1LMO	1.50	1.05	DQIN YWVDNAW
1PPL	1.52	1.52	EENDGSS YGDSQFLF IDGTTLL YLFI
1PSO	1.66	1.42	MEVDGST YGTFFIIY DGTSLQMLI
1BUG	1.71	1.21	HHFH HHIMGN F A FH
1HDC	1.72	1.63	STGMSLLT Y PGMTMTT W
1APT	1.75	1.49	ENDG YGDSFLF IDGTTLL YLFI
1POC	1.85	1.69	IYWCGHGCHDHTLFF VMYI
1CTR	1.87	1.09	E FILEMEAVMA
1AHA	1.89	1.27	VYIF GN Y IAER
1HEF	1.89	1.25	RDGADDVIGGIPVI
1ABE	1.96	1.96	KQE WF DDMLTRMNN
1ROB	1.97	1.39	Q H KVNTDR H F DAS
≤2.5 Å			
1MRG	2.05	1.41	IYIF GD Y IAER
1BYB	2.21	2.21	MLD WI HN VDAERYQWFKSGHW TCMEALLR

pdb code	rmsd of ^{1st} DS (Å) ^{a)}	best value (Å) ^{b)}	residues in the binding-site ^{c)}
1EPB	2.21	1.81	FIFWMVLAFKVVAIKY
1ICN	2.21	2.20	YFMMIKFVFFYLADLWFLQQY
1STP	2.24	1.57	NLSYSVGNWASTWWLD
1TLP	2.25	1.88	YNNAFWFLVHEHYELRDH
1ATL	2.26	1.32	EETLGTHEHH CIRPGL
1RDS	2.27	2.13	YHEYHDYEEPGARHGDDF
1CPS	2.32	1.79	HRERNRHSYLIYAGTEF
1ELA	2.34	1.85	HTVAWTGCQGDSTSFVSR
1TMN	2.34	1.98	YNNAFFLVHEHYEILRH
1HYT	2.36	2.16	NAFFLVHEHYEIGLRH
1RBP	2.42	1.65	LFLAFATAVLMVGMYLQHYFF
1RNT	2.43	2.43	NYHKYNNYEERHNNF
1AZM	2.46	2.42	FHHEHLLVSLTHW
>2.5 Å			
1SRJ	2.54	2.54	NLSYSAVGNAYWASTWWLD
1BBP	2.55	2.09	ENVEGWANYHYFIHLYNFYKFWL
1DIE	2.62	2.62	WHTTFVWEEHDD
1DID	2.64	2.48	WHMTFVWENEHDHD
1SNC	2.66	2.10	DTRLLEDKYLRY
1CBX	2.76	1.49	HERNRHSLIYAGTE
1HFC	2.89	1.62	GNLAHYVHEHHYPSY
1CIL	3.15	2.94	WNHQHHEHVFVLVSLTTPPW
1MRK	3.26	2.22	YIMFEGNYIEREW
1ACJ	3.31	2.85	GWGGYEFYWIHGY

The rmsd values are calculated for all atoms of each model (i.e. no averaging over structures). The models are treated as whole entities (i.e. no further optimisation by combining fragments from different models).

Rigid, space-filling amino acids (Phe, Tyr, Trp, His) are coloured red.

^{a)} rmsd of the binding-site model found on rank 1 by DrugScore, with respect to the crystal structure.

^{b)} rmsd of the binding-site model with the least deviation from the crystal structure.

^{c)} residues in the active site (given as one-letter code) for which new geometries were computed.

Table 3.3: Results for scoring multiple solutions of 46 protein-ligand complexes generated by MODELLER

	% of complexes with solutions exhibiting rmsd of the crystal structure			
	< 1.0 Å	< 1.5 Å	< 2.0 Å	≥2.0 Å
All ranks ^{a)}	9	43	70	30
^{1st} rank ^{b)}	44	47	66	34

^{a)} All solutions of each modelling experiment for the test data set (Table 3.5) are considered. The number expresses the portion of all complexes for which at least one solution with the given rmsd value was computed by MODELLER.

^{b)} Only the binding-site geometry scored to be on the first rank by DrugScore is considered. The numbers are related to the ones in the first line.

benzylsuccinate together with 100 generated models (yellow). The three best-scored models (DrugScore) are shown in Fig. 3.4b. Each individual model contains at least one residue rotamer that differs significantly from the crystal structure. In contrast, a combination of rotamers considering only the individually best-scored ones retrieved from the entire ensemble matches the crystal structure to a greater extent (Fig. 3.4c). To assess whether the thus generated binding pocket is capable to reproduce a correct ligand pose, we flexibly docked the ligand present in 1cbx into either the crystal structure or into the model. Prior to this, the modelled binding pocket was minimised with MOLOC in presence of the ligand. Of course, this procedure does not correspond to a realistic scenario in real-life modelling case studies because the modelling process was restrained with the ligand in its orientation known from the crystal structure, which will usually not be given. The Lamarckian genetic algorithm was applied in AutoDock using DrugScore grids to describe the protein binding-sites (Fig. 3.4d) [Sotriffer et al., 2002b] and ten independent runs were performed. In both cases, docking produced two different ligand placements (AutoDock scoring energies for crystal structures: -8.03 and -7.62 kcal/mol and model-built complex: -8.39 and -8.10 kcal/mol, respectively). In case of the experimentally protein structure, the first solution has an rmsd of 1.21 Å with respect to the crystal coordinates. For the model, the second solution deviates from the crystal coordinates by 0.75 Å rmsd. The close energy ranks and the small

rmsds indicate high similarity between the model and the original crystal structure. The ligand orientation, albeit found on rank 2, could be reproduced satisfactorily via docking into the generated model. This convincing result stimulated us to embark on some real-life modelling applications.

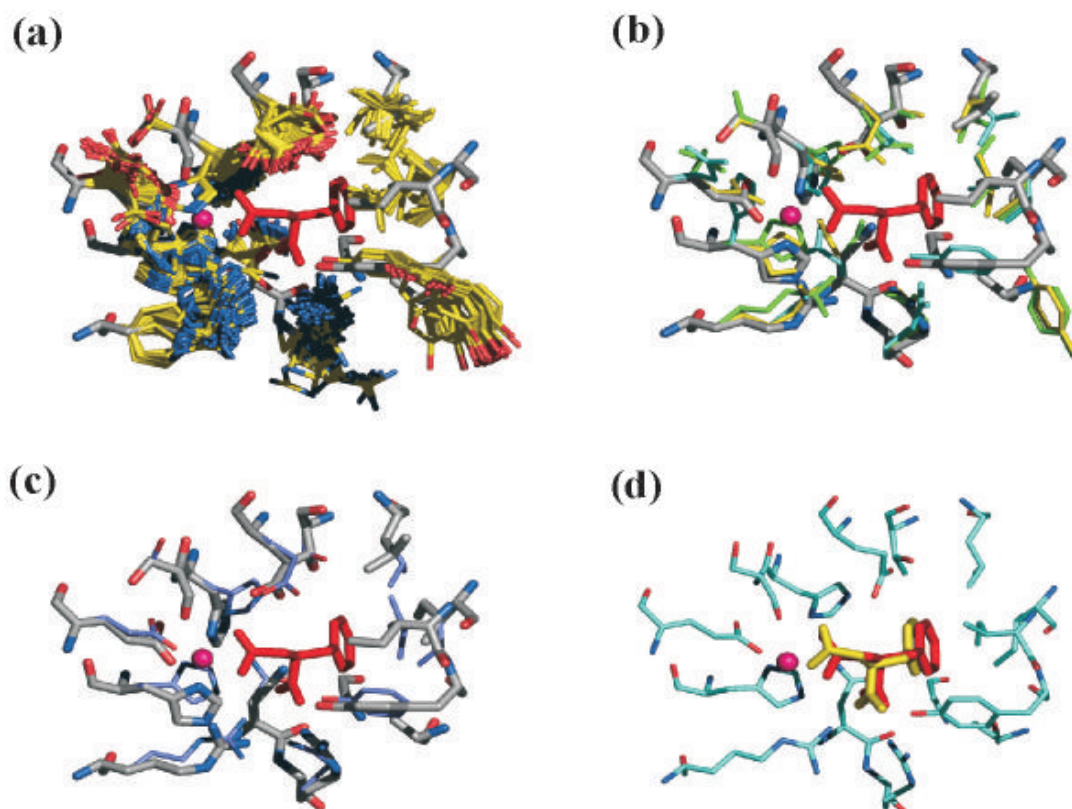


Figure 3.4: Binding site residues of carboxypeptidase A complexed with L-benzylsuccinate (1cbx). (a) shows the crystal structure (atom type coded colouring) together with an ensemble of 100 models (yellow). The ligand in its orientation as found in the crystal structure is coloured in red; (b) depicts the 3 models which obtained the best DrugScore values (green > cyan > yellow). The model which results as best combination of all binding-site residues retrieved individually from the generated 100 models is depicted in violet (c). The best docking solution obtained for the ligand based on the combined and subsequently minimised model is shown in (d). (Orientation of docked ligand in yellow, crystallographically determined orientation in red)

3.2.2 Modelling case studies

Modelling factor Xa based on trypsin

In the previous test examples entire binding pockets have been modelled. Furthermore, the modelling process was restrained by the ligand in its orientation known from the crystal structure. In real-life applications, the protein to be modelled may differ by only several mutations with respect to the template protein(s) and the ligand geometry might not be exactly known beforehand. To apply such a scenario, homology models of factor Xa were generated using bovine trypsin as template. Both proteins are members of the class of trypsin-like serine proteases and share 38 % sequence identity in common (see Fig. 3.5a). They are known to bind the ligand RPR128515 in a similar orientation [Maignan et al., 2000]. A sequence alignment was produced by MALIGN3D [Sali & Blundell, 1990]. In order to mimic a scenario where a ligand orientation is not given by crystallography, we generated 10 preliminary homology factor Xa models based on the crystal structure of trypsin (1f0u) excluding ligand information. The ligand RPR128515 was then flexibly docked into the merged binding pockets using AutoDock. The docking solutions, together with the ensemble of all modelled residues known to be crucial for the ligand binding, are depicted in Fig. 3.5b. Nine deviating docking solutions were obtained. The best solution (rmsd = 0.97 Å with respect to the ligand orientation in the crystal structure 1ezq) is found on rank six (AutoDock energy score: -14.35 kcal/mol). While the solution found on rank 1 (with an energy score of -14.80 kcal/mol) deviates by 3.33 Å rmsd, four additional solutions were generated with an rmsd value < 2.0 Å. The solution found on rank 2 differs by 1.64 Å from the crystal coordinates and has an energy score of -14.27 kcal/mol. As indicated in Fig. 3.5b, the orientations of some modelled binding-site residues (mainly those which are mutated compared to the original binding pocket of trypsin) are distributed over a large area. Nevertheless, the mapped configuration space for ligand orientations is rather restricted since all generated solutions cluster about the native orientation.

Following the strategy outlined above, subsequently new factor Xa models were generated explicitly considering ligand information. RPR128515 (see Fig. 3.5a)[Maignan et al., 2000] was included as additional restraint in the protein modelling process. Taking the crystal structure of bovine trypsin (1f0u)[Maignan et al., 2000] with the ligand in its co-crystallised orientation as template, 100 new factor Xa models were generated (Fig. 3.5c). Again, side-chain conformational space is considerably mapped, however,

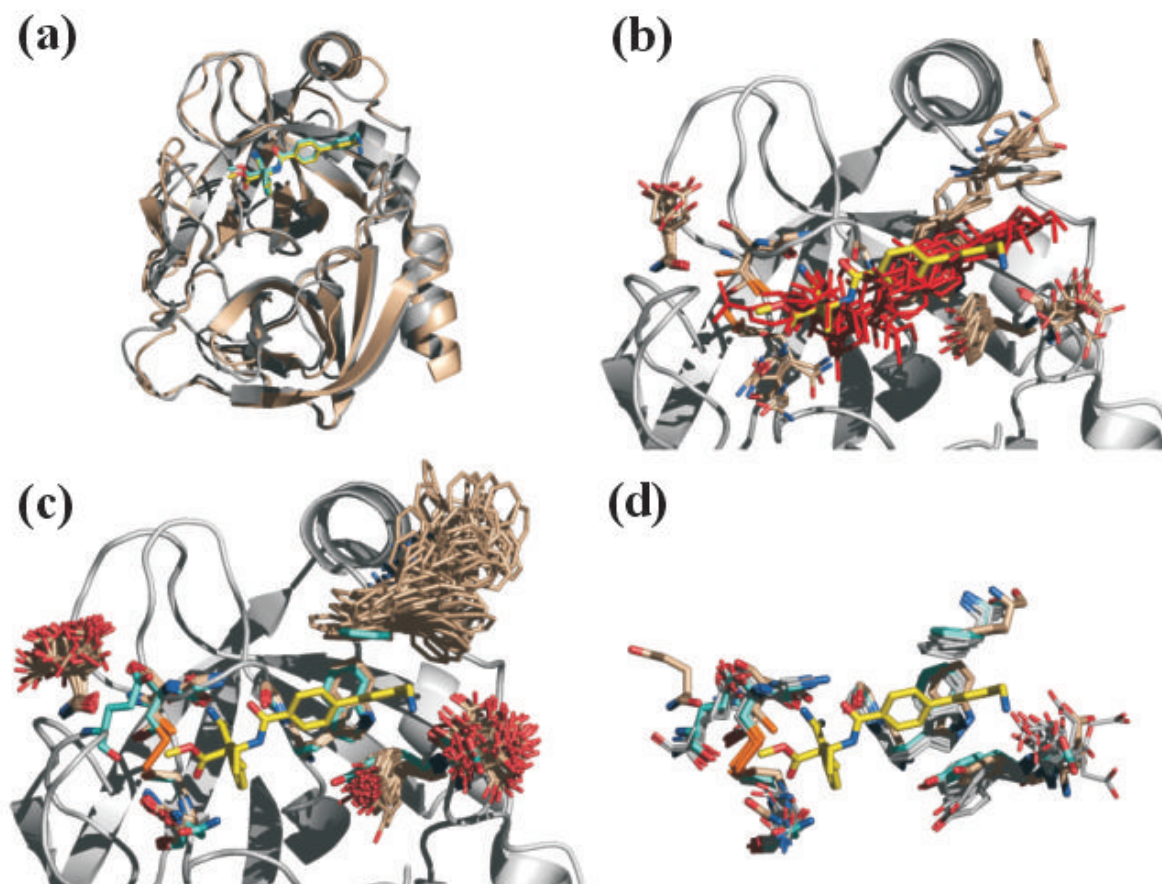


Figure 3.5: Superimposed crystal structures of trypsin (beige) and factor Xa (grey), both complexed with the ligand RPR128515 (a). In (b), the ensemble of relevant binding-site residues of factor Xa (modelled without considering ligand information using trypsin as template), together with the backbone of the crystal structure of factor Xa is shown. The ligand is depicted in yellow (native orientation known from the crystal structure) and red (solutions from docking into the ensemble of the homology models). Part (c) shows 100 new homology models of factor Xa that were generated based on trypsin regarding the ligand during the modelling process. The finally optimised binding-site model (generated by combining side-chains from different models) is shown in beige in (d), together with structures of factor Xa crystallised with RPR128515 (1ezq, grey) or with other ligands (cyan).

solutions penetrating into the ligand hardly occur. A final factor Xa model was obtained by combining rotamers retrieved from different homology models (Fig. 3.5d) (beige). For comparison, the crystal structure of factor Xa with bound RPR128515 (1ezq, cyan) and nine other crystal structures of factor Xa crystallised with different ligands (grey) are also shown [Maignan et al., 2000; Nar et al., 2001; Guertin et al., 2002; Kamata et al., 1998; Adler et al., 2000; Brandstetter et al., 1996]. Although the

rmsd between final model and crystal structure (1ezq) amounts to 1.66 Å, the features primarily responsible for binding are well reproduced, apart from Glu147, Gln192 and Glu97 which do not perfectly align with the crystal structure. In case of Glu147, this is due to the fact that the backbones of factor Xa and trypsin do not align in this area. However, as there is no specific interaction between the ligand and Glu147, this deviation is of no relevance. The ester group of the ligand forms an H-bond to Gln192-NH. Since the backbone traces match well in the template and the model, deviations in side-chain orientation of Gln192 are not important for the ligand pose. The same holds for Glu97, which establishes a strong H-bond (2.5 Å) through its backbone carbonyl oxygen and an amino group of the ligand. The other residues, in particular Tyr99 and Phe174 which contribute to binding (among others) and determine the specificity of the S4 pocket in factor Xa [Adler et al., 2000], are almost perfectly modelled. Regarding the fact that some of the discussed binding-site residues in factor Xa exhibit considerable side-chain flexibility upon binding of different ligands as indicated by multiple structure determinations (see Fig. 3.5d), the generated model appears rather convincing.

To assess whether the generated binding-site model could be used successfully for virtual screening, we tried to reproduce the binding mode of ten ligands which have been co-crystallised with factor Xa [Maignan et al., 2000; Nar et al., 2001; Guertin et al., 2002; Kamata et al., 1998; Adler et al., 2000; Brandstetter et al., 1996]. For reasons of comparison, we also docked these ligands into the binding pocket of a crystallographically determined factor Xa structure (1ezq).

The results with respect to rmsd and AutoDock energy score are listed in Table 3.4. The overall success rate is slightly higher while docking into the factor Xa crystal structure. Considering the solutions with the lowest rmsd value with respect to the experimental structure, in 6 (out of 10) cases a better solution is obtained while docking into the crystal structure instead of our model. However, the differences expressed in terms of rmsd are not large, in particular, taking the grid approximation within AutoDock and positional uncertainties in the experimentally determined structures into consideration. Also, the differences in energy scores are negligible. Only in 2 cases, they amount to more than 0.5 kcal/mol. Remarkably, docking RPR128515 either back into our model (rmsd: 0.33 Å, score: -17.24 kcal/mol) or into the crystal structure (rmsd: 0.78 Å, score: -16.17 kcal/mol) reveals a better result for the model. This shows, not unexpectedly, that the model is slightly tailored towards the ligand used to restrain the modelling. Nevertheless, since convincing results are obtained for all considered

Table 3.4: Statistics on the docking experiment on factor Xa

ligand pdb code ^{c)}	Crystal structure ^{a)}		Model ^{b)}	
	rmsd (Å)	energy (kcal/mol)	rmsd (Å)	energy (kcal/mol)
1EZQ	0.78	-16.17	0.33	-17.24
1F0R	1.81	-15.08	2.04	-15.35
1F0S	1.15	-13.50	1.75	-13.51
1FAX	1.98	-15.67	1.89	-15.20
1FJS	1.57	-15.48	2.21	-16.35
1G2L	1.96	-15.70	1.95	-15.76
1G2M	1.98	-15.03	1.83	-15.40
1KSN	1.03	-15.99	1.28	-16.32
1XKA	1.88	-14.71	2.38	-15.12
1XKB	1.90	-14.27	2.53	-14.30

All values refer to the least deviating solution with respect to the crystal structure.

^{a)} Results for docking the ligands into the crystal structure of factor Xa (1ezq).

^{b)} Results for docking the ligands into the homology model of factor Xa.

^{c)} Data set of 10 factor Xa ligands.

ligands, the generated model appears to be well suited for structure-based drug design purposes.

Modelling aldose reductase based on aldehyde reductase

The previous case study demonstrated that our approach generates sufficiently accurate geometries of protein residues to establish specific interactions with ligands. In the following example, we will investigate how well binding modes can be reproduced for a protein known from crystal structure analysis to exhibit pronounced induced-fit adaptations upon ligand binding.

Aldose reductase (AR), an NADPH-dependent enzyme, catalyses the reduction of glucose along the sorbitol pathway, and, therefore, represents a promising drug target in diabetes therapy of secondary complication.

AR shares 49.5 % sequence identity with aldehyde reductase. In particular, the co-

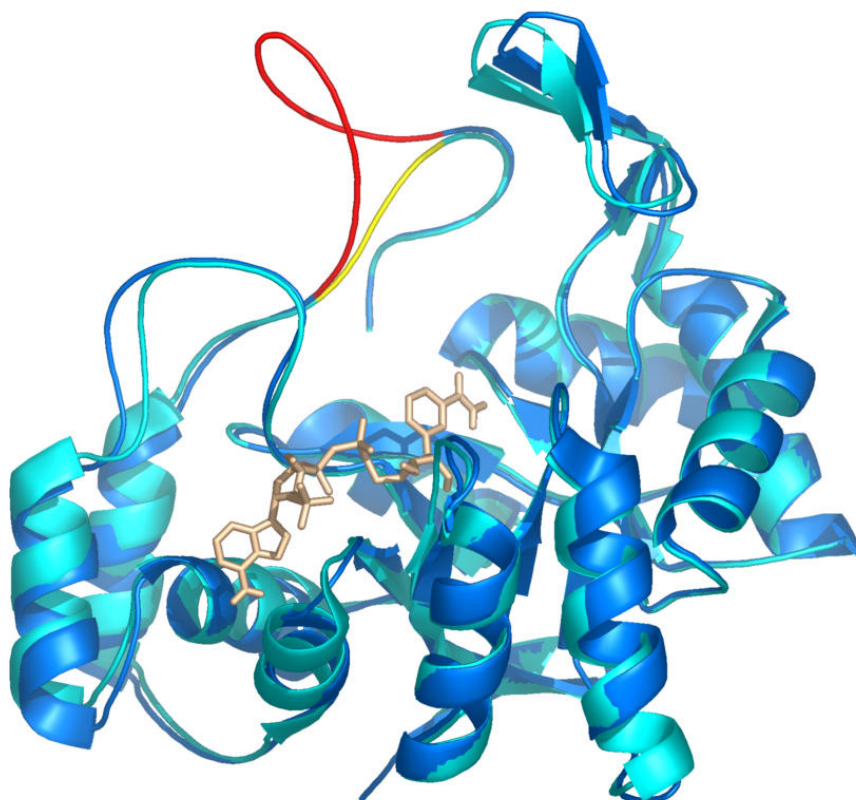


Figure 3.6: Superimposed crystal structures of aldose reductase (AR, cyan) and aldehyde reductase (marine) with the NADP+ cofactor (shown in beige) in its orientation from aldehyde reductase. The loop regions composing the specificity pockets are coloured yellow (AR) and red (aldehyde reductase), respectively.

factor binding-sites and the regions where the hydrid transfer from NADPH to the carbonyl carbon of the substrates occur (anion binding pocket), are structurally highly conserved (see Fig. 3.6). However, aldehyde reductase exhibits an additional loop, comprising 11 residues, that is responsible for differences in substrate specificity. Interestingly, in AR, this segment, is composed of only 4 residues (Ala299-Ser302). Here, it is part of the hydrophobic specificity pocket and shows the most striking adaptations upon ligand binding. An MD simulation performed on the ultra-high resolution crystal structure of human aldose reductase complexed with IDD59497 revealed the most pronounced flexibility in this region with the largest side-chain mobility exhibited by Leu300.[Sotriffer et al., 2003] A very distinct binding-site conformation (compared to the IDD594 complex) is observed for tolrestat binding (1ah3) to the porcine enzyme. Superimposition with the IDD594 complex (Fig. 3.7) reveals identical orientations of the ligand's carboxylates in the anion binding pocket, whereas tolrestat would clash into Leu300 in the IDD594 structure.

To examine whether these specific binding-site geometries could be modelled by the *MOBILE* approach, two different sets of AR models were generated including either tolrestat (1ah0) or IDD594 as ligand-derived restraints. According to our strategy, we initially generated 100 preliminary AR models based on the crystal structure of aldehyde reductase (1hqt) neglecting ligand information. The coordinates of the co-factor (being identical in AR and aldehyde reductase) were transferred from aldehyde reductase to the AR models. Next, we placed tolrestat and IDD594 into the ensemble of preliminary homology models using AutoDock. In case of tolrestat, a good docking solution (2.05 Å rmsd) was found on rank 2. For the IDD594 complex, a solution with 2.53 Å rmsd was obtained on rank 3. To further refine the modelled complexes, we performed an additional iteration of our approach. Therefore, the obtained ligand orientations of tolrestat and IDD594 were used to restrain the subsequent homology modelling. Considering each of the docked inhibitors separately, two sets of 100 homology models based on aldehyde reductase (1hqt) as template were generated. Next, we docked the two ligands into the ensembles of the produced protein models. For tolrestat, a solution with 0.84 Å rmsd with respect to the orientation observed in the crystal structure was obtained (found on rank 3), the best solution for IDD594 had 1.22 Å rmsd (also on rank 3). Compared to the respective rms deviations for the docking into the preliminary homology models (as shown above: 2.05 Å (tolrestat) and 2.53 Å (IDD594)), these improvements for the docking into the refined models are strongly significant.

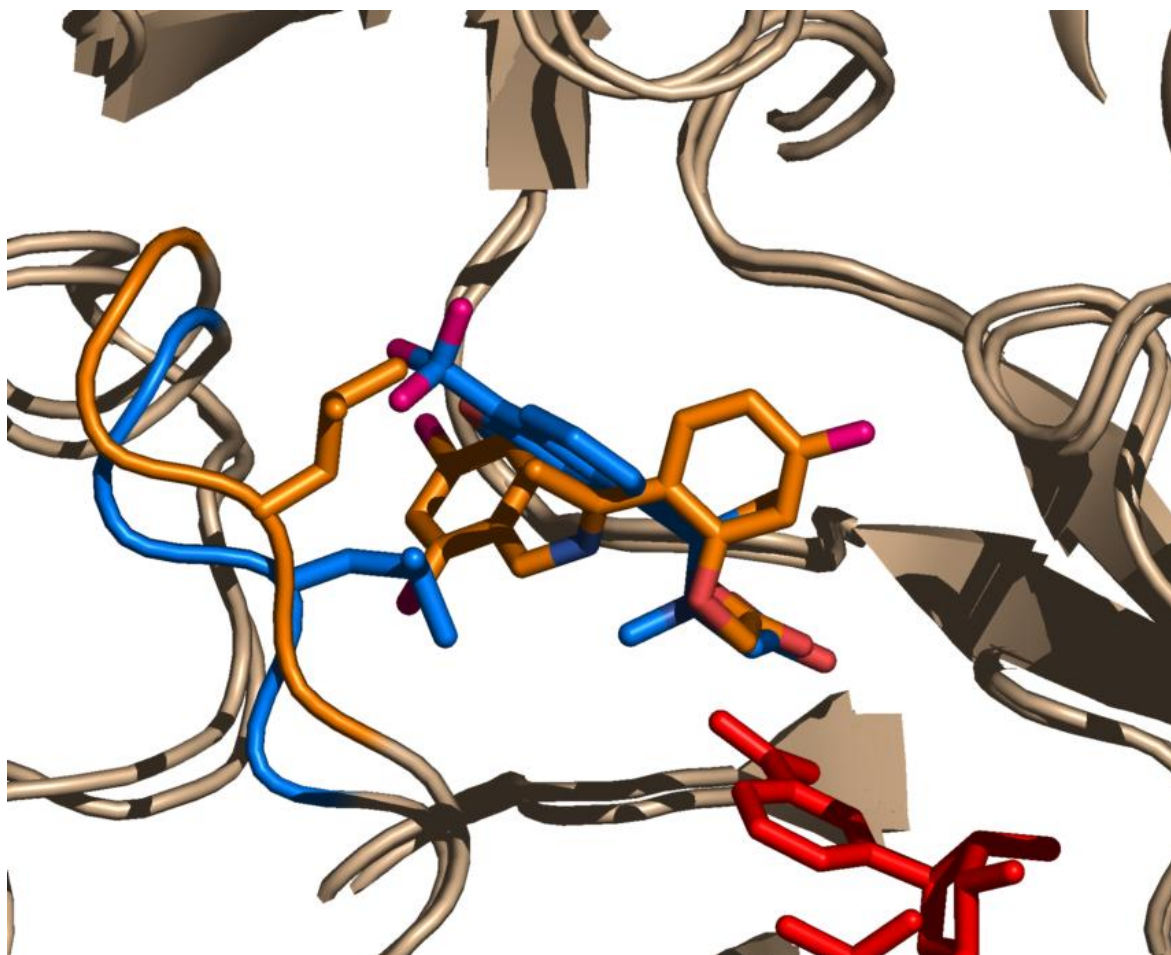


Figure 3.7: Conformational changes in the AR binding pocket in consequence of inhibitor binding. The nicotinamide ring of the cofactor is shown in red. In blue, the orientations of tolrestat and the side chains of Leu300 are displayed (as observed in the corresponding crystal structure 1ah3), the latter residue is mainly affected by the conformational rearrangement of the binding pocket upon ligand binding. The ligand IDD594, together with the obtained geometry of the corresponding binding-site residue, is depicted in orange.

Besides producing a near-native ligand geometry, the prime interest is focused on prediction of a correct loop geometry since the remaining part of the binding pockets in AR and aldehyde reductase are rather similar. Accordingly, we scored only the interactions formed between both docked ligands and the residues in the sequence stretch Ala299-Ser302 of the generated models using DrugScore. For both cases, loop geometries closely approximating the crystal structures (a comparison is shown in Fig. 3.8a and 3.8b) were found among the top-scored solutions. In case of tolrestat, the most convincing loop orientation was found on rank 2 (rmsd considering the side-chain atoms of Leu300: 1.22 Å), for IDD594 the loop conformer on rank 2 deviates by 1.49 Å.

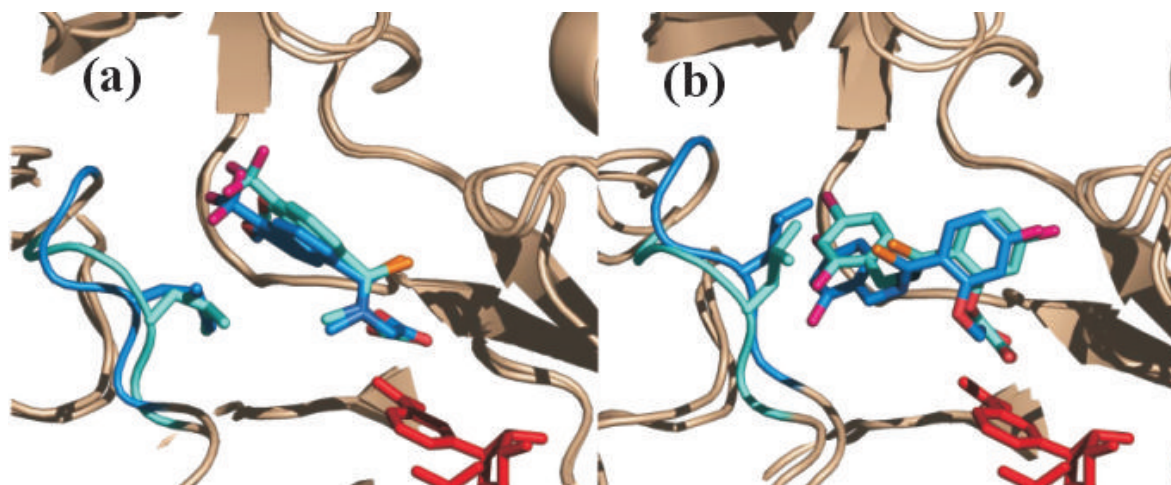


Figure 3.8: Superimposition of the crystal structures (blue) and the modelled complexes (cyan) of AR with (a) tolrestat and (b) IDD594. The side-chain orientations of Leu300 are indicated. The nicotinamide ring of the cofactor is shown in red.

AR provides an example for ligand-induced protein-adaptations affecting even the backbone conformation. This case study demonstrates that realistic protein-ligand geometries can be generated by applying the *MOBILE* approach to this rather complex system, where ligands reinforce different loop conformations upon binding. Furthermore, we have shown that the mutual orientations between the protein and a particular ligand can be adjusted in a stepwise fashion. Even though the initial starting protein-ligand geometries deviated considerably from the orientations found in the referring crystal structures, near-native geometries could be generated for both, the tolrestat- and the IDD594-complex after performing a second cycle of the *MOBILE* approach.

3.3 Summary and Outlook

In this contribution, a novel strategy (MOBILE) is presented to consider information about the binding mode of bioactive ligands during the homology modelling process. It starts with a combined set of homology models, and ligands are placed into a crude binding-site representation via docking onto averaged property fields derived from knowledge-based potentials. Once the ligands are placed, a new set of homology models is generated. However, in this step, ligand information is considered as additional restraint in terms of knowledge-based pair potentials. Consulting a large ensemble of produced models exhibiting different side-chain rotamers for the binding-site residues, a composite picture is assembled considering the individually best scored rotamers with respect to the ligand. After a local force-field optimisation, the obtained binding-site models are used for flexible docking. As a result, protein binding-site models of higher accuracy and relevance are generated. The application of DrugScore pair potentials proved efficient to restrain the homology modelling process and to score and optimise the modelled protein-ligand complexes. This was demonstrated using a test data set of 46 complexes and further validated by applying the new strategy to relevant modelling scenarios.

A commonly applied protocol for modelling binding-sites of unknown proteins usually starts with generating one preliminary homology model of the uncomplexed protein, occasionally optimised by molecular dynamics. After placing a ligand into the modelled active site, usually the entire complex is subjected to a further refinement. As pointed out by Schonbrun *et al.*, [Schonbrun et al., 2002] it is debatable whether a refinement by molecular dynamics actually improves the predicted structure. Indeed, none of the top comparative modelling groups involved in CASP4 [Moult et al., 2001] used such protocols, probably because previous experience did not suggest any advantage in predictive power. Most likely this is due to the limited sampling of configuration space using standard molecular dynamics although this limitation could be overcome by generalised-ensemble simulations. Accordingly, it is rather unlikely that the global minimum geometry of a protein-ligand complex can be obtained if the starting geometry is distant from the near-native one and separated by a high energy barrier. Supposedly, this limitation does not occur in our approach because the configuration space of the modelled binding pocket is more exhaustively sampled simultaneously considering the ligand in a near-native orientation. The global minimum is approximated as composite picture by identifying optimally scored rotamers from a large set of gen-

erated models. The individual rotamers are scored with respect to a given ligand pose using DrugScore. This function has been demonstrated to identify efficiently native and near-native protein-ligand configurations.[Gohlke et al., 2000a]

Similar to protein-ligand docking, the strategy to detect near-native complex geometries involves two equally important steps: (1) computing relevant geometries and (2) identifying the pose being most close to the experimentally given situation (scoring). The program MODELLER used in our approach produces relevant geometries of protein binding-sites even in the absence of a placed ligand, particularly, if the search space for side-chain rotamers is small and, thus, can be sampled efficiently. However, our approach also shows that the efficiency and accuracy of the modelling process is clearly enhanced by considering ligand information. The second goal - identifying complexes with near-native geometries - inevitably requires the presence of a ligand in a realistic orientation.

It has been shown that relevant binding modes can be produced by docking ligands into ensembles of protein structures.[Claussen et al., 2001] Smoothing the potential energy surface results in an even faster convergence of the docking problem. Nevertheless, due to the approximate nature of the binding-site representations derived from an ensemble of modelled protein geometries, the use of conformationally restricted ligands is advisable. If a 3D superposition of ligands, e.g. in the context of a previously performed 3D-QSAR study, is available, these aligned ligands could be docked rigidly into the homology models. This will further reduce the search space of the docking problem. The mutual similarity of different ligands in their docked orientations can be used as an additional criterion to assess the quality of the docking solutions.[Schafferhans & Klebe, 2001]

In the presented approach, ligand information is only used in structural terms. Additionally, affinity data of the ligands might be considered to assess the quality of the generated homology models. Such concepts will result in a "QSAR-refined homology modelling". The first option would be to use a given set of ligands, docked into several homology models, and the affinity of all resulting complexes is predicted. The model that yields the best correlation between calculated and experimental affinities is rendered prominent. A possible limitation of this strategy might be that the presently available scoring functions cannot predict affinities accurately enough. Interestingly, 3D-QSAR models based on superimposed ligands reveal surprisingly high predictive power in affinity estimation, provided a correct superimposition is given. In conse-

quence, a second alternative to assess the quality of the produced models would be to generate multiple QSAR models based on distinct ligand alignments obtained from the docking into the various homology models. In analogy to the procedure followed by several authors [Jalaie & Erickson, 2000; Schafferhans & Klebe, 2001; Kim, 1998], the statistical significance of the generated QSAR models is used to assess the relevance of the different protein models. A further possibility to reliably predict the affinities between homology models and ligands would be to establish an AFMoC model. [Gohlke & Klebe, 2002b] AFMoC tailors protein-specifically adopted DrugScore pair-potentials to one particular protein by considering additional ligand-based information in a CoMFA-type approach. The statistical significance of an AFMoC model thus explicitly reflects the quality of the underlying protein model. A further advantage is that AFMoC allows the user to gradually move from general knowledge-based potentials to protein-specifically adopted ones, depending on the confidence in the generated protein model and the amount of ligand data available for training.

To assess the predictive power of protein homology modelling techniques, usually the rmsd between the model-built and the corresponding crystal structure is determined. In the present contribution, we follow the same procedure. However, one has to regard intrinsic accuracy limits. X-ray structures obtained for the same protein in different laboratories or determined in two different crystal forms can show deviations in main-chain atoms of about 0.5 Å rmsd. The solvent-exposed side-chains can differ by as much as 1.5 Å, while for more buried side-chains, the difference can amount to 1.0 Å [Levitt et al., 1997]. Exploring the theoretical prediction limit of commonly applied force-fields, Petrella et al. suggested a limit for side-chain prediction of 0.8 Å [Petrella et al., 1998]. Xiang et al. assume accuracy limits of 0.7 Å for the side-chains of core residues [Xiang & Honig, 2001]. In light of these estimates, the accuracies achieved by our approach on the test set for binding-site residues (1.0 Å) are quite convincing. This becomes even more pronounced when considering that in the two above-mentioned studies, all residues were kept fixed except the one being predicted, whereas in our approach the orientations of all protein side-chains in the active site are predicted simultaneously. Finally, as noted by Tramontano et al. [Tramontano et al., 2001], the rmsd criterion is widely accepted, but not necessarily always a perfect figure-of-merit. Criteria that rank proper side-chain orientations with respect to neighbouring side-chains would be more conclusive. In particular, this is important while evaluating the side-chain orientations of a protein with respect to a preoriented ligand. In our approach, DrugScore convincingly supports in particular this step.

For proteins exhibiting pronounced induced-fit adaptations, homology modelling based on a single crystal structure of a related protein is difficult and results could be misleading. Even so crystal structures are our most reliable source to learn about protein geometry, they only provide a frozen snap-shot of a dynamically fluctuating system. Local effects such as the applied pH conditions or impacts imposed by crystal packing do influence binding modes [Stubbs et al., 2002]. Through ligand binding, different local minima experienced by the uncomplexed protein under dynamic conditions can be stabilised and observed as favourable binding-site geometries in a crystal [Sotriffer et al., 2003].

Homology modelling using MODELLER is based on a reference template structure and the approach tries to carry over as much information as possible from the template into the model, in particular in regions of high sequence identity and structural conservation. To perform an exhaustive side-chain screening by our approach, such regions must be excluded from the direct homology matching step. As an alternative, structural variability can also be introduced in the modelling process by considering multiple template structures exhibiting deviating conformations in the flexible regions.

3.4 Materials and Methods

3.4.1 Test data set

A test data set of 46 protein ligand complexes was compiled to validate the performance of our approach (see Table 3.5). This set has been extracted from the 91 protein-ligand complexes used for DrugScore validation [Gohlke et al., 2000a], considering the following criteria: (1) As MODELLER is primarily intended for homology modelling, we only selected structures which do not contain cofactors next to the active site. (2) The only hetero atoms allowed (besides those in the ligands) were metal ions. (3) We also eliminated all water molecules as their positions will generally not (yet) be predicted realistically in modelling scenarios. We used this test data set for scaling the DrugScore potentials to the MODELLER force-field, comparing homology models generated with and without ligand information, assessing the side-chain prediction accuracy of MODELLER, and evaluating DrugScore’s power to identify near-native complex geometries.

Table 3.5: PDB codes of proteins used for the parameterisation of the pair potentials

121P	1ABE	1ABF	1ACJ	1AHA	1APT	1ATL
1AZM	1BBP	1BLH	1BUG	1BYB	1CBX	1CIL
1CPS	1CTR	1DID	1DIE	1ELA	1EPB	1F3E
1HDC	1HEF	1HFC	1HSL	1HYT	1ICN	1IMB
1LAH	1LMO	1LNA	1MLD	1MRG	1MRK	1POC
1PPL	1PSO	1RBP	1RDS	1RNT	1ROB	1SNC
1SRJ	1STP	1TLP	1TMN			

3.4.2 Generation of binding-site models of the test data set

For validation studies performed with the test data set (see Table 3.5), only the geometries of residues next to the active site (within a distance of 4.5 Å to the ligand) were modelled. Here, the crystal structure of the respective PDB entry served itself as template for the modelling process (Fig. 3.9). The coordinates of all but the binding-site residues of the protein were kept unchanged with respect to the templates. New

geometries for binding-site residues (including side-chain and main-chain atoms) were forced to be generated by MODELLER by keeping the binding-site residues unmatched in the sequence alignment (see Fig. 3.9, panel C). For the de novo prediction of side-chain geometries only, the referring residues in the template structures were mutated to Gly (see Fig. 3.9, panel B).

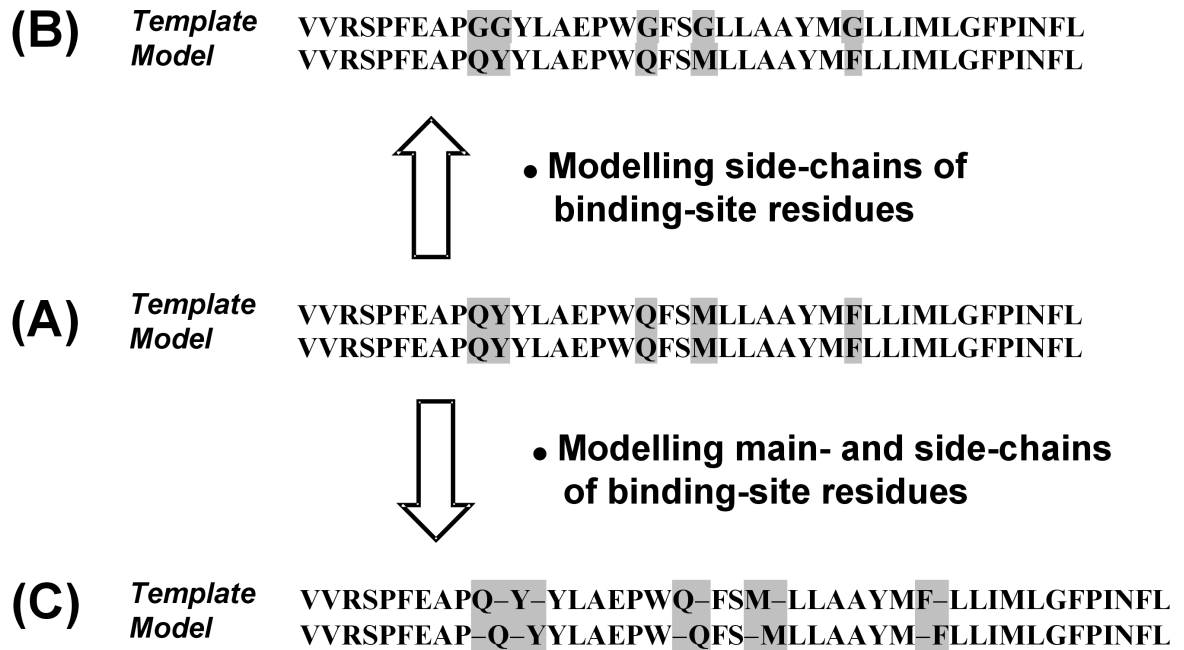


Figure 3.9: Generation of binding-site models for the proteins of the test set (Table 3.5). An aligned sequence stretch of a protein is shown in panel A. Five residues, belonging to the active site, are shaded grey. The procedure for generating binding-site models works as follows: The available structure of the *Template* protein serves as basis to model the structure of the *Model* sequence. Generating a homology model with the identically aligned sequences as represented in panel A would result in a *Model* structure that is identical to the *Template*. To generate a *Model* with new **side-chain** orientations of the binding-site residues (while keeping the remaining part of the protein as in the *Template* structure), sequence alignment (panel B) is used. Here, the binding-site residues in the *Template* structure are mutated to Gly. Thus, information about the respective side-chain coordinates cannot be inferred by homology and must be predicted *de novo* by MODELLER. For modelling the **main- and side-chain** orientations of the Model, the binding-site residues are unaligned in the sequence alignment, thus, leading to a complete neglect of information in MODELLER about these residues from the *Template* structure.

3.4.3 Scaling the DrugScore potentials with respect to the MODELLER force-field

In order to incorporate the DrugScore pair potentials as additional restraints into the MODELLER force-field, they were empirically scaled with respect to the remaining force-field terms. Apart from the terms adapted from CHARMM to constrain stereochemical properties, the MODELLER force-field consists of probability density functions of purely empirical origin. Therefore, it appears justified to weight the DrugScore potentials empirically. This adjustment was accomplished using our test data set of 46 complexes (see Table 3.5) by systematically varying the contribution of the DrugScore pair potentials to the MODELLER molecular probability density function. For this parameterisation study, the coordinates of the ligand atoms were adapted from the referring PDB entries. For each parameter setting (i.e. scaling factor), 10 homology models were generated. In these "models", all coordinates apart from residues within 4.5 Å distance to the ligand were kept identical to the crystal coordinates. The binding-site residues, however, were generated *de novo* by MODELLER, i.e. without considering information taken from the template structure. The models generated using the actual parameter setting were assessed with respect to their spatial deviation from the corresponding crystal structures by computing the rmsd between the residues of the modelled and crystallographically determined binding-site. As a further criterion to consider the similarity between model and crystal structure, grids based on DrugScore potentials were calculated in the modelled and crystallographic binding-sites and their mutual similarity has been assessed by evaluating the Hodgkin index [Hodgkin & Richards, 1987]. A scaling factor of $7.5 \bullet 10^{-5}$ was finally found as best solution to adjust the DrugScore potentials to the MODELLER force-field. A similar scaling factor was obtained by Sotriffer *et al.* scaling DrugScore to the intramolecular force-field implemented into AutoDock [Sotriffer *et al.*, 2002b].

4 Generation and validation of a ligand-supported homology model of the neurokinin-1 receptor by virtual screening for a submicromolar inhibitor

4.1 Biological target system

4.1.1 G-protein-coupled receptors

Introduction

G-protein-coupled receptors (GPCRs) form one of the largest superfamilies of cell-surface receptors mediating responses to diverse signals, for example, visual, olfactory, hormonal, and neurotransmitter signals [Marinissen & Gutkind, 2001]. A large percentage of prescription drugs (30 % of the top 50 sellers in 2001, 50 % of all currently launched drugs) target one or more GPCRs, with most major therapeutic areas being served to some extent by several GPCR-based drugs [Scussa, 2002]. Currently, there are more than 3000 sequences of this family known for different organisms, and GPCR genes correspond to 3 % of the genes in humans. The superfamily of GPCRs is subdivided into three main receptor subfamilies on the basis of the pharmacological nature of their ligands and sequence similarity. The division includes the class I (rhodopsin like) receptors (their ligands are biogenic amines, neuropeptides, chemokines and prostanoids), the class II receptors (secretin like) and the class III receptors (metabotropic glutamate receptor like), which represent about 89 %, 7 % and 4 %, respectively, of the known GPCRs [Menzaghi et al., 2002].

GPCRs play a key role in a whole series of processes in our body, establishing a functional and unidirectional link between the exterior of a cell and its cytoplasm [Baldwin, 1994; Nederkoorn et al., 1995; Oliveira et al., 1993; van Rhee & Jacobson, 1996; Selbie & Hill, 1998]. When a ligand interacts with the extracellular half of the GPCR, a signal is generated which is transmitted through the membrane to the



Figure 4.1: Family classification of G-protein coupled receptors. From [Chalmers & Behan, 2002].

cytosolic side where a G-protein is activated. This G-protein in turn activates one or more of a variety of secondary messengers. The mediation of the ligand action occurs through a conformational change of the GPCR [Oliveira et al., 1999; Birnbaumer & Birnbaumer, 1995]. A simple model suggests that GPCRs exist in an equilibrium of active and inactive states, R^* and R , respectively [Costa & Hertz, 1989; Samama et al., 1993; Lefkowitz et al., 1993]. Only R^* can bind to the G-protein and produce a cellular response. Under physiological conditions, the ligand-receptor interaction results in an increase in the ligand bound receptor LR^* , resulting in the production of an LR^* -G-protein complex and consequent cellular response.

GPCR Structural Information

It has long been known that GPCRs share a central core domain constituted of seven transmembrane helices (TM-I through -VII) connected by three intracellular (i1, i2 and i3) and three extracellular (e1, e2 and e3) loops (see Fig. 4.2) [Baldwin, 1993]. Albeit their common structural features, GPCRs do not share any overall sequence homology [Kolakowski Jr, 1994; Probst et al., 1992]. Significant sequence homology is found, however, within several subfamilies [Gether, 2000]. Despite the divergent overall sequence homology of GPCRs, the antagonist binding-site is often located in the same region where bovine rhodopsin binds the retinal ligand [Böhm et al., 1996]. On the other hand, it was shown by Jacoby *et al.* that for the monoamine GPCR $5HT_{1A}$ three distinct binding sites exist [Jacoby et al., 1999; Jacoby, 2001]. This was also suggested for other monoamine GPCRs.

A cartoon diagram of the recently resolved high-resolution structure of rhodopsin [Palczewski et al., 2000; Teller et al., 2001] is depicted in Fig. 4.3. The retinylidene chromophore (coloured cyan), which is covalently bound to the receptor, is located more toward the extracellular boundary of the plane of the putative membrane bilayer. Absorption of a photon by 11-cis-retinal causes its isomerisation to all-trans-retinal, initiating a conformational change in the receptor that leads to its activation. One of the most striking features of the structure of rhodopsin is the presence and positioning of the β 4-hairpin (coloured red in Fig. 4.3) within the second extracellular loop which holds the chromophore very firmly in place by many contacts. The β 4-hairpin is stabilised by a disulfide bond between Cys110 and Cys187 which is highly conserved in the family of the GPCRs [Sakmar, 2002]. The role of the E2 loop in rhodopsin and the other GPCRs is not known and still a question of debate. Experiments in

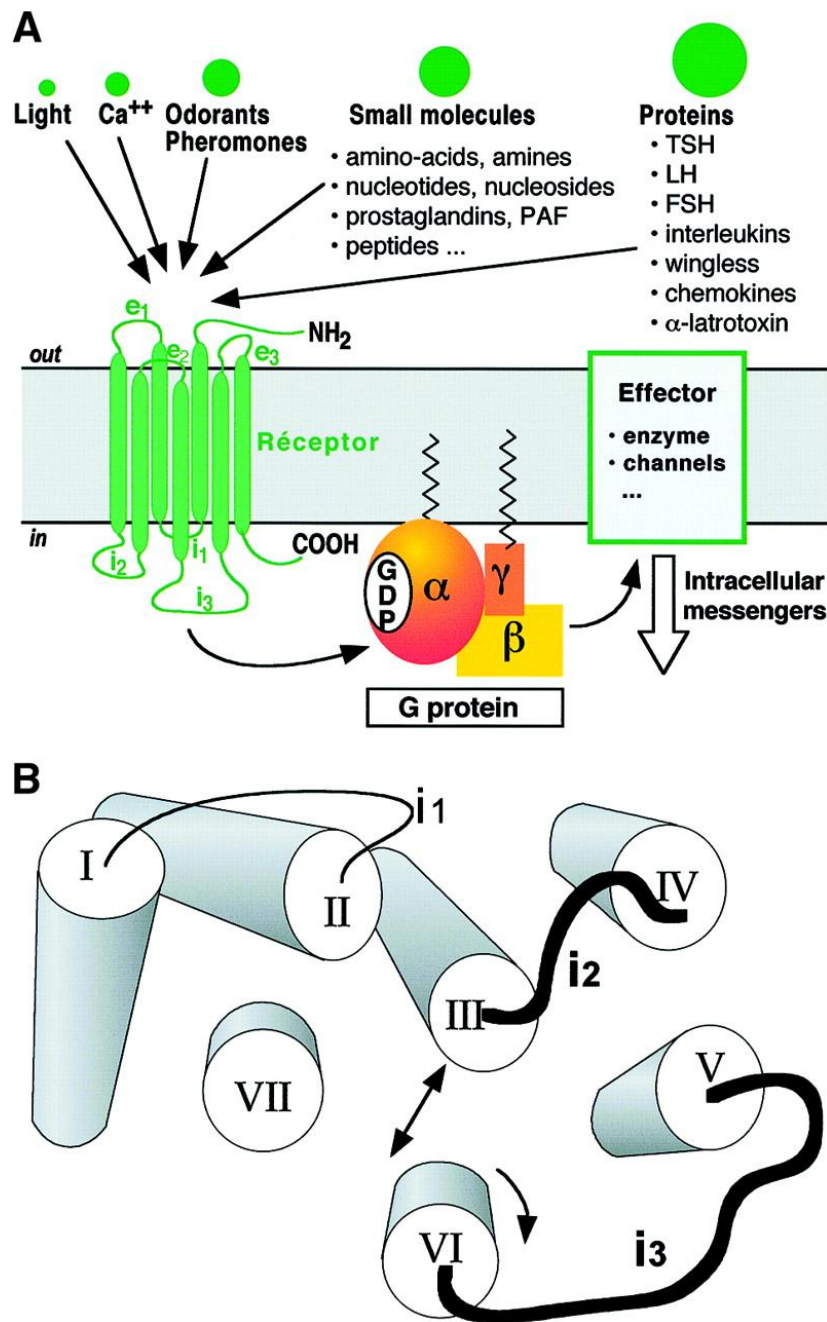


Figure 4.2: (A) GPCRs comprise a central common core composed of seven trans-membrane helices (TM-I to -VII) connected by three intracellular (i1, i2, i3) and three extracellular (e1, e2, e3) loops. The diversity of messages which activate these receptors is an illustration for their evolutionary success. (B) **Illustration of the central core of rhodopsin, viewed from the cytoplasm.** The core is represented under its 'active conformation'. The TM-VI and -VII lean out of the structure, the TM-VI turn by 30 % on its axis (clockwise as viewed from the cytoplasm) [Bourne, 1997]. This opens a cleft in the central core in which G proteins can find their way. i2 and i3 loops are the two main loops engaged in G-protein recognition and activation. From [Bockaert & Pin, 1999]

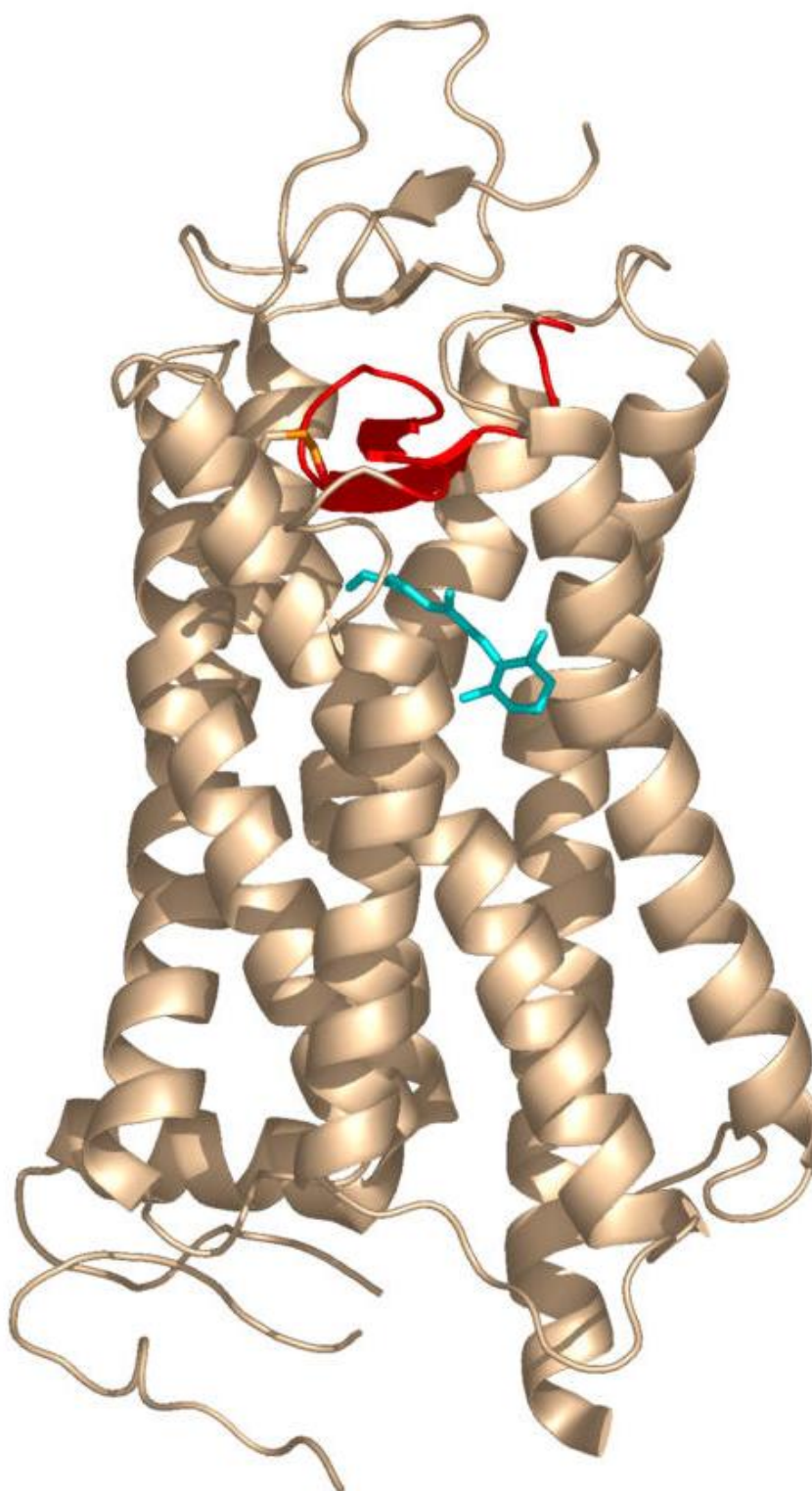


Figure 4.3: Cartoon representation of the high-resolution X-ray structure of bovine rhodopsin. The E2 loop which contains the β 4-hairpin is coloured red. It is in close contact to the retinal ligand (coloured cyan).

the laboratory of Sakmar *et al.* indicate that its main role might be to regulate the stability of the active state of the receptor in which the all-trans chromophore acts, suggesting that the E2 loop in other GPCRs also might provide contacts with agonist ligands [Sakmar, 2002]. The presence of the $\beta 4$ -hairpin raises the question how the ligand gets from the external medium into the binding pocket. Oliveira *et al.* suggest the following two options [Oliveira et al., 2002]:

1. Through the space between the helices VII, I and II, and the beta turn 182-185 in the IV-V hairpin, through a space flanked by the 178-182 strand, the part of the IV-V hairpin directly after Cys187, and the helices VI-VII
2. Motion of the IV-V hairpin leaves the entry to the binding cavity open long enough for a ligand to enter

The authors furthermore note that the $\beta 4$ -hairpin sequence is highly variable among the family GPCRs. It is therefore unlikely that any function can be found in this area that is common to many receptors. Accordingly, it is questionable if the $\beta 4$ -hairpin in the orientation observed in bovine rhodopsin provides a reasonable structural template for homology models of other GPCRs.

Considering the fact that the global sequence identity between bovine rhodopsin and other GPCRs is generally below 25 %, might lead to the conclusion that rhodopsin might not provide a proper template for homology modelling. Normally, when the sequence identity between the model and the template is below 30 %, the sequence alignment is the main bottleneck in the modelling procedure. GPCRs form an exception to this rule [Oliveira et al., 2002]. Each helix contains one or more highly conserved residues that allow for an unambiguous alignment of the helices of the model and the bovine rhodopsin template. Therefore, it seems reasonable to assume that the geometry of the transmembrane region is very similar among GPCRs.

Until the high-resolution X-ray structure of bovine rhodopsin became available [Palczewski et al., 2000; Teller et al., 2001; Okada et al., 2002], inferences about the structure of GPCRs have been based on cryo-microscopy studies of rhodopsin [Unger & Schertler, 1995; Schertler & Hargrave, 1995; Davies et al., 1996] and the high-resolution structures of bacteriorhodopsin [Henderson et al., 1990; Pebay-Peyroula et al., 1997; Luecke et al., 1999]. Recent pairwise comparisons of the high-resolution structures of rhodopsin and bacteriorhodopsin show that helices IV and V do not superimpose

[Teller et al., 2001]. The twists and kinks in the helices create substantial differences between the two proteins. Furthermore, bacteriorhodopsin, even though it belongs to the family of 7-TM receptors, does not couple through G-proteins and thus is not a member of the GPCR family. Before the high-resolution structure of bovine rhodopsin became available, a common approach for generating hypotheses about binding mechanisms was the generation of preliminary GPCR models based on bacteriorhodopsin, subsequently refined by experimental findings from structure-activity, mutagenesis and affinity labelling studies [Gershengorn & Osman, 2001]. Although models based on bacteriorhodopsin could thus be useful for studying the functional architecture of GPCRs, they are probably not reliable enough for precise structure-based ligand design.

GPCRs and drug design

Nowadays, the search and optimisation of lead structures for GPCRs predominantly relies on ligand-based drug design techniques. They usually start by establishing a pharmacophore model. If only limited information about ligands is available, such a model can be deduced directly from the natural ligand and its analogues. It can then be exploited for virtual screening. This technique has been applied successfully for the search of novel lead structures by researchers at Merck for the *sst* receptor [Yang et al., 1998] and Aventis for the urotensin II receptor [Flohr et al., 2002]. If information about other ligands is available, a common strategy is to spatially superimpose several structurally diverse, preferably rigid ligands to identify common features responsible and essential for binding the target receptor. These features are subsequently translated into a pharmacophore model used for virtual screening (for example, see [Marriott et al., 1999]). Another new successful strategy for finding novel leads is based on the fact that many GPCR ligands share common structural motifs, although binding to different receptors. Based on this observation, so-called *targeted libraries* for GPCRs have been compiled [Balakin et al., 2002, 2003]. These libraries are assembled and continuously enriched by compounds with structural features required to bind to members of the target family. Indeed, such preselected libraries provide significantly higher hit rates compared to the screening of randomly selected libraries.

Another conceivable option to discover novel leads for GPCRs would be a docking screen using a homology model of the target receptor. To our knowledge, such procedure has not yet been described in literature. This is possibly due to the fact that the sequence identity among the active sites of bovine rhodopsin and other GPCRs are

often too low to allow for reliable prediction of side-chain geometries at the binding-sites. Aspects concerning homology modelling of GPCRs for drug design have been discussed in recent review articles ([Klabunde & Hessler, 2002; Flower, 1999; Ballesteros et al., 2001; Gether, 2000]). Flower suggests that improvement of GPCR models for the purpose of structure-based drug design can be achieved by refining preliminary models using information from mutational data. Prediction accuracy can supposedly be optimised by combining QSAR with receptor modelling [Flower, 1999].

Indeed, the knowledge about ligands is often combined with is data of GPCRs, which is increasingly communicated via the internet, e.g. the GPCRDB [Horn et al., 1998], the GRAP Mutant Database [Beukers et al., 1999; Kristiansen et al., 1996; Edvardsen & Kristiansen, 1997], and the Olfactory Database [Crasto et al., 2002]. Ligand binding information on GPCRs and their effectors is available at the PDSP (Psychoactive Drug Screening Program) database [Roth et al., 2000]. Combining the receptor information with structure-activity data of ligands, the relevant amino acids and the composite interacting moieties of the ligands can be identified and, thus, provide a starting point for further optimisation of binding affinity.

Examples of GPCR models based on the crystal structure of bovine rhodopsin, refined or validated by ligand and mutational data, can be found in literature (for example [Jöhren & Hölte, 2002; Chambers & Nichols, 2002; Shim et al., 2003; Lopez-Rodriguez et al., 2001]). As mentioned above, no such homology model was ever used to screen for novel compounds, although Bissantz *et al.* recently demonstrated that homology models of GPCRs, based on bovine rhodopsin, are reliable enough to be used for virtual screening of chemical databases [Bissantz et al., 2003]. This was shown by successful retrieval of known antagonists of the dopamine D3 receptor, the muscarinic M1 receptor, and the vasopressin V1a receptor via docking from a database, which additionally included randomly collected drug-like compounds.

An important issue in homology modelling of GPCRs for the purpose of structure-based drug design is the fact that the bovine rhodopsin structure resolved by Palczewski *et al.* is that of the *inactive* state of rhodopsin. This justifies the usage of bovine rhodopsin as structural template as long as ligands are searched which preserve the addressed GPCR in its inactive conformation (i.e., antagonists).

The task of finding agonists for GPCRs represents a more demanding challenge because in bovine rhodopsin, the activation, i. e. the coupling with the G-protein involves an

alteration in the relative orientation of TM III and TM VI, with an accompanying rotation of TM VI (see Fig. 4.2B). These movements probably result in an alteration in the position of the third intracellular loop (i3), which uncovers residues related to G-protein coupling [Chalmers & Behan, 2002]. Once the molecular details about the activation mechanism are resolved, it could be possible to transfer these steps to other GPCRs, assuming that the activation process is similar among all GPCRs [Oliveira et al., 2002]. However, considering the divergence of the extracellular loops which interact with the natural ligands and taking into account the diversity of these ligands in terms of molecular size and chemical diversity, it is in question whether all GPCRs follow exactly the same activation mechanism, or whether the coupling with the G-protein can be achieved by slightly different structural rearrangements. In any case, as long as the molecular mechanism of activation at an atomic resolution is still unresolved, it is probably unreasonable to search for agonists based on such a homology model. Consequently, ligand-based drug design is probably a more promising strategy to discover new agonists for GPCRs. This assumption is supported by the docking results revealed by Bissantz *et al.*, who successfully identified GPCR antagonists from large compound databases via docking, but the models were not accurate enough for retrieving known agonists [Bissantz et al., 2003].

Homology Modelling of GPCRs

Due to the enormous importance of GPCRs as drug targets and to support structure-based drug design, several approaches have been developed especially with respect to the modelling of GPCRs.

- Shacham *et al.* presented a technology, named *PREDICT*, which models the 3D structure of any GPCR based on its amino acid sequence, without the use of a structural template [Shacham et al., 2001]. The modelling procedure reveals a low-energy conformation by optimising a model considering a large number of properties, including helical-packing geometry, multihelical tilts, helix orientations, sidechain rotamers, helix membrane-surface crossing, and helical kinks. The huge size of the protein conformational space is covered through a hierarchical design, starting with a coarse representation and gradually increasing its complexity until reaching a full atomistic model. *PREDICT* was capable of reproducing the experimental structure of bovine rhodopsin.

Furthermore, *PREDICT* has been incorporated into the *Predix Drug Discovery Platform* [Predix]. Within this platform, *PREDICT*-generated models of GPCRs have been used for computational screening of virtual molecular libraries. Using these models, the authors report hit rates of 85 to 100 % when screening for known binders and hit rates of 10 to 24 % for unknown binders (with an experimental binding affinity of $< 5 \mu\text{M}$) [Becker et al., 2003]. To the best of our knowledge, GPCR models generated by *PREDICT* are the only ones reported to yield novel GPCR binders identified by docking.

- Another method for predicting the structure of GPCRs (*MembStruk*), and the binding mode of ligands classified in terms of relative binding affinities (*HierDock*) was developed by Vaidehi *et al.* [Vaidehi et al., 2002]. Similar to *PREDICT*, *MembStruk* predicts 3D structures by using only the amino acid sequence of the target GPCR by applying a hierarchical modelling strategy. Starting with the prediction of the TM regions, the individual helices are constructed and oriented according to the 7.5 Å electron density map of bovine rhodopsin. After a coarse grain optimisation of the TM bundle, interhelical loops are added and the full structure is generated.

The *HierDock* ligand screening protocol also follows a hierarchical strategy to examine ligand binding conformations and calculating their binding energies. *MembStruk* and *HierDock* were successfully used for predicting the retinal-bound structure of bovine rhodopsin and applied to the $\beta 1$ -adrenergic receptor, endothelial differential gene 6, mouse and rat l7 olfactory receptors, and the human sweet receptor.

- An integrated GPCR modelling approach was introduced by Müller [Müller, 2000]. Based on a sequence alignment established by exhaustive sequence similarity searches over all sequence databases, an in-plane projection structure for the seven transmembrane helices is derived assisted by calculated vectorial property moments. After mapping this 2D topology representation onto the $C\alpha$ -trace suggested by Baldwin [Baldwin et al., 1997], side-chains are added in favourable conformation. Such models are refined by molecular dynamics simulation under explicit consideration of the non-isotropic environment for energetic relaxation. This procedure was applied to the human CCK-B receptor, followed by docking studies of nonpeptide antagonists utilising the DragHome concept introduced in section 2.2.1 [Escherich et al., 2001].

- Another program for building the transmembrane domains of GPCRs is *BUNDLE* [Filizola et al., 1998]. The following steps are involved in the construction of a GPCR model: After identifying the helices, their centres are arranged according to the low-electron density map of rhodopsin, followed by the computation of the tilt of each helix. After defining a local coordinate axis for each of the helices, they are oriented in an antiparallel fashion and rotated along their axes. In the next step, each helix is rotated along an axis perpendicular to the helical one; finally, each helix is translated to its centre deduced from the projection map. The described procedure was used to model rhodopsin and other GPCRs. At the time of publication, the low resolution crystal structure of bovine rhodopsin was not available, thus a detailed validation of generated models was not possible.

4.1.2 The Neurokinin Receptors

The tachykinins are a family of neuropeptides comprising substance P, neurokinin A and neurokinin B, which share the C-terminal sequence Phe-X-Gly-Leu-Met- NH_2 in common. The preferred receptors for these neuropeptides are named (respectively) neurokinin-1 (NK1), neurokinin-2 (NK2), and neurokinin-3 (NK3) (see Table 4.1). They belong to the family of the GPCRs and share a high sequence homology. For example, the human (h)NK3 receptor sequence has 74 % and 68 % homology with the hNK1 and hNK2 receptors, respectively, with maximal sequence conservation in the putative transmembrane regions and minimal identity at the amino- and carboxy-termini. Whereas NK1 and NK2 are widely distributed in the central nervous system (CNS) and peripheral tissue, NK3 seems to be stronger localised in the CNS. Tachykinins have been described to be implicated in numerous physiological and pathological processes such as neuronal modulation, plasma protein extravasation, mast cell degranulation, stimulation of mucus secretion, or neurotropic and mitogenic effects. Modulation of their functional properties is suggested for the treatment of at least 6 major disease groups including CNS disorders, pain, airway disease, urinary incontinence, emesis and intestinal disfunction [Patacchini & Maggi, 2001].

The percentage sequence identity between bovine rhodopsin and the NK1 receptor is 21 %, considering only the transmembrane regions, it rises to 27 %. In the case of soluble proteins, pronounced structural similarity is generally given among proteins with > 25% sequence identity, but below this level of sequence identity, structural

divergence increases rapidly [Wilson et al., 2000; Yang & Honig, 2000a]. Nevertheless, recent work has demonstrated that a number of different soluble proteins can achieve the same fold with different sequence patterns [Yang & Honig, 2000b]. This holds also for GPCRs [Oliveira et al., 2002]. In GPCRs, each helix contains one or more highly conserved residues that allow for an unambiguous alignment of the helices of the model and the bovine rhodopsin template. Therefore, it seems reasonable to assume that the transmembrane region is very similar among all GPCRs.

4.1.3 NK1 antagonists

Meanwhile, a variety of NK1 antagonists have been developed based on several diverse lead structures (for an overview, see [Giardina et al., 1997]). Although distinct in their chemical scaffolds, they are similar in their interaction geometries. A generally accepted pharmacophore model for non-peptidic NK1 antagonists is depicted in Fig. 4.4 [Boks et al., 1997]. It consists of at least two aromatic rings kept in fixed orientation by various scaffolds, and contains at least one hydrogen-bond acceptor.

Table 4.2 gives some examples for antagonists binding with high affinity to the NK1 receptor. In Fig. 4.4, the key pharmacophoric elements are coloured according to the given pharmacophore definition. Besides these pharmacophoric elements, a remarkable structure-activity relationship for the ligands binding with high affinity is given with respect to the substitution pattern at the phenyl rings. In compounds **4**, **5**, **6**, **7**, and **8** (Table 4.2), the phenyl ring is 3,5-bis(trifluoromethyl)-substituted. On the other hand, compounds **2** and **3** are 2-methoxylated. The substitution pattern at the phenyl ring seems to take a strong impact on binding affinity.

The influence of an appropriate phenyl substitution is demonstrated by examples given

Table 4.1: Tachykinins and their receptors

Receptor	endogenous ligand	sequence of the ligand
NK-1	substance P	Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met- NH_2
NK-2	Neurokinin-A	His-Lys-Thr-Asp-Ser-Phe-Val-Gly-Leu-Met- NH_2
NK-3	Neurokinin-B	Asp-Met-His-Asp-Phe-Phe-Val-Gly-Leu-Met- NH_2

in Table 4.3. If the phenyl ring of the tryptophane-benzyl-esters is 3,5-bis-methyl-substituted, affinity increases at least by 6-fold, replacing the methyl groups by trifluoromethyle even results in a 40-fold affinity increase. In another series, compound **14** has a 40-fold increased affinity compared to the unsubstituted phenyl ether **13**. Here, no remarkable change in affinity was observed when substituting the methyl by trifluoromethyl groups (compound **15**). Furthermore, the compounds in Table 4.3 show that affinity can be increased by substituting the exocyclic nitrogen atom. Introduction of a N-acetyl group to compound **10** enhances its affinity by more than 20-fold (**11**), the same increase is observed when attaching the carboxamidomethyl to compound **15** (cf **16**).

The first active NK1 antagonist was CP-96345, which was found in 1991 in a high-throughput screening at Pfizer [Snider et al., 1991]. It is probably the best studied NK1 antagonist. The binding of several derivatives of CP-96345 was measured to obtain insight into the features responsible for binding (for example, see Table 4.4, 90) [Swain et al., 1995; Lowe et al., 1992; Seward et al., 1993; Fong et al., 1993, 1994b; Lowe et al., 1994]. Mutational studies and affinity measurements of derivatives of CP-96345 revealed the key roles of Gln165 [Fong et al., 1994a], His265 [Fong et al., 1994b], and His197 [Fong et al., 1993] in binding of the quinuclidine antagonists. Several groups have constructed a putative interaction model based on the skeleton of CP-96345 and the amino acids essential for binding have been highlighted [Elliott et al., 1998; Takeuchi et al., 1998; Swain et al., 1995; Jacoby et al., 1997; Vedani et al., 2000; Boks et al., 1997]. An interaction model, constructed in our group, is depicted in Fig. 4.5. It is based on the mutational data available in literature and is in agreement with the other

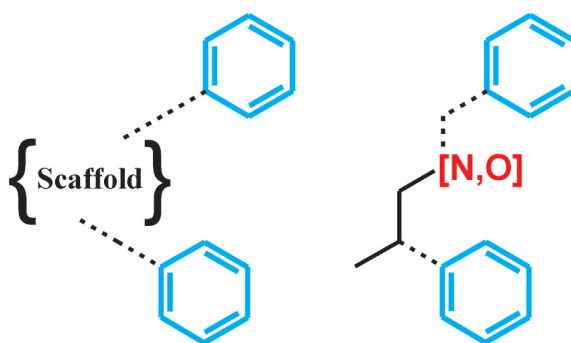


Figure 4.4: Generalised pharmacophore for non-peptidic NK1 antagonists. Two aromatic rings are connected via various scaffolds (left). A more detailed pharmacophore contains at least one hydrogen-bond acceptor within the scaffold (right).

published models.

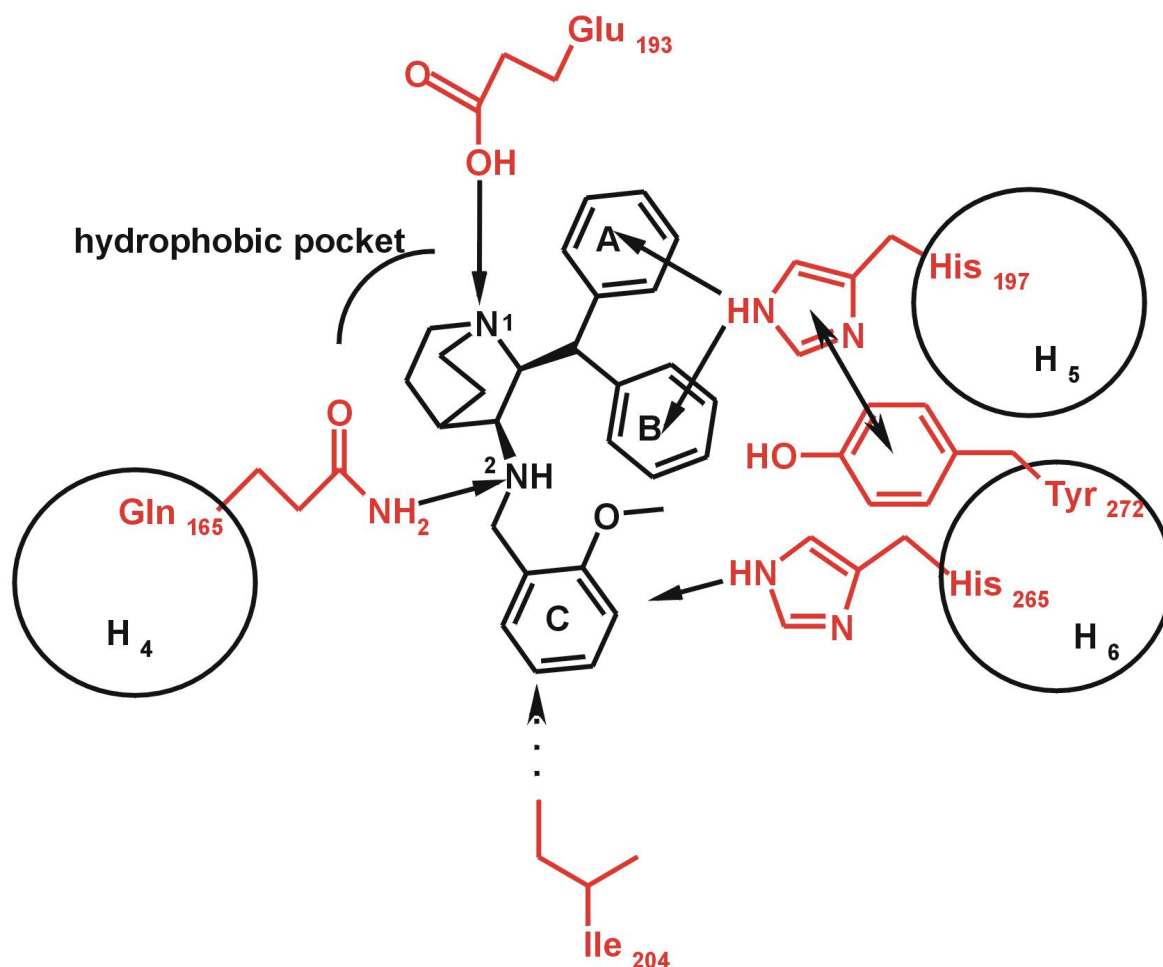


Figure 4.5: Schematic representation of the postulated NK1-receptor-ligand interactions for CP-96345. The arrows indicate proposed key interactions between the receptor and the ligand.

In this interaction model, Gln₁₆₅ establishes a hydrogen bond with the exocyclic secondary amine. As high affinity is also obtained when this secondary amine is replaced by an oxygen, Gln₁₆₅-NH probably acts as a donor and the amine-nitrogen as acceptor. The benzhydryl group of CP-96345 performs an amino-aromatic interaction with His₁₉₇ which is kept in place by an aromatic-aromatic interaction with Tyr₂₇₂ [Fong et al., 1993]. The interactions of the aromatic moiety C with Ile₂₀₄ [Holst et al., 1998; Greenfeder et al., 1998] and His₂₆₅ [Fong et al., 1994b] are obviously not specific. Instead, these residues seem to be part of a hydrophobic pocket. The aromatic moiety A occurs in several NK1 antagonists, but is obviously not necessarily required for high affinity. Supposedly, it serves as conformational anchor, but does not experi-

ence any specific interactions with the receptor. Furthermore, an ionic interaction (or a charged-assisted hydrogen-bond) between the positively charged quinuclidine nitrogen and a corresponding counterpart in the receptor is possible [Boks *et al.*, 1997]. A putative candidate is Glu193, however, mutational data do not definitely support this assumption [Elling *et al.*, 1995; Greenfeder *et al.*, 1998]. Another explanation could be that this part of the ligand interacts with residues from the extracellular loop or it is exposed to the solvent. The latter hypothesis is supported by the observation that a range of diverse polar substituents at the nitrogen is well tolerated [Giardina *et al.*, 1997].

A detailed study of the bioactive conformation of CP-96345 (and other antagonists) was performed by Boks *et al.* [Boks *et al.*, 1997]. They analysed small molecule crystal structures of NK1 antagonists with respect to the intermolecular interactions of their pharmacophoric groups with neighbouring molecules in the crystal packing. The most striking feature is the relative orientation of the two aromatic rings to each other (B and C in Fig. 4.5). Three distinct conformations are observed in crystal structures for CP-96345 and two closely related derivatives, which exhibit distinct orientations of the aromatic ring C. In one case, a conformation with parallel orientation is observed, in the other two cases orientations with perpendicularly oriented rings are found (see Fig. 4.6).

In another study, a conformational search was undertaken by Swain *et al.* to identify the most favourable conformation of CP-96345 [Swain *et al.*, 1995], however, no clear preference for any of the two possible orientations has been found. Modelling studies performed by Sisto *et al.* on a series of peptides and nonpeptides indicate that the aromatic moieties exhibit a parallel stacking with respect to each other [Sisto *et al.*, 1995]. This assumption was further evidenced by ultraviolet absorption and fluorescence measurements. These findings suggest that the conformation adopted in the crystals of LEWCUL (CSD refcode, see Fig. 4.6) matches best the given requirements at the binding site of the NK1 receptor.

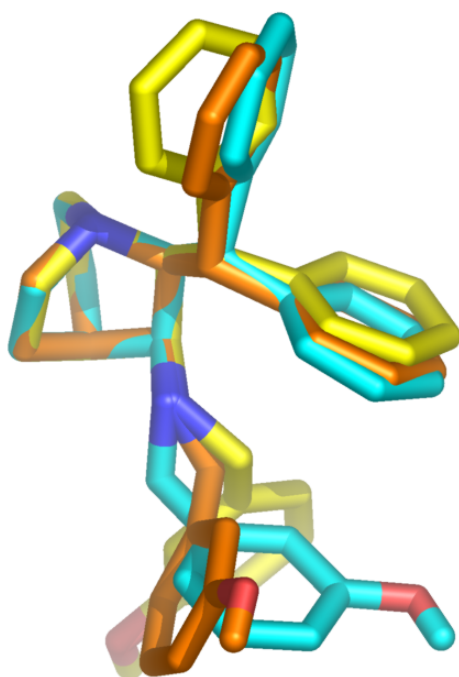
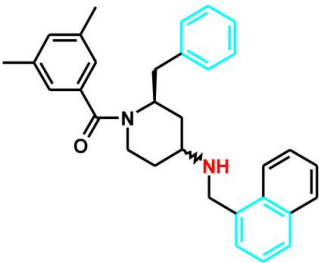
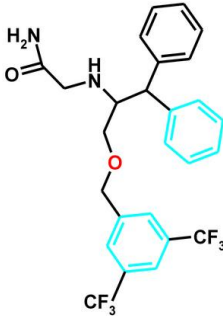
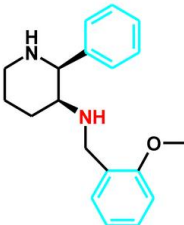
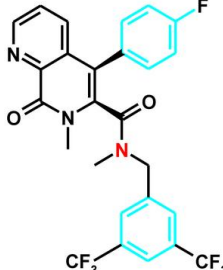
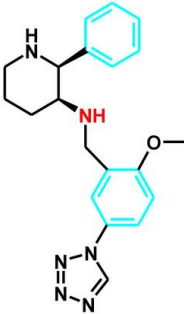
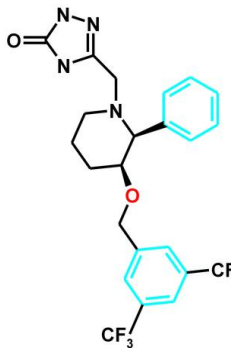
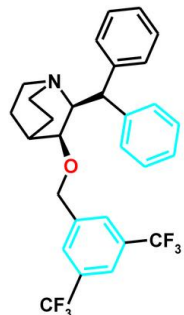
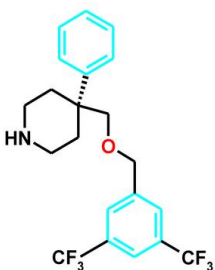


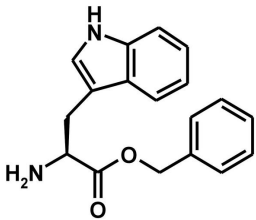
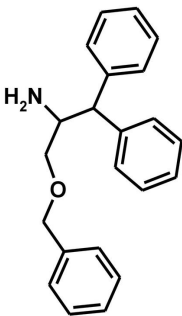
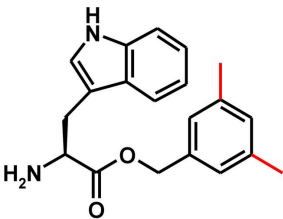
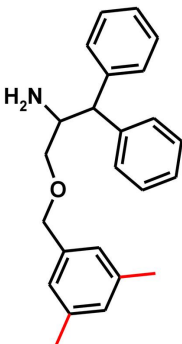
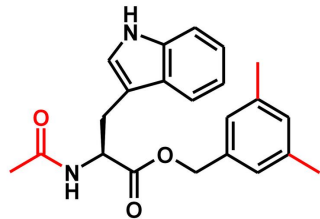
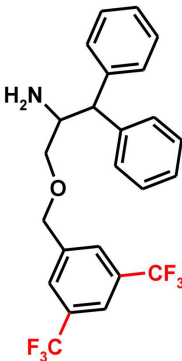
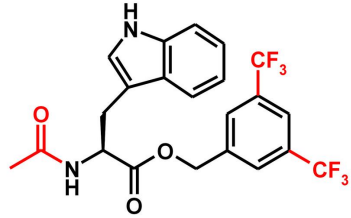
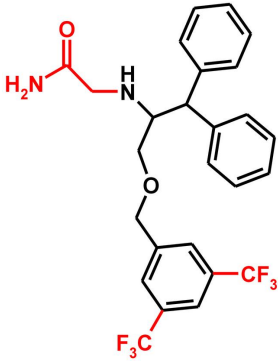
Figure 4.6: Superimposed small molecule crystal structure conformations of CP-96345. The CP-96345 free base (coloured yellow, CSD Refcode YAFJOE, [Lowe et al., 1992]), the CP-96345 dimesylate salt (orange, YAFJUK [Lowe et al., 1992]), and the N-methyl analogue of CP-96345 (cyan, LEWCUL, [Natsugari et al., 1995]) are shown.

Table 4.2: Examples of NK1 antagonists from diverse structural classes.

No.	compound	IC_{50} [nM]	No.	compound	IC_{50} [nM]
1		12	5		0.53
2		0.6	6		0.21
3		0.06	7		0.05
4		1.3	8		1.0

The compounds are extracted from the following references: 1: [Ofner et al., 1996], 2: [Williams et al., 1994], 3: [Desai et al., 1992, 1993], 4: [Natsugari et al., 1995], 5: [Ward et al., 1995], 6: [Ladduwahetty et al., 1996], 7: [Swain et al., 1993], 8: [Stevenson et al., 1995].

Table 4.3: Effect of introducing appropriate substituents to increase affinity to the NK1 receptor.

No.	compound	IC_{50} [nM]	No.	compound	IC_{50} [nM]
9		> 10000	13		400
10		1533	14		9.3
11		67	15		10.7
12		1,6	16		0.53

The compounds are extracted from the following references: 9, 10, 11, 12: [MacLeod et al., 1993], 13, 14, 15, 16: [Williams et al., 1994].

4.2 Modelling the neurokinin-1 receptor

In chapter 3, we have presented *MOBILE*, an approach for homology modelling of protein binding-sites including information about bioactive ligands. In the following, we describe in detail the application of this approach to the NK1 receptor, which belongs to the family of the G-protein coupled receptors.

The quality of the NK1 model and, thus, a critical evaluation of the *MOBILE* approach, is accomplished by probing the ability to find novel antagonists with this homology model. This task is realised by performing a virtual screening based on the modelled receptor structure and a subsequent biochemical testing of a limited number of selected hits.

Considering the fact that G-protein coupled receptors represent one of the most relevant classes of pharmaceutic drug targets, the present study provides a particular challenge for homology modelling with respect to structure-based drug design. This holds in particular for the NK1 receptor, as its overall sequence identity to bovine rhodopsin is only 21 % and in the region of the NK1 antagonist binding pocket no homology is given. A successful application of the *MOBILE* procedure to the NK1 receptor would therefore open a new perspective for the discovery of novel antagonists for any member of the GPCR family.

4.2.1 Generation of protein-ligand complexes using the *MOBILE* approach

Sequence alignment

The sequence alignment (see Fig. 4.7), generated automatically with PSI-BLAST [Altschul et al., 1997] and IMPALA,[Schaffer et al., 1999] was taken from a previously produced homology model of the NK1 receptor deposited in ModBase [Pieper et al., 2002] (database accession number P25103). The alignment was slightly modified in the region of the β 4-hairpin following the recommendations of Oliveira *et al.* [Oliveira et al., 2002] (see section 4.1.1).

As apparent from the mutual alignment, the sequences are most divergent in the ex-

tracellular regions. No gaps or insertions are predicted in the TM regions. However, in the antagonist binding-site region (residues marked grey in Fig. 4.7), no matching residues are conserved.

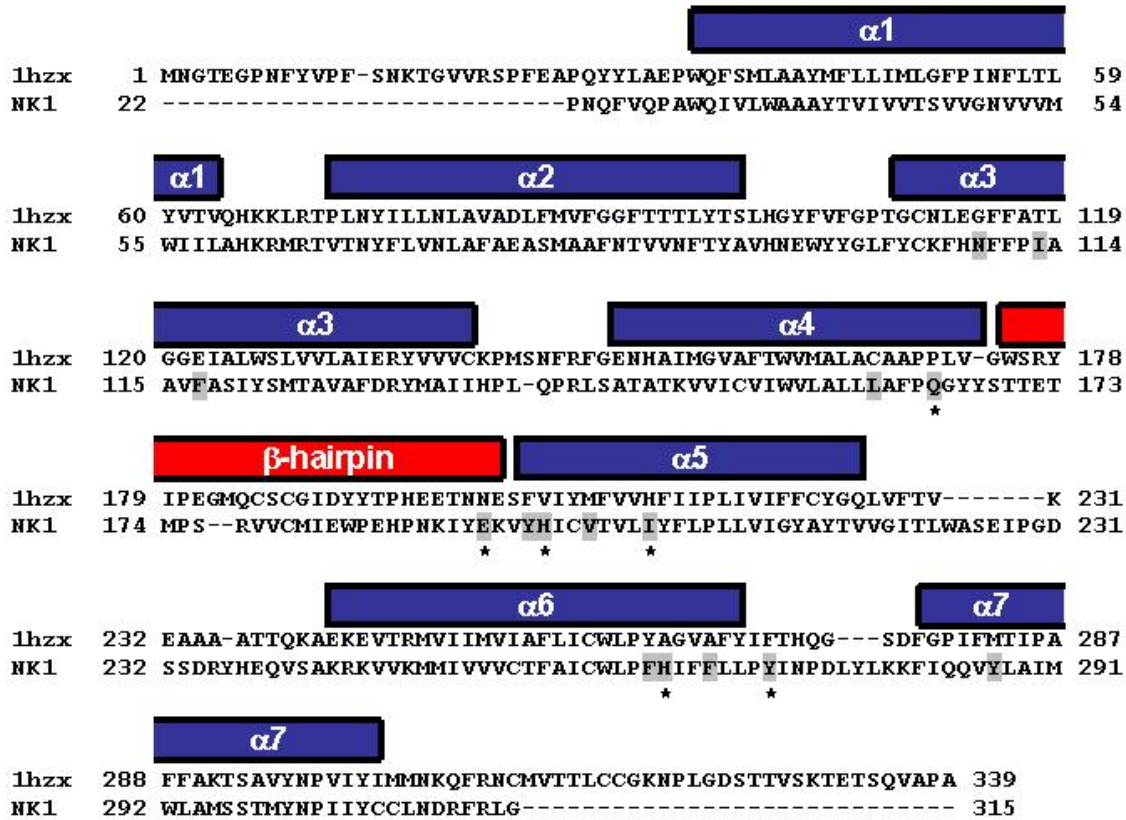


Figure 4.7: Sequence alignment of bovine rhodopsin and the NK1 receptor. All residues comprising the putative binding pocket of CP-96345 are marked grey. The amino acids known from mutational studies to be essential for CP-96345 binding are labelled with an asterisk.

Generation of preliminary NK1 models

A set of 100 initial protein models of the NK1 receptor were generated using MODELLER [Marti-Renom et al., 2000; Sali & Blundell, 1993; Fiser et al., 2000]. According to the algorithms implemented into MODELLER, structures of slightly deviating geometry are subjected to the optimisation step. In total, 100 different structures are finally obtained, which reflect - to some extent - the conformational variability in of those regions which differ in the sequence alignment. For those regions which align

in the sequence space, but with different amino acids (i.e. no insertion or gap), the backbone coordinates will remain close to those of the template coordinates, whereas the conformational space of the side-chain atoms is exhaustively sampled. Thus, an ensemble of 100 binding-site models of the NK1 receptor was produced (see Fig. 4.8).

Visual inspection of the generated models in particular of the residues known by mutational studies to be involved in antagonist binding [Garret et al., 1991; Snider et al., 1991; Cascieri et al., 1992; Fong et al., 1992; Gether et al., 1993a; Fong et al., 1993; Zoffmann et al., 1993; Gether et al., 1994b,a; Huang et al., 1994; Fong et al., 1994a,b; Gether et al., 1993b; Holst et al., 1998; Greenfeder et al., 1999] confirmed the relevance of our sequence alignment: The spatial arrangement of the modelled binding-site residues is in good agreement with the pattern of the proposed interaction models used to describe CP-96345 binding (Fig. 4.5). Fig. 4.8 (left) shows the backbone of one NK1 model together with an ensemble of orientations of those residues which are known to be involved in antagonist binding. The backbone trace of the putative β 4-hairpin is coloured red.

Since the NK1 receptor model was generated by homology, the geometry generated for the β 4-hairpin has been directly transferred from the bovine rhodopsin template structure. Its geometry appears unlikely for the modelled receptor since it would prevent antagonist binding in the present close-up conformation. This detail of the modelled receptor points to the limitations in protein structure prediction by homology, however it is not crucial for the present modelling attempt since mutational studies do not give hints for specific interactions with CP-96345.

Placing the ligand into the preliminary homology models

The next step following our *MOBILE* approach, outlined in chapter 3, would be to dock a ligand flexibly into an averaged ensemble of preliminary NK1 models. However, this procedure is not reasonable in the case of the NK1 receptor. As the conformational space of the residues defining the NK1 antagonist binding-site is exhaustively sampled, the averaged ensemble is too diffuse to allow for precise docking. Therefore, the ligand was docked into each single NK1 model using AutoDock 3.0 [Morris et al., 1996] with DrugScore pair potentials serving as objective function. The E2 loop (containing the β 4-hairpin) as described is not essential for the present modelling study, thus it was removed from the models for the following docking procedure.

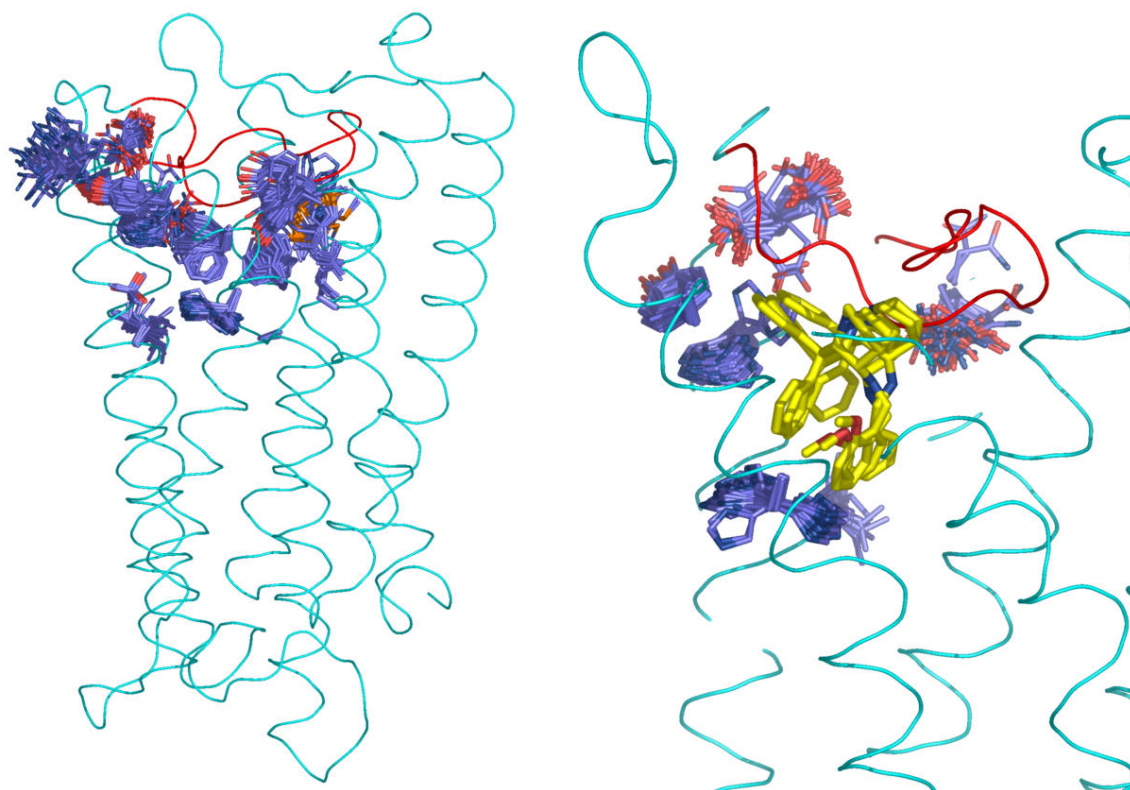


Figure 4.8: Homology models of the NK1 receptor. *left:* The backbone (cyan) and an ensemble of 100 side-chain residue conformers involved in binding CP-96345 is shown; *right:* Four docking solutions of CP-96345. Their orientations agree well to the mutational data and the proposed interaction model shown in Fig. 4.5. The ensemble of the binding-site residues crucial for binding CP-96345 is depicted: Gln165, Glu193, His197, Ile204, His265, and Tyr272.

To reduce the search space, CP-96345 was kept rigid for the docking into the initial homology models, simultaneously assuming that its bound conformation is similar to that observed in the crystal structure (LEWCUL) [Natsugari et al., 1995], i.e. the aromatic rings are supposed to be oriented in parallel to each other (see Fig. 4.6).

Finally, the docking solutions were inspected visually, evaluating the obtained agreement with the interaction model based on the published mutational data [Elliott et al., 1998; Takeuchi et al., 1998; Swain et al., 1995; Jacoby et al., 1997; Vedani et al., 2000; Boks et al., 1997] (4.5). A total of four solutions with alternative orientations of the side-chains of the protein binding-site residues (see Fig. 4.8 (right)) were selected for the subsequent protein modelling step.

Generation of refined NK1 models including ligand information and optimisation of the modelled protein-ligand complexes

In the next step, for each of the four selected docking poses, 100 new homology models were generated. According to the *MOBILE* approach, ligand information was considered as additional restraint in the homology modelling procedure. The 400 thus generated protein-ligand complexes were further refined. First, to each amino acid a DrugScore value was assigned to describe the interaction with the ligand. It was then used to restrain the protein modelling step. Subsequently the best individual solutions from the different models were combinatorially assembled, and finally the composed complex was selected which yielded the best total DrugScore value avoiding any intramolecular clashes between the individual amino acid side-chains from the different models. In order to relax the composed model, the entire binding pocket was minimised with the MAB force-field available in Moloc [Gerber, 1998; Gerber & Müller, 1995], keeping the ligand and the protein residues flexible.

4.2.2 Analysis of the model

Analysis of the global fold of the model

As we applied the principle of homology modelling, the length and geometry of the transmembrane region is produced very similarly to that in bovine rhodopsin. As outlined in section 4.1.1, considering the conservation of key residues in the family of

GPCRs, there are good reasons to assume that the geometry of the transmembrane region will be conserved among these receptors. This assumption is supported by the fact that all GPCRs couple to the same protein (the G-protein). Furthermore, it is likely that all GPCRs follow a similar G-protein activation mechanism [Oliveira et al., 2002]. In consequence, it can be assumed that the inactive state of the NK1 receptor (as for all GPCRs) corresponds to the inactive state of bovine rhodopsin captured in the crystal structure and which was used as template in our modelling process. As we are interested in finding antagonists, which should stabilise the NK1 receptor in its inactive state, the crystal structure of bovine rhodopsin as reference appears well suited.

The confidence in the geometry of the extracellular loops (including the β 4-hairpin) is very low, in contrast to the transmembrane region. This holds in particular for the extra-cellular regions. Considering that all endogenous agonists interact with the extracellular region of their receptor and recalling that these agonists are rather diverse with respect to different GPCRs, it is very likely that GPCRs are very different in this region. From this point of view, it seems very unlikely that the extracellular loop regions adopt similar orientations in all GPCRs. Furthermore, it must be considered that the conformation adopted by the extracellular loops in the crystal structure of bovine rhodopsin is supposedly largely determined by crystal packing forces. However, it has been noticed that for the attempted modelling of the antagonist binding site the assumed loop conformations are of no direct relevance since no specific interactions with NK1 antagonists are suggested in literature.

Analysis and validation of the active site and modelled protein-ligand complex

The complex of the NK1 model, including the six residues known to be crucial for binding CP-96345, is depicted in Fig. 4.9. The proposed interactions (as suggested by the interaction model, Fig. 4.5) are displayed as dashed lines. As mentioned above, no sequence identity between NK1 receptor and bovine rhodopsin is given in the antagonist binding pocket. Thus, the backbone trace of the NK1 model is closely related to the rhodopsin template, but no restrictions for the orientation of the active-site side-chains is provided by the homology model. Since the NK1 model was generated and optimised to produce a binding pocket that exhibits optimal interactions with CP-96345, the arrangement of these side-chains is predominantly determined by the docked

binding mode of this ligand. The finally achieved spatial arrangement of the binding-site residues agrees well with the topographic interaction model depicted in Fig. 4.5: All residues known to be crucial for CP-96345 binding are in direct contact with the ligand in our model. This fact retrospectively confirms the relevance of the assumed sequence alignment.

As indicated by the mutational studies, the most important interaction between NK1 receptor and antagonists is the hydrogen-bond between the Gln165-NH and a hydrogen-bond acceptor of the antagonist. This interaction is well reproduced by our model as indicated in Fig. 4.10. The distance between the Gln165-NH and the nitrogen N2 (dashed lines, see Fig. 4.9) amounts to 3.21 Å. Also, our model reproduces the suggested amino-aromatic interactions [Fong et al., 1993], which are established between the aromatic rings A and B and His197, which is in turn stabilised by an aromatic-aromatic interaction with Tyr272. The role of His267 and Ile204 and their interactions with the ligand are not clear. In agreement with mutational data, instead of specific interactions, hydrophobic interactions between the aromatic moiety C and His264/Ile204 are established. A possible ionic interaction between the nitrogen N1 and a correspond-

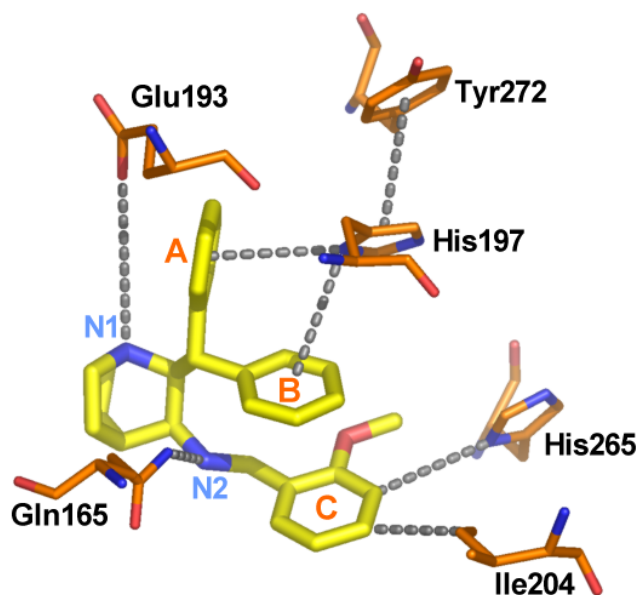


Figure 4.9: Modelled complex of the NK1 receptor with CP-96345 The dashed lines indicate the key interactions of the proposed interaction model depicted in Fig. 4.5.

ing protein partner is not fully evidenced, neither by the mutational data nor by our model (the distance to the Glu193-carboxy oxygen amounts to 5.58 Å). The mutational data show that a slight change in affinity is detected when Glu193 is mutated to Ala or His [Elling et al., 1995; Greenfeder et al., 1998], however, as this decrease is not significant (less than 3-fold and 7-fold), and the ionic interaction cannot be reproduced by our model, N1 is possibly exposed to the solvent and does not directly interact with Glu193 or another amino acid.

As indicated above, it is unlikely that the orientation of the the β 4-hairpin is similar in bovine rhodopsin and the NK1 receptor. Indeed, in our modelled complex, CP-96345 would clash with the β 4-hairpin if it would adopt the same orientation as the crystal structure of bovine rhodopsin (see Fig. 4.8). It was shown by Cavasotto *et al.* that retinal can be docked accurately into the bovine rhodopsin pocket even if the N-terminus, C-terminus, and extracellular (including the β 4-hairpin) and intracellular loops are removed [Cavasotto et al., 2003]. Even though retinal is in contact with the β 4-hairpin in the rhodopsin crystal structure, the majority of the ligand is deeply buried and binding is sufficiently determined by contacts established to the transmembrane part of the receptor. This fact suggests that these features can be used to identify ligands on the basis of a sufficiently accurate homology model of the NK1 receptor. This assumption is further supported by the observation that mutations within the β 4-hairpin of the NK1 receptor did not result in any specific interactions with CP-96345. Consequently, one option for structure-based drug design would be to fully neglect this region in the search for new antagonists. This, however, could provoke docking solutions for ligands that are artificially oriented into unoccupied region. Therefore, to restrict the binding pocket to some degree, however simultaneously to avoid specific interactions with the binding site of CP-96345, we manually placed the β 4-hairpin loop in a somewhat distant region. Subsequently, the corresponding residues were relaxed using the AMBER force-field [Weiner et al., 1986] to avoid unreasonable interaction geometry. In the following virtual screening procedure, the relevance of the docked ligands was evidenced not by solely evaluating the interactions between protein- and ligand-atoms but in particular by considering their similarity with known NK1 antagonists (see section).

To further assess the relevance of the produced model, we examined its ability to accommodate other known NK1 antagonists. Two different sets of antagonists were docked into the active site. The first data set comprised known NK1 antagonists listed in Table 4.2. A second set has been composed by quinuclidine derivatives of CP-96345 (see Tab. 4.4). The ligands were placed using FlexX-Pharm [Hindle et al., 2002],

restraining the docking procedure by defining an essential hydrogen-bond interaction between the Gln165-NH and a hydrogen-bond acceptor in the ligand. As base fragment for the incremental construction algorithm of FlexX, the aromatic ring B was used in the docked orientation of CP-96345 (Fig. 4.9) using the *mapref* mode in FlexX. In all cases, reasonable orientations were obtained. In addition, based on the docking solutions, the trends in binding affinity could be reproduced qualitatively.

4.2.3 Generation of a protein- and ligand-based pharmacophore

Based on the modelled NK1–CP-96345 complex, a structure-based pharmacophore model has been generated considering the mutational data from literature and the common features of known NK1 antagonists, which are assumed as prerequisite for binding.

This pharmacophore model is characterised by the following three features (see Fig. 4.5 and 4.10):

- A hydrogen-bond between the Gln165-NH and a corresponding acceptor of the ligand (N2 in CP-96345). The hydrogen-bond between Gln165-NH and the NK1 antagonists are indicated to be essential by Fong *et al.* [Fong et al., 1994a]. This assumption is best demonstrated by the observation that replacing N2 in CP-96345 by a carbon atom (compound **24** in Table 4.4) results in a dramatic decrease in affinity (0.52 to > 32000 nM).
- The aromatic moiety B which interacts with His197 via amino-aromatic interactions [Fong et al., 1993].
- A second aromatic group (C) which falls next to His265 [Fong et al., 1994b] and Ile204 [Holst et al., 1998; Greenfeder et al., 1998]. Obviously it does not form specific interactions to His265, but favourable interactions are observed for CP-96345 analogs that show substituents at ring C. Although the kind of interaction is not clear, analysis of known NK1 antagonists reveals that an aromatic moiety C is essential for high-affinity binding to the NK1 receptor.

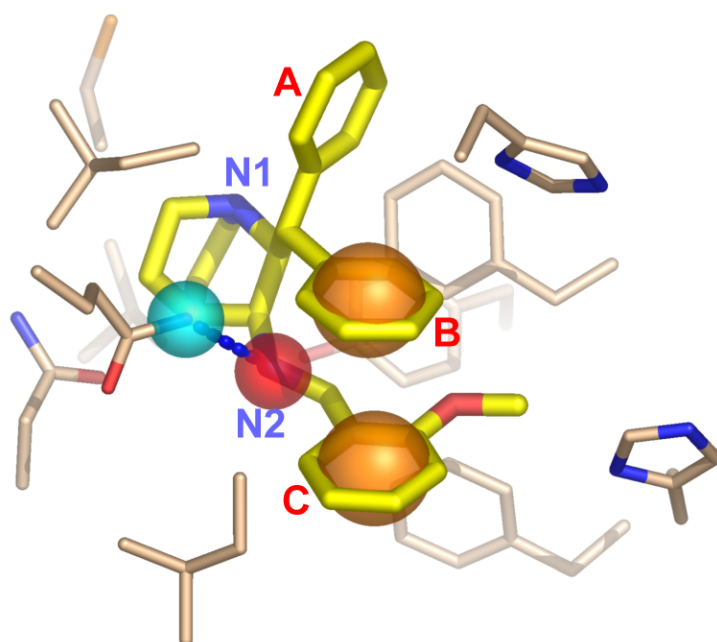


Figure 4.10: Structure-based Pharmacophore hypothesis The H-bond interaction between GLN165-NH (cyan sphere) and a corresponding H-bond acceptor (red) is considered essential for all NK1 antagonists. The aromatic moieties are indicated by yellow spheres.

4.2.4 Virtual Screening

In the present study, about 800000 candidate molecules, assembled from seven different databases, were screened to search for NK1 antagonists. The same compounds were previously screened in our group to search tRNA-guanine transglycosylase (TGT) inhibitors [Brenk et al., 2003].

Similar to the studies of Brenk *et al.* [Brenk et al., 2003] and Grüneberg *et al.* [Grüneberg et al., 2002], the screening has been performed in a stepwise fashion using Selector [Selector, 1996], Unity [UNITY, 2001], and FlexX-Pharm [Hindle et al., 2002] and included several hierarchical filters of increasing complexity with respect to their computational requirements.

The initial step, a rather unspecific and target-independent filter, was already applied by Brenk *et al.* [Brenk et al., 2003]: Only compounds with up to seven rotatable bonds and a molecular weight of less than 450 Da have been considered. The reason for these criteria is to retrieve hits small enough to allow for further optimisation, thus focusing on "leadlike" hits [Oprea et al., 2001; Hann et al., 2001]. Furthermore, highly flexible ligands are avoided as they possibly (1) reveal reduced binding affinity due to entropic considerations and (2) increase the complexity of the attempted 3D search procedures. A further rationale to restrict flexibility arises from the experience that the reliability of accurate docking diminishes with increasing degrees of freedom to be considered in the search procedure. Almost 50 % of the initial compounds were eliminated by this filter.

In a second step, a topological filter was applied according to the pharmacophore requirements given in Fig. 4.5. Only candidate molecules comprising at least (a) two phenyl rings and (b) one hydrogen-bond acceptor were further considered. This reduced the list of prospective compounds to about 16 % of the initial set.

The 3D protein- and ligand-based pharmacophore model (Fig. 4.10) was used in the following step to constrain the mutual spatial arrangement of the aromatic rings and the hydrogen-bond acceptor. In a fourth step, receptor information was explicitly included by restraining the directionality of the hydrogen-bond (to interact with the Gln165-NH) and by considering excluded volume constraints. The number of hits in agreement with this filter contained 11109 compounds. Accordingly, the hierarchical filtering procedure reduced the databases to 1.34 % of their original size (Table 4.5).

Table 4.4: Quinuclidines and analogues that were used for model validation.

No.	compound	IC_{50} [nM]	No.	compound	IC_{50} [nM]
17		0.8	22		> 1000
18		246	23		85
19		20	24		> 32000 ^a
20		1.6	25		12.2 ^b
21		106	26		332 ^b

Affinities of compounds 17 – 23 were derived from inhibition of [¹²⁵I]SP binding (IC_{50} of CP-96345: 0.8 nM); [Swain et al., 1995]. ^a) compound 24 was measured against [³H]SP (IC_{50} of CP-96345: 0.77 nM) [Lowe et al., 1992] and ^b) affinities of compounds 25 and 26 were measured against [¹²⁵I]BHSP (IC_{50} of CP-96345: 0.52 nM) [Lowe et al., 1994].

Table 4.5: Statistical Overview of the Results from Sequential Application of a Series of Hierarchical Filters on the Seven Considered Databases

filter step	ACD		AMBINTER		AEGC		AEPC	
	no. of compds	[%]	no. of compds	[%]	no. of compds	[%]	no. of compds	[%]
	215212	100.00	115815	100.00	182485	100.00	44549	100.00
1. rotatable bonds/ MW	135502	62.96	59877	51.70	91677	50.24	9417	21.14
2. requested no. of hydrophobic, donor and acceptor properties	30878	14.34	19764	17.07	36302	19.89	2740	6.15
3. pharmacophore hypothesis	8645	4.02	5353	4.62	10534	5.77	1018	2.29
4. excluded volumes	3084	1.43	1510	1.30	2998	1.64	334	0.75

continuation Table 4.5

filter step	ChemStar		IBS		LeadQuest		[Σ]	
	no. of compds	[%]	no. of compds	[%]	no. of compds	[%]	no. of compds	[%]
	57927	100.00	158942	100.00	52002	100.00	826952	100.00
1. rotatable bonds/ MW	28712	49.57	76321	48.02	18231	35.04	419747	50.76
2. requested no. of hydrophobic, donor and acceptor properties	11229	19.38	24571	15.46	6483	12.47	131967	15.95
3. pharmacophore hypothesis	3547	6.12	5463	3.44	2144	4.12	36704	4.44
4. excluded volumes	1226	2.12	1362	0.86	595	1.14	11109	1.34

The remaining compounds were docked into the binding-site of our NK1 homology model. To obtain only hits which agree with our 3D pharmacophore model, we used FlexX-Pharm for docking, which allows the incorporation of constraints derived from pharmacophore features. Based on our 3D protein- and ligand-based pharmacophore model (Fig. 4.10), the following features were included: The phenyl ring B was defined as base fragment for the incremental construction algorithm of the docking procedure. This was accomplished using the phenyl ring coordinates of the initially modelled orientation of CP-96345 applying the *mapref* command in FlexX. The hydrogen-bond between Gln165-NH and a composite hydrogen-bond acceptor was constrained as *essential hydrogen-bond interaction*. The orientation of the aromatic ring (C) was not constrained in order to reduce the bias on the system and to assess whether reasonable orientations for this moiety can be generated by FlexX-Pharm using other molecular skeletons and considering the protein environment as constraints.

All docking solutions were scored with DrugScore [Gohlke et al., 2000a]. DrugScore scales within the size of the ligands in contact with the protein. We therefore normalised the score with respect to the number of non-hydrogen atoms in each placed candidate ligand [Pan et al., 2002].

For the 1000 best-ranked ligands, the best docking solutions were minimised with the MAB force-field available in Moloc [Gerber, 1998; Gerber & Müller, 1995] keeping the ligand and the binding-pocket (i.e., all residues within 6 Å around the ligand) flexible. The purpose of this procedure was (1) to optimise the local interactions and (2) to account for protein flexibility induced by ligand binding.

The minimised solutions were very quickly checked visually rejecting those poses which, did not show the aromatic ring C in parallel orientation to ring B as depicted in Fig. 4.10.

The remaining ~ 250 solutions were inspected more carefully considering the following aspects. Ideally, the selection of virtual screening hits could be solely based on the ranking of the scoring function used to evaluate the interaction geometry of the docked ligands. However, many binding features in the protein-ligand interface are yet not fully understood and certain observations cannot be reproduced adequately enough by current scoring functions. Furthermore, it has been shown that the performance of a scoring function possibly depends on binding characteristics present in a particular protein-ligand interface, such as hydrophobicity, hydrophilicity, dominance

of electrostatic/H-bond properties, etc. [Stahl & Rarey, 2001; Schulz-Gasch & Stahl, 2003]. In addition, general observations from quantitative structure-activity relationships prompted us to carefully inspect the best hits from the virtual screening with special regard to the following characteristics:

- An amino-aromatic interaction should be given between His197 and the aromatic moiety B. This type of interaction is not yet parameterised and validated in the current scoring functions.
- Most scoring functions do not consider intramolecular interactions when evaluating protein-ligand interactions, in particular aromatic-aromatic interactions. The described $\pi - \pi$ -stacking, as observed between the aromatic moieties B and C, seems to have a favourable impact on binding, thus its occurrence has been requested.
- As the binding pocket and most interactions between the NK1 receptor and its antagonists are mainly hydrophobic, the hydrogen-bond between Gln165-NH and a corresponding acceptor was carefully analysed. This interaction seems to be of utmost importance: as mentioned, upon replacement of N2 in CP-96345 by a carbon atom (compound 24 in Table 4.4) affinity is dramatically reduced (0.52 to > 32000 nM). This is probably beyond the scope of any scoring function.
- As the model is not reliable next to the region of the $\beta 4$ -hairpin, parts of the ligands placed into this region were evaluated with respect to a given similarity with known NK1 antagonists.

We furthermore focused on ligands with a limited number of rotatable bonds to avoid entropically disfavoured binding due to pronounced conformational immobilisation. Applying these criteria in a thorough visual inspection of the retrieved candidates, the seven compounds listed in Table 4.6 were selected for biochemical testing.

4.2.5 Testing for binding

As assay a radioligand binding assay on whole CHO (Chinese Hamster Ovary) cells (with substance P as radioligand) has been performed. This assay is only sensitive to detect ligands of at least $1 \mu\text{M}$ potency. Any binding beyond this rather stringent limit

cannot be detected. Out of the seven selected compounds, one (**29**) shows 0.25 μM affinity.

Compound **29** agrees well with the 3-dimensional pharmacophore model. Comparing compound **29** and CP-96345 in their docked orientations reveals that the aromatic moieties and the position of the hydrogen-bond acceptor superimpose well. Even the postulated amino-aromatic interaction between His197 and the aromatic ring is matched similarly to CP-96345. The hydrogen-bond with Gln165-NH is established via the peptide carbonyl oxygen (see Fig. 4.11). Besides, the sulfur of the thioether-group could be involved as further hydrogen-bond acceptor. The peptide bond rigidifies the ligand's skeleton and possibly has a favourable impact on the entropic contribution to binding. Considering the docked geometry of **29**, both aromatic moieties (B and C in the pharmacophore model) exhibit a parallel arrangement, however they are shifted with respect to each other in a way that they do not establish a $\pi - \pi$ -stacking interaction. Instead, according to our model, the aromatic ring B stacks upon the π -face of the peptide bond. In addition, **29** exhibits an aromatic moiety A (Fig. 4.10) that is present in several NK1 antagonists. It is obviously not mandatory for high affinity. As mentioned before, it probably serves as conformational anchor, but does not perform any specific interaction with the receptor. Interestingly enough, besides **27**, it is the only one out of the seven tested compounds that possesses this moiety.

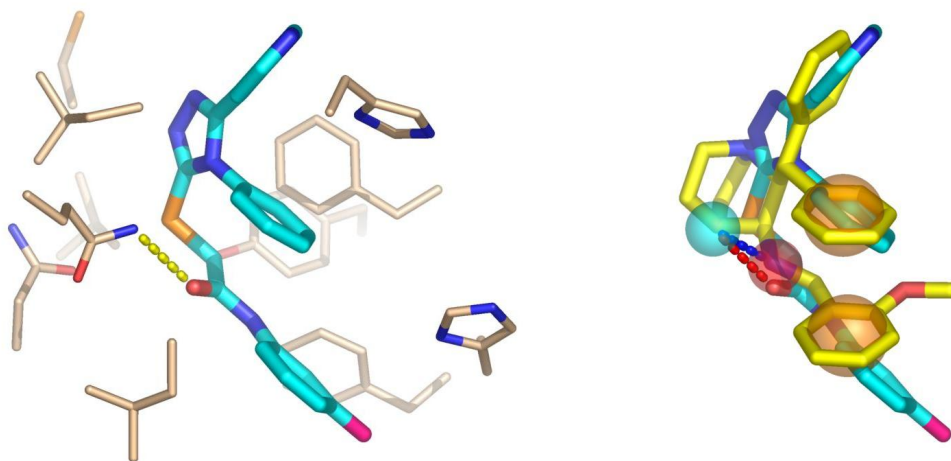


Figure 4.11: Modelled binding mode of ASN-1377642, which was identified as novel submicromolar NK1 antagonist by virtual screening. The orientation with respect to the modelled binding pocket is shown (left). On the right, the modelled superimposition of ASN-1377642 and the known NK1 antagonist CP-96345 is shown.

4.2.6 Discussion

In this contribution, we present a strategy for the computer screening of large compound libraries for the NK1 receptor. This receptor belongs to the family of G-protein coupled receptors, which represents one of the most important pharmaceutical drug target classes. Our approach, which can in principle be applied to any member of the GPCR family to produce a geometry of the receptor in an inactive state, is based on a homology model generated on the basis of the crystal structure of bovine rhodopsin as structural template. The NK1 model was constructed under explicit consideration of ligand information applying our recently developed *MOBILE* (Modelling Binding Sites Including Ligand Information Explicitly) approach described in chapter 3 of this thesis. The model was validated by reproducing experimental information such as mutational data and corresponding affinity data of known ligands. It was successfully used to screen seven databases containing in total about 800000 compounds. Docking yielded one novel compound (out of seven selected, biochemically tested hits) that binds to the receptor in the submicromolar range. Any binding of the other tested hits in an affinity range beyond 1 μ M could be possible, however it is out of the detection limit of the applied assay conditions.

Similar computer-aided screening approaches have previously been performed in our group to discover novel inhibitors for the tRNA-guanine transglycosylase (TGT) [Brenk et al., 2003] and carbonic anhydrase II [Grüneberg et al., 2002]. Both studies were based on available high-resolution crystal structures of protein-ligand complexes. The search strategy applied in the present contribution is somewhat different in certain steps of the screening to account for potential structural uncertainties of the homology model. We started with the generation of preliminary protein models and subsequent docking into these crude models and finally generated a refined protein-ligand complex consistent with experimental data. This modelled complex served as a platform to generate a hybrid protein- and ligand-structure-based pharmacophore model and as structural grounds for the following database search. Thus, in our approach, ligand information was not only explicitly included in the protein modelling step but also considered in the screening and scoring procedure. In the initial screening only those compounds out of 800000 database entries were selected, that agreed to simple 2D pharmacophore features established as minimal requirement due to the analysis of known NK1 antagonists. In subsequent steps, ligand information was taken into account by applying 3D pharmacophore features derived from the analysis of the putative binding mode of

the ligand CP-96345, complexed with our NK1 model. In contrast to the approach followed by Brenk *et al.* and Grüneberg *et al.*, these pharmacophore features were also used to constrain the docking procedure within FlexX-Pharm. As a further difference to the above-mentioned studies, the entire protein-ligand complexes were minimised using the MAB force-field to consider possible adaptations of the protein induced by ligand binding and to subsequently account for potential structural deficiencies of the model. Finally, the docked solutions were carefully inspected considering the agreement of their pharmacophoric features with the putative interaction features present in known active NK1 antagonists. Considering the fact that CP-96345 and compound **29** are similar with respect to their pharmacophoric features, but different with respect to their molecular skeletons, we believe that our hit would not be retrieved as one of the top hits using a solely ligand-based or a solely protein-based screening approach.

Primary focus of this study was to demonstrate that ligand-supported homology modelling of the target receptor can be accomplished successfully for GPCRs using our *MOBILE* approach. The discovered lead structure **29** has yet not been further optimised. This could involve, as outlined in section 4.1.3, a remarkable affinity increase by introduction of a 3,5-bis(trifluoromethyl)-substitution at the aromatic moiety C. Swain *et al.* suggested as possible explanation of this observation that a particular arrangement of the aromatic moieties B and C with respect to each other is favoured or that lipophilic contacts with His265 can be established [Swain et al., 1995]. Further optimisation of binding of **29** could be attempted by appropriate substitution of one of the nitrogen atoms at the triazole ring by attaching a N-acetyl or carboxamidomethyl group (cf examples in Table 4.3).

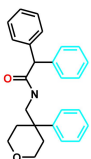
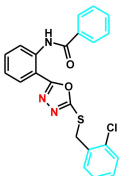
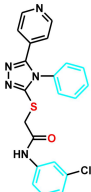
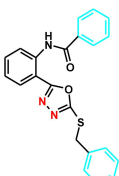
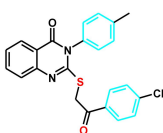
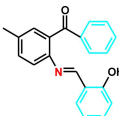
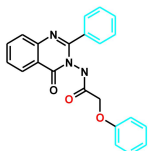
The applied procedure of combining information from bioactive ligands and the knowledge about the 3D structure of bovine rhodopsin, along with mutational data, provides a new perspective to drug discovery of GPCR ligands. Remarkably enough, the global sequence identity between bovine rhodopsin and the NK1 receptor is 21 %. Considering the transmembrane region only, the identity increases to 27 %, however, regarding only the modelled binding site of CP-96345, no sequence identity is given. Usually, if sequence identity falls below 35 %, the accuracy of a homology model is considered not sufficient enough to allow for virtual screening and docking of small ligands [Baker & Sali, 2001]. In light of this non-conserved antagonist binding area among NK1 receptor and rhodopsin it is even more remarkable that our *MOBILE* approach produced a binding-site geometry reliable and relevant enough to discover a submicromolar antagonist via structure-based screening methods.

However, it must be noted that precise affinity prediction is not possible with our model. To a certain extent, this is due to shortcomings of the currently available scoring functions. Further limitations probably arise from the fact that towards the β -hairpin, our model is very crude and approximative and not correctly reflecting the native configuration of the protein.

An important binding determinant is the H-bond formed to Gln165. The mutational and ligand data provide clear evidence that Gln165 acts as a donor [Bieler, 1998]. It could furthermore serve as H-bond acceptor, depending on the side-chain orientation of the terminal amide group and the composite group in the ligand. This has consequences on the definition of the pharmacophore model used to specify the search queries. In our search we requested an acceptor site in putative ligands. An alternative pharmacophore model could define an acceptor **or** donor site at this position and serve as additional basis for the search of NK1 antagonists. Besides, in the 3D pharmacophore model, the arrangement of the aromatic ring systems was assumed to be parallel as observed in the small molecule crystal structure LEWCUL and confirmed by experimental studies based on ultraviolet absorption and fluorescence measurements [Sisto et al., 1995]. Nevertheless, the other two small molecule crystal structures indicate that a perpendicular arrangement of these aromatic moieties to each other also corresponds to a low-energy conformer. Accordingly it cannot be ruled out that the latter geometry is also of relevance for the arrangement at the binding pocket. Neither the available mutational data nor the local contacts to adjacent protein residues next to both aromatic rings favours one of these arrangements.

Furthermore, it has to be mentioned that originally 15 compounds were selected for biochemical testing. Due to inaccessibility or delivery problems of the commercial suppliers, we could only obtain half of the requested compounds (7). The affinity determination was performed using an assay with a detection limit beyond 1 μ M affinity. Since testing at higher concentrations was impossible to perform we cannot decide whether the remaining six hits from our selection antagonise the NK1 receptor with micro- or millimolar affinity. For the present feasibility study such information would be desirable but even the test results on seven compounds would not allow for statistics on the success rates of the present method. To demonstrate that homology modelling using our *MOBILE* approach is capable to produce models of relevance for structure-based virtual screening, the discovered hit which is in full agreement with the search hypothesis is a remarkable and convincing result.

Table 4.6: List of compounds that were tested for inhibition.

No.	compound	label (Database)	DS ^{a)} (rank)	DS2 ^{a)} (rank)	K_B [nM]
27		ASN-2069941 (AEPC)	-54.2 (10)	-17.7 (9)	> 1000
28		ASN-2069935 (AEPC)	-52.7 (19)	-17.3 (16)	> 1000
29		ASN-1377642 (AEPC)	-53.4 (15)	-17.4 (15)	251
30		STOCK2S-25832 (IBS)	-48.1 (42)	-16.7 (29)	> 1000
31		STOCK2S-20468 (IBS)	-49.3 (35)	-16.2 (37)	> 1000
32		STOCK2S-74056 (IBS)	-55.0 (6)	-17.9 (7)	> 1000
33		STOCK1S-23930 (IBS)	-51.2 (24)	-16.7 (30)	> 1000

^{a)} The original DrugScore values are divided by 10000.

4.3 Modelling the neurokinin-2 receptor

Both, NK1 and NK2 receptors, appear to be involved in pulmonary pathophysiology. Accordingly, agents that simultaneously antagonise NK1 and NK2 may have therapeutic applications.

To understand the molecular basis of factors relevant to neurokinin selectivity, we generated ligand-supported homology models of the NK2 receptor including knowledge about mutational studies and the binding of the NK2-selective antagonist SR-48968 [Emonds-Alt et al., 1992]. Different sets of models were constructed, based (1) on the structure of the previously generated homology model of the NK1 receptor (section 4.2) and (2) on the crystal structure of bovine rhodopsin [Palczewski et al., 2000; Teller et al., 2001; Okada et al., 2002].

To obtain insights into the features determining selectivity between NK1 and NK2, mutational and affinity data with respect to the following three ligands (Table 4.7) are analysed: (1) the NK2-selective compound SR-48968 [Emonds-Alt et al., 1992], (2) the dual NK1/NK2 antagonist ZD-6021 [Bernstein et al., 2001], and (3) the NK1-selective CP-96345 [Snider et al., 1991].

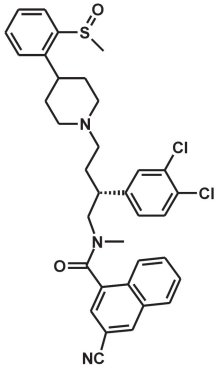
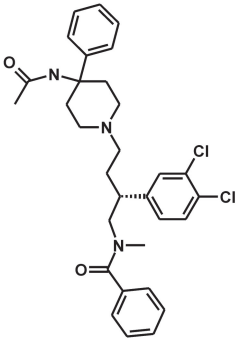
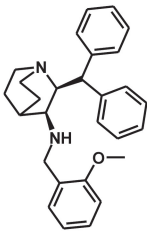
4.3.1 Generation of protein-ligand complexes using the MOBILE approach

Sequence alignment

Different sets of homology models were based on the sequence alignment depicted in Fig. 4.12. This alignment was generated by CLUSTALW [Thompson et al., 1994] keeping the relative positions of the NK1 receptor and bovine rhodopsin as in the alignment depicted in Fig. 4.12 and realigning only the sequence of the NK2 receptor. According to this alignment, the global sequence identity amounts to 56 %. According to the overall strategy of our approach, initial sets of NK2 models were generated without including ligand information into the homology modelling process.

Considering the corresponding 15 residues defining the antagonist binding pocket in the NK1 receptor, the sequence identity between the binding sites of the NK1 and NK2

Table 4.7: Ligands binding to the neurokinin receptors

ligand	structure	K_B [nM] (NK1)	K_B [nM] (NK2)
ZD-6021		1.0	5.5
SR-48968		> 10000	0.7
CP-96345		8.1	> 10000

Affinity measurement of ZD-6021 ([Bernstein et al., 2001]) was performed with a different assay than measurement of SR-48968 and CP-96345 ([Gether et al., 1993b]). The absolute affinity values can thus not be compared directly with each other.

receptor is 75 %. As in the case of the NK1 receptor, the thus defined binding pockets of the NK2 receptor and bovine rhodopsin lack any similarity.

Generation of preliminary NK2 models

Using MODELLER, 100 homology models of the NK2 receptor were constructed. For the generation of these preliminary models, the homology model of the NK1 receptor

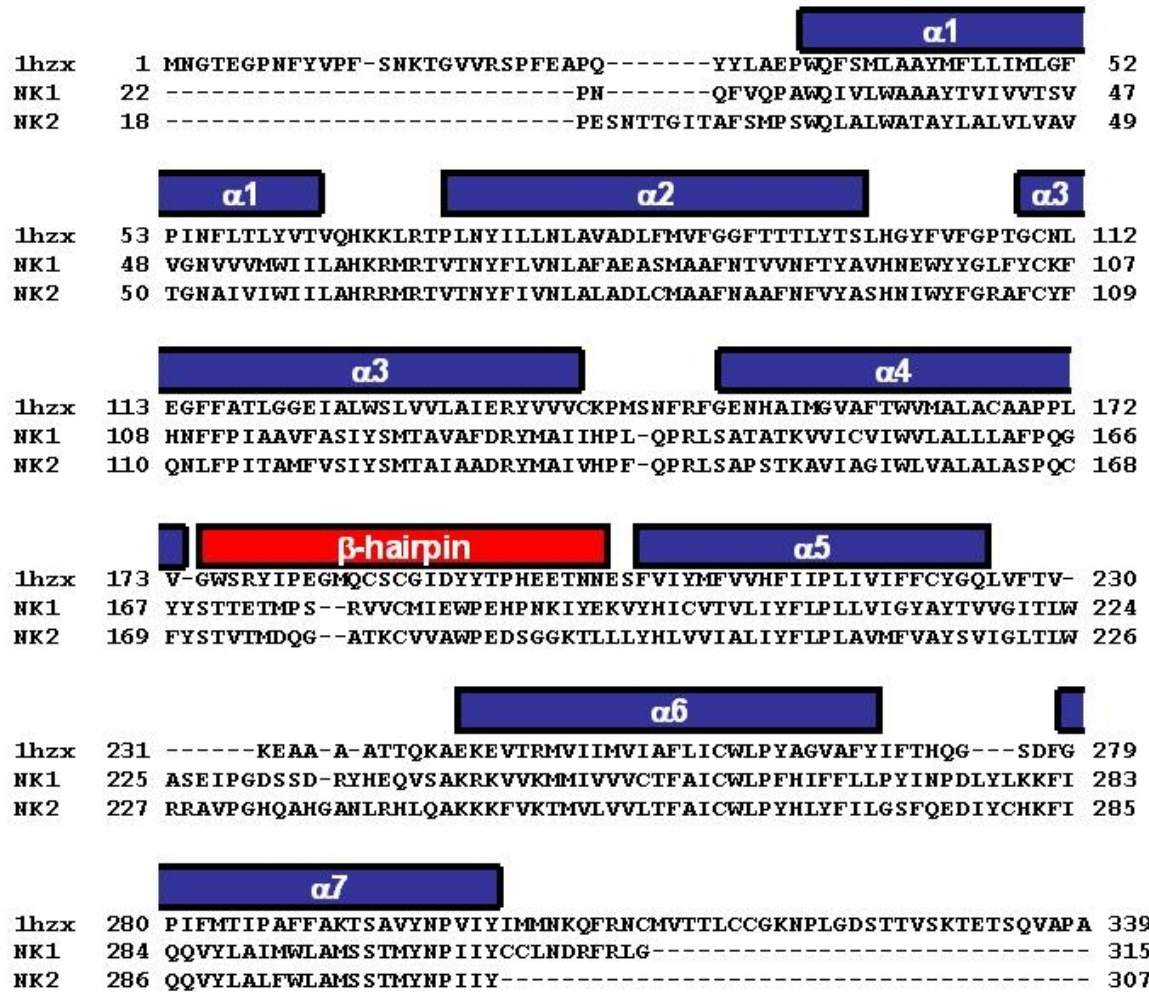


Figure 4.12: Sequence alignment of bovine rhodopsin, the NK1, and NK2 receptor.

(see section 4.2) served as a template. The rationale for such doing is the fact that the sequence identity of the antagonist binding pockets between NK1 and NK2 is rather high (75 %), whereas no identity is given among the binding pockets of the NK2 receptor and rhodopsin. This reduces the search space for the modelling procedure. Not surprisingly, a binding cavity is defined by those amino acids, which are known from mutational studies to be involved in SR-48968 binding (Gln166, His198, Tyr266, His267, Phe270, Tyr289, and Phe293) (see Fig. 4.13).

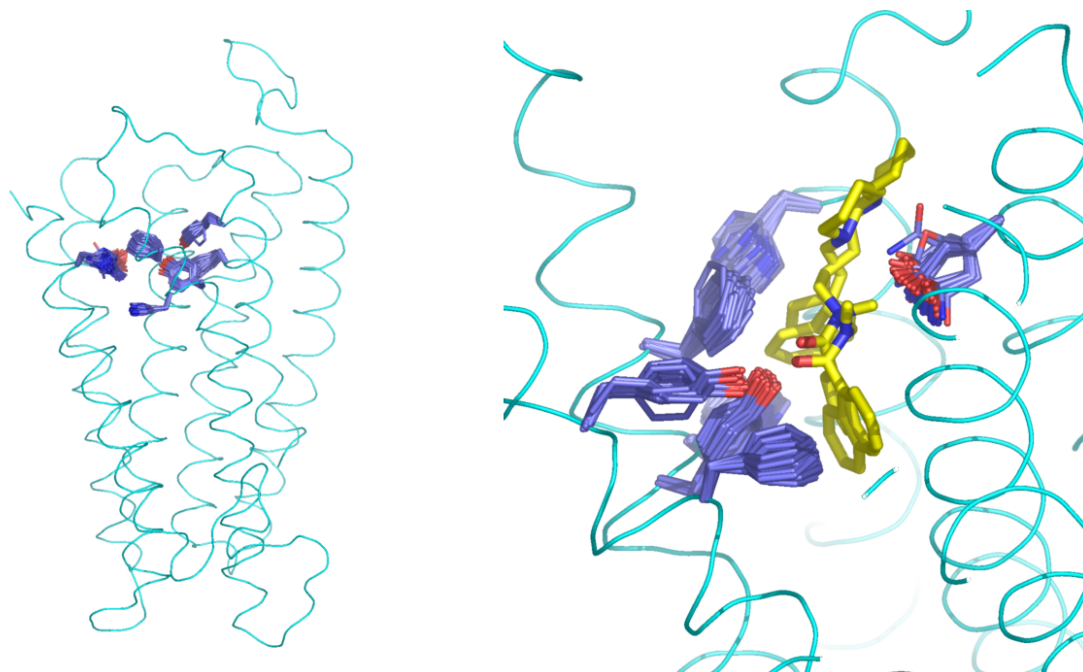


Figure 4.13: Homology model of the NK2 receptor. left: the backbone (cyan) and the ensemble of 100 conformers of 7 residues relevant for binding the NK2-selective antagonist SR-48968 is shown: Gln166, His198, Tyr266, His267, Phe270, Tyr289, and Phe293; right: binding pocket with two docking solutions of the core fragment of SR-48968.

Placing the ligand into the preliminary homology models

A core fragment of SR-48968 (see Table 4.7) was flexibly placed into each individual homology model. For this experiment, SR-48968 was docked with an unsubstituted piperidine moiety, because experimental observations indicate that these substituents interact with regions of the extracellular loops for which, due to the low sequence identity with rhodopsin, no reliable conformation can be obtained by homology modelling.

Accordingly, the corresponding extracellular sequence stretch in the NK2 receptor was not considered for the docking experiment.

The docking solutions were visually analysed. The major criterion to assess the relevance of the modelled complex was the question whether the generated interactions are in agreement with mutational data and preliminary assumptions about a pharmacophore model: Analysing selective NK1, NK2 as well as dual NK1/NK2 antagonists (e.g. [Giardina et al., 1997]), the aromatic ring B in our pharmacophore model for the NK1 receptor (see Fig. 4.10, p. 88) seems to be an essential pharmacophore element with respect to both receptors, probably establishing an amino-aromatic interaction with His197 (NK1), His198 (NK2), respectively. Also an H-bond seems to be essential between SR-48968 and Tyr287. As most NK2 antagonist possess H-bond acceptors at corresponding positions, this feature also appears mandatory for binding. On the other hand, mutational data suggest that an H-bond with Gln166 (Gln165 in NK1) is not essential. Two similar ligand geometries could be identified, which fulfilled the required criteria (see Fig. 4.13).

Generation of refined NK2 models including ligand information and optimisation of the modelled protein-ligand complexes

Assuming that the selected docking solutions resemble the native orientation in the NK2 binding pocket, new ensembles of NK2 models were generated. In order to allow for more conformational sampling of the binding site residues, this modelling step was performed with bovine rhodopsin as structural template. Due to the fact that no sequence identity between NK1 and bovine rhodopsin is given in the active site, the conformational space of the referring residues was exhaustively covered.

The binding sites of the NK1 models were then further refined by optimising the interactions with the two ligand orientations of SR-48968 retrieved from the initial docking experiments. First, each amino acid was assigned a score (the DrugScore value for the interaction with the actual ligand). Then, the best individual solutions from the different models were combinatorially merged, and finally the model was selected which yielded the best DrugScore value and did not show any intramolecular clashes between the individual amino acid side-chains. In order to relax the system, the entire binding pocket was minimised with the MMFF94 force-field [Halgren, 1996], keeping the ligand and the protein residues flexible.

4.3.2 Analysis of the model

Analysis of the global fold of the model

Expectedly, the same observations which were evident for the NK1 model (section 4.2.2) also apply for the global structure of the NK2 model: the global fold is similar to that of the template structure (see Fig. 4.13). As no insertions or deletions occur in the transmembrane regions, the C α trace is completely identical. The geometry of the loops, in particular of the β -hairpin, which interacts with NK2 antagonists, is unlikely to correspond to the actual orientation in NK1 as the sequence identity between both receptors is below 25 % in this sequence stretch.

Analysis of the modelled protein-ligand complexes

The binding at the NK2 receptor is not as extensively supported by mutational studies in literature as binding to the NK1 receptor. Most binding (along with mutational) data for antagonists are available for compound SR-48968 [Emonds-Alt et al., 1992].

According to the mutational data published in refs [Huang et al., 1995; Labrou et al., 2001; Renzetti et al., 1999a,b; Bhogal et al., 1994; Giolitti et al., 2000; Donnelly et al., 1999] the most essential residues for binding SR-48968 are His198, Tyr266, His267, Phe270, Tyr289, Phe293, and Gln166. In particular, Tyr289 (TM VII), along with Tyr266 and Phe270 (TM VI), are proposed to form part of the SR-48968 binding pocket [Huang et al., 1995]. Giolitti *et al.* [Giolitti et al., 2000] propose as prime interactions of SR-48968 the contacts to the aromatic ring of Tyr266 and both, the aromatic ring and the OH group of Tyr289. According to their data, the binding site should comprise Gln166 on TM IV, His198 on TM V, Tyr266 and Phe270 on TM VI for significant but not essential interactions, and Tyr289 as a crucial residue. Although binding is not totally abrogated by the mutations Phe270Ala and Phe270Cys [Labrou et al., 2001; Renzetti et al., 1999b; Huang et al., 1995; Giolitti et al., 2000], the decrease in affinity is significant. Since the mutation Phe270Tyr [Labrou et al., 2001] results only in minor affinity changes, the role of Phe270 could be (1) a hydrophobic interaction with a corresponding part of the ligand, and (2) it stabilises the orientation of the neighbouring aromatic residues Tyr266 and Tyr289 to perform favourable interactions with the ligand. Furthermore, although not directly, Phe293 seems to play an important

role in binding SR-48968. Interestingly, the single mutations Phe270Ala or Phe293Ala do not affect SR-48968 binding but the double mutation Phe270Ala-Phe293Ala fully abolishes SR-48968 binding. Similarly, with respect to SR-48968 binding, simultaneously mutating His267Phe and Phe293Ala produces an effect that was more than the sum of those resulting from two point mutations. Surprisingly, in contrast to the NK1 receptor, Gln166 (Gln165 in the NK1 receptor, which performs an essential H-bond in the NK1 receptor) does not seem to be crucially essential for binding, since the mutation Gln166Val only slightly affects SR-48968 binding. In summary, in the NK2 receptor, the crucial interactions with the analysed ligands seem to be performed via the hydrophobic patch defined by Tyr266, His267, Phe270, Tyr289, and Phe293.

Our model reproduces these observation. As shown in Fig. 4.14a, an H-bond is established between the amide-O of SR-48968 and the Tyr266-OH (dashed lines). According to our model, Tyr289-OH can also forms an H-bond with Tyr266-OH and possibly stabilises its orientation and the interaction with the ligand. Although not exactly observed in our model, it could well be that an H-bond network is established between Tyr266, Tyr289, and the ligand (see Fig. 4.14a). This assumption is supported by the fact that all NK2 antagonists show an H-bond acceptor at the corresponding position. Also the role of Phe270 and Phe293 can be explained by our model. Indeed, no direct, specific interaction is observed for Phe293. Phe270, on the other hand, performs a hydrophobic interaction with SR-48968. It furthermore seems to hold Tyr266 and, eventually, Tyr289 in position to interact with the ligand. As concluded from the mutational data, Phe293 does not interact directly with the ligand, but it borders the hydrophobic binding pocket and, similarly to Phe270, it seems to keep Tyr266 and Tyr289 in position. According to our model, an H-bond with Gln166 is also possible, however, as mentioned above, this interaction does not seem to be as essential for binding as in the NK1 receptor (Fig. 4.14b).

Differences between the NK1 and NK2 receptor models

Fig. 4.15a displays those residues of the NK1 receptor, which are within 6.0 Å distance to CP-96345 (residues coloured yellow) and the corresponding residues of the modelled NK2 receptor (cyan). Fig. 4.15b depicts all those binding-site residues which are conserved among NK1 and NK2 (in total 24 residues). Those residues which are mutated among the NK1 and NK2 receptor are displayed in Fig. 4.15c. In the extracellular region (residues at the top of the ligand), the degree of conservation seems to be lower

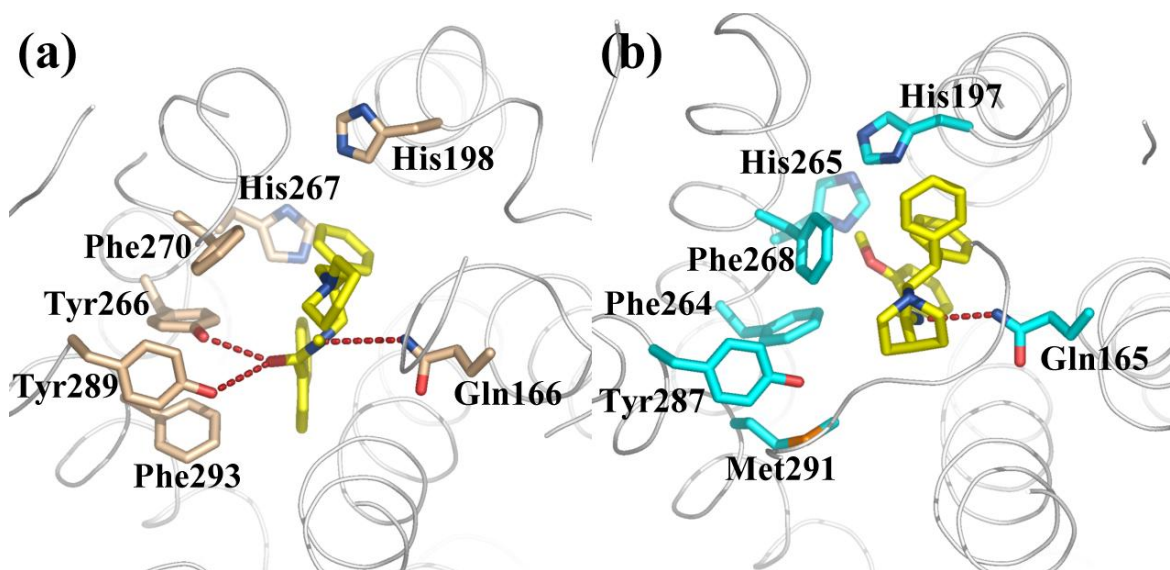


Figure 4.14: Binding pockets of the (a) NK2 and (b) NK1 receptors and their selective ligands SR-48968 (NK2) and CP-96345 (NK1)

than in the transmembrane region. Fig. 4.15d depicts all those residues which do not belong to the extracellular region **and** are not conserved among the NK1 and NK2 receptor. From these (seven) residues, three (Lys194, Thr201, and Leu269 in the NK1 receptor) point away from the ligand, whereas Glu193 could possibly perform specific interactions with antagonists (see also section 4.2). On the other hand, the residues Phe264 and Met291, which are not conserved with respect to the NK2 receptor, are part of the hydrophobic pocket depicted in Fig. 4.14. As discussed above, this part of the binding pocket is essential for binding NK2 antagonists and could be relevant with respect to selectivity between NK1 and NK2 antagonists.

Due to the low sequence identity of the β -hairpin region among the neurokinin receptors and to rhodopsin, the homology models are probably too unprecise to allow for reliable structural comparisons. It is very likely that this sequence stretch adopts distinct conformations in the NK1 and NK2 receptor. This suggests that selectivity can be modulated by addressing the corresponding region of the receptors.

Analysis of ligand data

Whereas CP-96345 binds selectively to the NK1 receptor, SR-48968 is selective for the NK2 receptor (see Table 4.7, p. 100). ZD-6021 shows similar affinities to both

receptors. Interestingly, ZD-6021 and SR-48968 are relatively similar, differing only in the substitution of the piperidine ring and the phenyl ring of the benzamide group. It was indeed shown that by modifying this phenyl ring the degree of NK1 and NK2 activity can be modulated (for example, refs [Shih et al., 2002; Burkholder et al., 1996; Qi et al., 1998; Bernstein et al., 2001; Albert et al., 2002]). For example, as discussed for the NK1 homology model (see section 4.1.3), 3,5-disubstitutions on the benzyl group, especially by CF_3 can markedly enhance the activity for the NK1 receptor. However, these effect could not be rationalised quantitatively by the NK1 homology model.

Another option to modulate selectivity to the NK1 and NK2 receptor is the substitution of the piperidine ring. According to the NK2 model, substituents at this ring interact with the extracellular region of the receptor. As the model is probably not reflecting the native structure in this region, insights into structure-activity relationships should not be based on the model but on the analysis of ligand data, for example cf [Gerspacher et al., 2001; Ting et al., 2000, 2001, 2002; Nishi et al., 1999; Qi et al., 1998; Mah et al., 2002; Bernstein et al., 2001; Albert et al., 2002; Reichard et al., 2002].

4.3.3 Discussion

Computer-aided inhibitor design is concerned with the task of (1) finding and (2) optimising lead structures. In this thesis, it was shown for the NK1 receptor that the derived homology model was sufficient for identifying a lead structure with submicromolar affinity. However, trends in binding affinity for known inhibitors could not be reproduced at a quantitative level. A possible approach for optimising our novel NK1 lead structure, is to consider structure-activity relationships by analysis of known potent NK1 antagonists (see 4.2.6).

Our approach of combining protein- and ligand-based information in a computer-aided screening procedure should also be followed to discover novel dual antagonists for the NK1 and NK2 receptor. Based on the insights of the NK2 homology model (in complex with SR-48968) and on the analysis of ligand data, a 3D pharmacophore model could be derived that is general enough to filter potential NK1 and NK2 antagonists. This pharmacophore model would comprise the phenyl ring B of the NK1 model as given in Figures 4.5 and 4.10 (p. 74 and 88). Furthermore, an H-bond acceptor is required. The spatial tolerances should be defined large enough to enable an H-bond to Gln165 (for binding the NK1 receptor) and (simultaneously) to Tyr266 (for binding the NK2

receptor, see Fig. 4.14). Finally, the pharmacophore should include the aromatic moiety C (Fig. 4.5, p. 74).

Since sufficient data about ligand binding affinities are available for the NK1 and NK2 receptor, lead structure optimisation should be followed by analysing affinity data about known ligands. This holds in particular for the modulation of NK1 and NK2 selectivity. In the case of the NK1 and NK2 receptors, this approach is probably an attractive procedure since a vast number of information is available in the literature. As inferred by analysis of the homology models and ligand data, NK1 and NK2 selectivity can be modulated by variations in the region (1) pointing to the extracellular region or (2) pointing to the hydrophobic subpocket of the receptors.

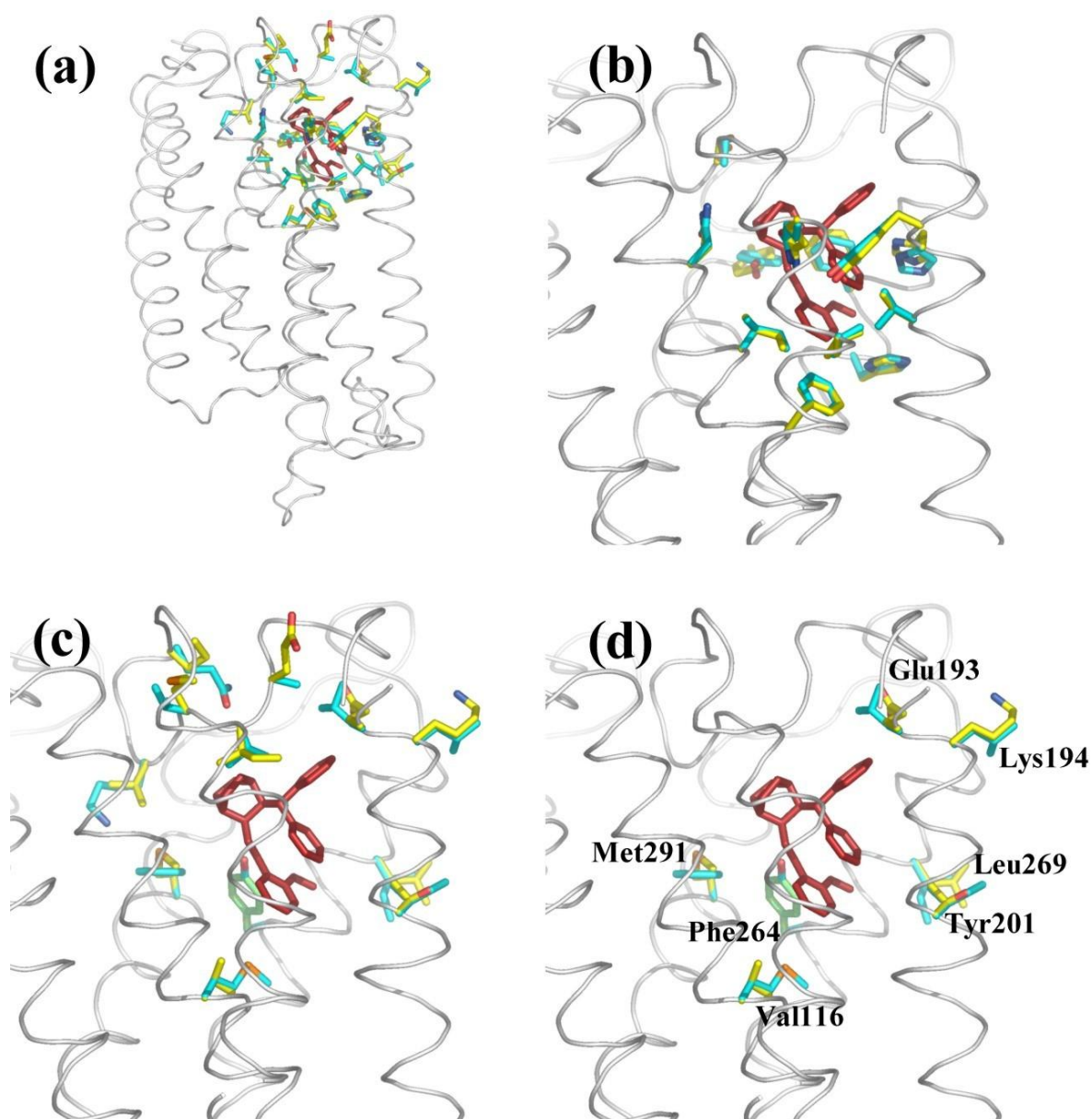


Figure 4.15: Comparison of the NK1 and NK2 receptor models. The 24 antagonist binding-site residues of the NK1 (yellow) and NK2 receptor (cyan) are shown (a). (b) depicts all residues that are conserved among the NK1 and NK2 receptor (12 residues). (c) shows the 12 residues that are mutated among the NK1 and NK2 receptor, (d) depicts only those mutated residues that do not belong the extracellular region (7 residues). The residues of the NK1 receptor are labelled.

5 Summary, Zusammenfassung

5.1 Summary

Recent advances in the various genome sequencing projects have opened the floodgates to thousands of protein sequences, possibly coding for new targets in drug discovery. Thus increasingly, for a considerable part of putative drug targets, the three-dimensional structure will not be readily available. In such cases, the most reliable computer-based technique to generate a three-dimensional protein structure is homology modelling. As homology modelling only considers information available from the related protein structures, it has to remain on a rather approximate level if in the target protein several amino acids of the active site are replaced with respect to those in the template protein(s).

In this thesis, a novel strategy (*MOBILE* (Modelling Binding Sites Including Ligand Information Explicitly)) was developed that models protein binding-sites simultaneously considering information about the binding mode of bioactive ligands during the homology modelling process (chapter 3). As a result, protein binding-site models of higher accuracy and relevance can be generated.

Starting with the (crystal) structure of one or more template proteins, in the first step several preliminary homology models of the target protein are generated using the homology modelling program MODELLER. Ligands are then placed into these preliminary models using different strategies depending on the amount of experimental information about the binding mode of the ligands. (1.) If a ligand is known to bind to the target protein and the crystal structure of the protein-ligand complex with the related template protein is available, it can be assumed that the ligand binding modes are similar in the target and template protein. Accordingly, ligands are then transferred among these structures keeping their orientation as a restraint for the subsequent modelling process. (2.) If no complex crystal structure with the template is available, the ligand(s) can be placed into the template protein structure by docking, and the resulting orientation can then be used to restrain the following protein modelling process. Alternatively, (3.) in cases where knowledge about the binding mode cannot be inferred by the template protein, ligand docking is performed into an ensemble of

homology models. The ligands are placed into a crude binding-site representation via docking into averaged property fields derived from knowledge-based potentials. Once the ligands are placed, a new set of homology models is generated. However, in this step, ligand information is considered as additional restraint in terms of the knowledge-based DrugScore protein-ligand atom pair potentials. Consulting a large ensemble of produced models exhibiting different side-chain rotamers for the binding-site residues, a composite picture is assembled considering the individually best scored rotamers with respect to the ligand. After a local force-field optimisation, the obtained binding-site models can be used for structure-based drug design.

The incorporation of the knowledge-based DrugScore pair-potentials as additional restraints into the MODELLER force-field was adjusted using a test data set of 46 protein-ligand complexes extracted from the Protein Data Bank (PDB). The use of the DrugScore pair potentials proved not only efficient to restrain the homology modelling process, they were also successfully applied to score and optimise the modelled complexes (section 3.2.1). For 70 % of the complexes in the test set, near-native binding-site geometries were produced (root-mean-square deviation (rmsd) ≤ 2.0 Å) with MODELLER considering a bound ligand in its native orientation. Scoring the resulting complexes with DrugScore revealed, in 66 % of the cases a near-native binding mode (rmsd ≤ 2.0 Å) on rank 1.

MOBILE has been applied to two case studies (sections 3.2.2 and 3.2.2). In the first example, factor Xa was modelled based on the crystal structure of trypsin. After docking the ligand RPR128515 flexibly into preliminary factor Xa models, new homology models were generated with the *MOBILE* approach including the ligand in its orientation crystallised with trypsin. To validate the correctness of the model, 10 known factor Xa inhibitors were docked into the factor Xa model and, for reasons of comparison, into a crystallographically determined factor Xa structure. Since similar docking poses and energy scores were obtained for all considered ligands, the generated model appears to be well suited for structure-based drug design purposes.

In another example, different sets of models of aldose reductase were generated based on the crystal structure of aldehyde reductase and (1) the ligand tolrestat or (2) IDD594, which are known to induce conformational changes in the protein affecting even the backbone. This study demonstrated that relevant protein-ligand geometries can even be generated by *MOBILE* in case of a rather complex system, where the bound ligands reinforce different loop conformations upon binding. Furthermore, it was shown that

the mutual orientations between the protein and a particular ligand can be adjusted in a stepwise fashion. Even though the initial starting protein-ligand geometries deviated considerably from the orientations found in the corresponding crystal structures, near-native geometries could be generated for different complexes after performing a second cycle with the *MOBILE* approach.

As a real-life test scenario we applied *MOBILE* to the neurokinin-1 (NK1) receptor, a member of the superfamily of G-Protein coupled receptors (GPCRs) which mediate responses to, e.g. visual, olfactory, hormonal, or neurotransmitter signals (4). This class represents one of the most relevant target families for small-molecule drug design. Due to the fact that GPCRs are membrane-bound proteins, their expression, purification, crystallisation and structure determination remain a major enterprise. So far, only the structure of bovine rhodopsin could be determined to sufficiently high resolution. The NK1 receptor belongs to the series of tachykinin-binding receptors (NK1, NK2, and NK3). They selectively bind the peptide neurotransmitters substance P, neurokinin A, and neurokinin B, respectively. Substance P plays a role in the transmission of pain and is involved in inflammation and immune response. The probably best studied NK1 antagonist is CP-96345. Considerations about its bioactive conformation were achieved through several theoretical and experimental studies. Through mutational studies and comparative affinity determinations based on CP-96345 binding, the essential amino acids involved in ligand recognition could be identified and translated into a crude topographical interaction model (Fig. 4.5, p. 74). The sequence identity between bovine rhodopsin and the NK1 receptor amounts to 21 %. Considering only the transmembrane regions, this figure increases to 27 %. However, regarding only the antagonist binding site of the NK1 receptor, no sequence identity can be detected.

In the first step, 100 preliminary homology models based on the crystal structure of rhodopsin were generated. Next, CP-96345 was docked rigidly into the modelled binding pockets with AutoDock, using the conformation observed in its small-molecule crystal structure (CSD Refcode: LEWCUL). From the set of docked ligand poses, we selected that solution which satisfied the key interactions of our topographical model best. Subsequently, new homology models were generated explicitly considering bound CP-96345. The best-scored side-chain orientations of the individual models were merged in a combinatorial fashion. Finally, the model was selected that yielded the best total DrugScore value avoiding any unfavourable intramolecular contacts among individual amino acid side-chains. The binding-site of the finally composed complex (Fig. 4.9, p. 85) reproduced convincingly well the proposed interactions observed in the

topographical model. The relevance of the model was validated by probing its ability to accommodate other known NK1 antagonists from structurally diverse compound classes.

Mutational data and features shared in common by known NK1 antagonists were then used to establish a pharmacophore hypothesis to retrieve candidate ligands from seven databases containing in total about 800000 compounds. A hierarchical search strategy consisting of 2D and 3D filters of increasing complexity was applied using the programs Selector and UNITY. The 11109 compounds remaining after these filtering steps were flexibly docked into the modelled NK1 binding pocket using the program FlexX-Pharm and scored with DrugScore. After minimising the produced complexes of the best-ranked solutions using MOLOC, the best hits were inspected visually analysing their adopted binding modes, the agreement with the proposed topographical interaction model, and the mutual surface complementarity between protein and accommodated ligand. Finally, seven compounds were selected for biochemical testing. Out of these, one (**29**) showed at 251 nM binding affinity. Considering that CP-96345 and **29** are similar with respect to their pharmacophore features (Fig. 4.11, p.94), but distinct with respect to their molecular skeletons, suggests that this hit would not be matched as one of the top hits using a solely ligand-based or solely protein-based screening approach.

A further step towards homology modelling GPCRs for structure-based drug design was undertaken by generating a ligand-supported homology model of the neurokinin-2 (NK2) receptor (section 4.3). Both, NK1 and NK2 receptors appear to be involved in pulmonary pathophysiology. Accordingly, agents that simultaneously antagonise NK1 and NK2 may have therapeutic applications. Following the *MOBILE* approach, ligand-supported NK2 models were generated that were restrained by the NK2-selective antagonist SR-48968. The modelled antagonist binding pockets of the NK1 and NK2 receptors, showing an overall sequence identity of 75 %, were compared and the features responsible for binding the NK1-selective antagonist CP-96345, the NK2-selective antagonist SR-48968, and the dual antagonist ZD-6021 were rationalised on the basis of the homology models. These insights, together with the generated models, might be a useful platform for the design of further ligands antagonising NK1 and NK2.

In summary we can conclude that our ligand-supported homology modelling produces binding-site models that can be successfully applied as platform for structure-based drug design. The fact that the post-genomic era will provide us with about a factor

of ten more new drug targets of unknown structure, emphasises the relevance and demand for such method. This is in particular the case for the family of the GPCRs where structure determination will probably remain a major enterprise in the next future. Consequently, the successful application of the *MOBILE* procedure to the NK1 receptor opens a perspective field for the discovery of novel antagonists for a large subset of members of the GPCR family.

5.2 Zusammenfassung

Durch die Fortschritte in den Genomsequenzierungsprojekten werden in nächster Zeit tausende von Proteinsequenzen entschlüsselt werden, die für neue Zielproteine ("Targets") für die Arzneimittelforschung kodieren. Für einen beträchtlichen Anteil dieser neuen Targets wird die dreidimensionale Struktur nicht unmittelbar zur Verfügung stehen. Die Homologiemodellierung ist die zuverlässigste computergestützte Methode, um die dreidimensionale Struktur eines Proteins vorherzusagen. Voraussetzung für die Anwendung dieser Methode ist das Vorhandensein einer experimentell aufgeklärten Templat-Proteinstruktur mit ausreichend hoher Sequenzidentität zum Zielprotein. Sie verbleibt bis zu einem gewissen Grad ungenau, insbesondere dann, wenn sich Templat- und Zielprotein in ihrer Sequenz im aktiven Zentrum unterscheiden.

In dieser Arbeit wurde eine neue Methode (*MOBILE* (Modelling Binding Sites Including Ligand Information Explicitly)) entwickelt, die Modelle von Proteinbindetaschen erstellt (Kapitel 3). Durch die explizite Mitberücksichtigung von Liganden während der Homologiemodellierung gelingt es, Bindetaschenmodelle höherer Genauigkeit und Relevanz zu erstellen.

Ausgehend von der Kristallstruktur eines oder mehrerer Templatproteine werden zunächst vorläufige Homologiemodelle des Zielproteins erstellt. Im nächsten Schritt werden bioaktive Liganden in diese vorläufigen Modelle eingepaßt. Dabei werden, je nach vorhandenem Informationsgehalt über die Ligandenbindung, unterschiedliche Strategien verfolgt. (1.) Gibt es einen Liganden, der an Ziel- und Templatprotein bindet und ist darüber hinaus die Kristallstruktur des Protein-Ligand Komplexes mit dem Templatprotein vorhanden, wird der Ligand zwischen diesen Strukturen transferiert und in dieser Orientierung als Randbedingung für die folgende Homologiemodellierung vorgegeben. (2.) Wenn keine Kristallstruktur des Templatproteins mit dem

Liganden gegeben ist, wird der Ligand durch *Docking* in das Templatprotein plaziert und in der resultierenden Orientierung als Randbedingung für die folgende Proteinmodellierung verwendet. (3.) Können Erkenntnisse über die Ligandenbindung nicht über das Templatprotein gewonnen werden, wird der Ligand in angenäherte Binde-taschenrepräsentationen des modellierten Zielproteins eingepaßt. Diese Bindetaschenrepräsentationen werden durch sogenannte gemittelte Eigenschaftsfelder beschrieben, die durch wissensbasierte Potentiale abgeleitet werden. Basierend auf der erhaltenen Ligandenorientierung werden im nächsten Schritt mehrere neue Homologiemodelle erstellt. Dabei wird der Ligand als zusätzliche Randbedingung in Form der wissensbasierten DrugScore Protein-Ligand Atom-Paarpotentiale berücksichtigt. Bei dem resultierenden Ensemble der generierten Homologiemodelle weisen die Aminosäuren im aktiven Zentrum unterschiedliche Seitenkettenkonformationen auf. Daraus wird ein optimiertes Modell zusammengesetzt, indem zunächst die individuellen Rotamere bezüglich ihrer Wechselwirkung zum Liganden bewertet und anschließend so miteinander kombiniert werden, daß die Gesamtbewertung der zusammengesetzten Bindetasche maximal ist. Nach einer lokalen Kraftfeldminimierung können die resultierenden Bindetaschenmodelle für strukturbasiertes Wirkstoffdesign verwendet werden.

Die Einbindung der wissensbasierten DrugScore Paarpotentiale als zusätzliche Randbedingung in das MODELLER Kraftfeld erfolgte durch eine Parameterisierung an einem Datensatz von 46 Protein-Ligand Komplexen aus der Proteindatenbank (PDB). Die Verwendung der DrugScore Potentiale erwies sich nicht nur als geeignet für die Generierung von Homologiemodellen, sie wurden auch erfolgreich zur Bewertung und Optimierung der modellierten Protein-Ligand Komplexe eingesetzt (Abschnitt 3.2.1). Für den Testdatensatz wurden in 70 % der Fälle nativ-ähnliche Bindetaschenmodelle ($\text{rmsd} \leq 2.0 \text{ \AA}$) erstellt, wenn der Ligand in seiner Orientierung aus der Kristallstruktur als zusätzliche Randbedingung in die Homologiemodellierung einbezogen wurde. Bei der Bewertung dieser Modelle mit den DrugScore Paarpotentialen wurden in 66 % der Fälle Orientierungen mit einer mittleren quadratischen Abweichung $\leq 2.0 \text{ \AA}$ auf Rang 1 gefunden.

MOBILE wurde für zwei Teststudien verwendet (Abschnitte 3.2.2 and 3.2.2). Im ersten Fall wurde die Struktur von Faktor Xa basierend auf der Kristallstruktur von Trypsin modelliert. Nachdem der Ligand RPR128515 flexibel in vorläufige Faktor Xa Modelle eingepaßt wurde, wurden neue Homologiemodelle erstellt. Dabei wurde der Ligand explizit in seiner Orientierung aus der Trypsinkristallstruktur miteinbezogen. Um die Relevanz des Modells zu validieren, wurden 10 weitere bekannte Inhibitoren in

das Faktor Xa Model eingepaßt. Aus Vergleichsgründen wurden dieselben Verbindungen zusätzlich in eine Faktor Xa Kristallstruktur plaziert. Für alle Liganden wurden ähnliche Lösungen sowohl hinsichtlich ihrer Orientierung als auch hinsichtlich ihrer Energiewerte erhalten. Folglich scheint das Model für strukturbasiertes Wirkstoffdesign gut geeignet zu sein.

In der zweiten Studie wurden zwei unterschiedliche Serien von Aldose Reduktase Homologiemodellen erstellt. Die Liganden (1) Tolrestat bzw. (2) IDD594, die konformationelle Änderungen im Protein hervorrufen, wurden jeweils in die Homologiemodellierung einbezogen. Dieses Beispiel demonstrierte, daß *MOBILE* auch dann erfolgreich realistische Protein-Ligand Geometrien generiert, wenn zwei Liganden durch ihre Bindung unterschiedliche Konformationen der Loop-Regionen induzieren. Darüber hinaus wurde gezeigt, daß die Orientierungen zwischen dem Protein und einem spezifischen Liganden in einem schrittweisen Verfahren aneinander angepasst werden können. Obwohl die Startgeometrien zwischen Protein und Ligand beträchtlich von der jeweiligen Kristallstruktur abwichen, wurden nach Durchführung eines zweiten *MOBILE* Zyklus nativ-ähnliche Komplexgeometrien erhalten.

Im Rahmen eines realen Modellierungsproblems wurde *MOBILE* für die Erstellung von Homologiemodellen für den Neurokinin-1 (NK1) Rezeptor verwendet. Dieser Rezeptor gehört zur Familie der G-Protein gekoppelten Rezeptoren (GPCRs), die Reaktionen auf verschiedenste Signale (z.B. optische, geruchliche, hormonelle oder neurotransmittervermittelte) übermitteln. Die Klasse der GPCRs ist eine der relevantesten Zielfamilien des aktuellen Wirkstoffdesigns. Da GPCRs membran-gebundene Proteine sind, ist ihre Expression, Aufreinigung, Kristallisation und Strukturaufklärung immer noch nicht gelöst. Bisher konnte nur die Kristallstruktur von Rinderrhodopsin in genügend hoher Auflösung bestimmt werden. Der NK1 Rezeptor gehört zu den Tachykinin bindenden Rezeptoren (NK1, NK2 und NK3), die selektiv an die peptidischen Neurotransmitter Substanz P, Neurokinin A und Neurokinin B binden. Substanz P spielt eine wichtige Rolle in der Schmerzübertragung und ist an Entzündungsprozessen und Immunantworten beteiligt. Der wahrscheinlich am besten untersuchte NK1 Antagonist ist CP-96345. Erkenntnisse über dessen bioaktive Konformation wurden durch eine Reihe von theoretischen und experimentellen Untersuchungen gewonnen. Durch Mutationsstudien und vergleichende Affinitätsbestimmungen, basierend auf der Bindung von CP-96345, konnten die essentiell an der Ligandenbindung beteiligten Aminosäuren identifiziert und in ein topographisches Wechselwirkungsmodell übersetzt werden (Fig. 4.5, S. 74). Die Sequenzidentität zwischen Rinderrhodopsin und dem NK1 Rezep-

tor beträgt 21 %. Betrachtet man nur die Transmembranregion, steigt sie auf 27 %. Werden jedoch nur die Aminosäuren der Antagonistenbindungsstelle berücksichtigt, so beobachtet man überhaupt keine Sequenzidentität.

Basierend auf der Kristallstruktur von Rinderrhodopsin wurden zunächst 100 vorläufige Homologiemodelle erstellt. Im nächsten Schritt wurde CP-96345 mit AutoDock in die modellierten Bindetaschen plziert. Dabei wurde der Ligand rigide in seiner Orientierung aus der Kristallstruktur (CSD Referenzcode: LEWCUL) eingesetzt. Die am besten bewerteten Aminosäuren der individuellen Proteinmodelle wurden kombinatorisch zusammengesetzt und das Modell mit der besten gesamten DrugScorebewertung ausgewählt. Die im topographischen Modell (Fig. 4.9, S. 85) vorgeschlagenen Wechselwirkungen werden durch das resultierende Modell gut reproduziert. Die Relevanz dieses NK1 Modells wurde durch die Vorhersage der Bindungsmodi weiterer bekannter NK1 Antagonisten aus strukturell verschiedenen Klassen bestätigt. Mutationsdaten und gemeinsame strukturelle Eigenschaften bekannter NK1 Antagonisten wurden anschließend in eine Pharmakophor-Hypothese umgesetzt. Basierend auf dieser Pharmakophor-Hypothese wurden in einer Datenbanksuche Liganden aus sieben Datenbanken mit insgesamt 800000 Verbindungen durchmustert. Dieses virtuelle Screening wurde schrittweise mit den Programmen Selector und UNITY durchgeführt. Es bestand aus mehreren hierarchischen 2D und 3D Filterschritten mit zunehmender Komplexität. Die erhaltenen 11109 Verbindungen wurden anschließend mit dem Programm FlexX-Pharm in die modellierte NK1 Bindetasche plziert und mit DrugScore bewertet. Nach einer lokalen Kraftfeldminimierung der resultierenden Komplexe mit MOLOC wurden die besten Treffer einer visuellen Inspektion unterzogen. Dabei wurden die Bindungsmodi insbesondere hinsichtlich ihrer Übereinstimmung mit dem topographischen Wechselwirkungsmodell und hinsichtlich ihrer Oberflächenkomplementarität zwischen Protein und Ligand beurteilt. Sieben Verbindungen wurden für eine biochemische Testung ausgewählt. Von diesen zeigte eine (**29**) eine Bindungsaffinität von 251 nM. CP-96345 und **29** ähneln sich bezüglich ihrer Pharmakophoreigenschaften (Fig. 4.11, p.94), sie weisen aber unterschiedliche molekulare Gerüste auf. Folglich wäre dieser Ligand vermutlich nicht als einer der besten Treffer in einem rein protein- oder rein ligand-basierten Screening gefunden worden.

Ein weiterer Schritt in Richtung "Homologiemodellierung von GPCRs für strukturbasiertes Wirkstoffdesign" wurde durch die Erstellung eines liganden-verfeinerten Homologiemodells für den Neurokinin-2 (NK2) Rezeptors unternommen. Sowohl der NK1 als auch der NK2 Rezeptor sind offenbar in Lungenerkrankungen involviert. De-

mentsprechend könnten Substanzen, die gleichzeitig NK1 und NK2 antagonisieren, therapeutische Anwendung finden. Mit *MOBILE* wurden unter Berücksichtigung des NK2-selektiven Antagonisten SR-48968 ligandengestützte Homologiemodelle des NK2 Rezeptors erstellt. Die modellierten Bindetaschen des NK1 und NK2 Rezeptors, die eine Sequenzidentität von 75 % aufweisen, wurden hinsichtlich ihrer Bindung zu dem NK1-selektiven Antagonisten CP-96345, dem NK2-selektiven Antagonisten SR-48968 und dem dualen Antagonisten ZD-6021 verglichen. Die gewonnenen Erkenntnisse können, zusammen mit den generierten Homologiemodellen, als nützliche Grundlage für die Entwicklung weiterer NK1 und NK2 Antagonisten dienen.

Zusammenfassend kann festgestellt werden, daß durch liganden-unterstützte Homologiemodellierung Proteinbindetaschenmodelle generiert werden, die erfolgreich als Grundlage für strukturbasiertes Wirkstoffdesign verwendet werden können. Die Tatsache, daß die Genomprojekte uns mit zahlreichen neuen Arzneistofftargets versorgen wird, verdeutlicht die Relevanz und Notwendigkeit einer solchen Methode. Dies gilt insbesondere für die Familie der GPCRs, die für die experimentelle Strukturbestimmung voraussichtlich auch in der Zukunft eine bedeutende Hürde darstellen wird. Die erfolgreiche Anwendung der *MOBILE* Methode für den NK1 Rezeptor eröffnet eine neue Perspektive für die Entdeckung von Antagonisten auch für andere Mitglieder der GPCR Familie.

Bibliography

- Adler, M., Davey, D. D., Phillips, G. B., Kim, S. H., Jancarik, J., Rumennik, G., Light, D. R., & Whitlow, M. (2000). Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. *Biochemistry*, 39(41):12534–42.
- Albert, J. S., Aharony, D., Andisik, D., Barthlow, H. G., Bernstein, P. R., Bialecki, R., Dedinas, D. F., Dembofski, B. T., Hill, D., Kirkland, K., Koether, G., Kosmider, B. J., Ohnmacht, C. J., Palmer, W., Potts, W., Rumsey, W. L., Shen, L., Shenvi, A., Sherwood, S., Warwick, P. J., & Russell, K. (2002). Design, Synthesis, and SAR of Tachykinin Antagonists: Modulation of Balance in NK_1/NK_2 Receptor Antagonist Activity. *J Med Chem*, 45:3972–3983.
- Allen, F. H., Bellard, S. A., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R., & Watson, D. G. (1979). The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information. *Acta Crystallographica*, B35:2331–2339.
- Altschul, F. A., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25(17):3389–3402.
- Andrade, M. A., Brown, N. P., Leroy, C., Hoersch, S., de Daruvar, A., Reich, C., Franchini, A., Tamames, J., Valencia, A., Ouzounis, C., & Sander, C. (1999). Automated genome sequence analysis and annotation. *Bioinformatics*, 15(5):391–412.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181:223–230.
- Baker, D. & Sali, A. (2001). Protein Structure Prediction and Structural Genomics. *Science*, 294(5): 93–96.
- Balakin, K. V., Lang, S. A., Skorenko, S. E., Tkachenko, S. E., Ivashchenko, A. A., & Savchuk, N. P. (2003). Structure-Based versus Property-Based Approaches in the Design of G-Protein-Coupled Receptor-Targeted Libraries. *J Chem Inf Comput Sci*, 43(5):1553–1562.
- Balakin, K. V., Tkachenko, S. E., Lang, S. A., Okun, I., Ivashchenko, A. A., & Savchuk, N. P. (2002). Property-Based Design of GPCR-Targeted Library. *J Chem Inf Comput Sci*, 42:1332–1342.
- Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *EMBO J*, 12:1693–1703.
- Baldwin, J. M. (1994). Structure and function of receptors coupled to G proteins. *Curr Opin Cell Biol*, 6:180–190.
- Baldwin, J. M., Schertler, G. F., & Unger, V. M. (1997). An alpha-carbon template for the trans-membrane helices in the rhodopsin family of G protein-coupled receptors. *J Mol Biol*, 272:144–164.

- Ballesteros, J. A., Shi, L., & Javitch, J. A. (2001). Structural Mimicry in G Protein-Coupled Receptors: Implications of the High-Resolution Structure of Rhodopsin for Structure-Function Analysis of Rhodopsin-Like Receptors. *Mol Pharmacol*, 60:1–19.
- Bates, P. A., Kelley, L. A., MacCallum, R. M., & Sternberg, J. E. (2001). Enhancement of Protein Modeling by Human Intervention in Applying the Automatic Programs 3D JIGSAW and 3D PSSM. *Proteins*, Suppl. 5:39–46.
- Bathelt, C., Schmid, R. D., & Pleiss, J. (2002). Regioselectivity of CYP2B6: homology modeling, molecular dynamics simulation, docking. *J Mol Model (Online)*, 8(11):327–35.
- Becker, O. M., Shacham, S., Marantz, Y., & Noiman, S. (2003). Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr Opin Drug Discov Devel*, 6(3):353–361.
- Bergner, A., Gunther, J., Hendlich, M., Klebe, G., & Verdonk, M. (2001). Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers*, 61(2):99–110.
- Bernstein, P. R., Aharony, D., Albert, J. S., Andisik, D., Barthlow, H. G., Bialecki, R., Davenport, T., Dedinas, D. F., Dembofski, B. T., Koether, G., Kosmider, B. J., Kirkland, K., Ohnmacht, C. J., Potts, W., Rumsey, W. L., Shen, L., Shenvi, A., Sherwood, S., Stollman, D., & Russell, K. (2001). Discovery of Novel, Orally Active NK_1/NK_2 Antagonists. *Bioorg Med Chem Lett*, 11:2769–2773.
- Beukers, M. W., Kristiansen, K., Ijzerman, A. P., & Edvardsen, O. (1999). TinyGRAP database: a bioinformatics tool to mine G protein-coupled receptor mutant data. *TiPS*, 20(12):475–477.
- Bhogal, N., Donnelly, D., & Findlay, B. C. (1994). The Ligand Binding of the Neurokinin 2 Receptor. *J Biol Chem*, 269(44):27269–27274.
- Bieler, M. Modellierung des humanen Neurokinin1-Rezeptors unter Berücksichtigung von Mutationsergebnissen und Struktur-Wirkungsbeziehungen ausgewählter Liganden. Master's thesis, Martin-Luther-Universität Halle Wittenberg, 1998.
- Birnbaumer, L. & Birnbaumer, M. (1995). Signal transduction by G proteins: 1994 edition. *J Recep Signal Transduc Res*, 15:213–252.
- Bissantz, C., Bernard, P., Hibert, M., & Rognan, D. (2003). Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins*, 50(1):5–25.
- Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J Med Chem*, 43:4759–4767.
- Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., & Thornton, J. M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*, 326:347–352.
- Bockaert, J. & Pin, J. P. (1999). Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J*, 18(7):1723–1729.

- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 31:365–370.
- Böhm, H.-J. (1992a). The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors. *J Comput Aided Mol Des*, 6:61–78.
- Böhm, H.-J. (1992b). LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des*, 6:593–606.
- Böhm, H. J. (1993). A novel computational tool for automated structure-based drug design. *J Mol Recognit*, 6(3):131–7.
- Böhm, H. J. (1994a). 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 Aided Mol Des*, 8(3):243–56.
- Böhm, H. J. (1994b). On the use of LUDI to search the Fine Chemicals Directory for ligands of proteins of known three-dimensional structure. *J Comput Aided Mol Des*, 8(5):623–32.
- Böhm, H. J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des*, 12(4):309–23.
- Böhm, H.-J. & Klebe, G. (1996). What can we learn from Molecular Recognition in Protein-Ligand Complexes for the Design of New Drugs? *Angew Chem Int Ed Engl*, 35:2588–2614.
- Boks, G. J., Tollenaere, J. P., & Kroon, J. (1997). Possible ligand-receptor interactions for NK1 antagonists as observed in their crystal structures. *Bioorg Med Chem*, 5(3):535–47.
- Bourdon, H., Trumpp-Kallmeyer, S., Schreuder, H., Hoflack, J., Hibert, M., & Wermuth, C. G. (1997). Modelling of the binding site of the human m1 muscarinic receptor: experimental validation and refinement. *J Comput Aided Mol Des*, 11(4):317–32.
- Bourne, H. (1997). How receptors talk to trimeric G proteins. *Curr Opin Cell Biol*, 9:134–142.
- Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016):164–70.
- Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., & Engh, R. A. (1996). X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J Biol Chem*, 271(47):29988–92.
- Braun, W. & Go, N. (1985). Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol*, 186(3):611–26.
- Brenk, R., Naerum, L., Grädler, U., Gerber, H.-D., Garcia, G. A., Reuter, K., Stubbs, M., & Klebe, G. (2003). Virtual screening for submicromolar leads of TGT based on a new unexpected binding mode detected by crystal structure analysis. *J Med Chem*, 46:1133–1143.

- Brenner, S. E. (2000). Target selection for structural genomics. *Nature Struct Biol*, Seiten 967–969.
- Brenner, S. E. & Levitt, M. (2000). Expectations from structural genomics. *Protein Sci*, 9:197–200.
- Broder, S. & Venter, J. C. (2000). Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. *Annu Rev Pharmacol Toxicol*, 40:97–132.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comp Chem*, 4(2):187–217.
- Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., & Hill, R. C. (1969). A possible three-dimensional structure of bovine α -lactalbumin based on that of hen's egg-white lysozyme. *J Mol Biol*, 42:65–86.
- Burkholder, T. P., Kudlacz, E. M., Le, T.-B., Knippenberg, R. W., Shatzer, S. A., Maynard, G. D., Webster, M. E., & Horgan, S. W. (1996). Identification and Chemical Synthesis of MDL 105,212, a non-peptide tachykinin antagonist with high affinity for NK_1 and NK_2 Receptors. *Bioorg Med Chem Lett*, 6(8):951–956.
- Bursavich, M. G. & Rich, D. H. (2002). Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles. *J Med Chem*, 45:541–558.
- Bursi, R. & Grootenhuis, P. D. (1999). Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J Comput Aided Mol Des*, 12:221–232.
- Böhm, H.-J., Klebe, G., & Kubinyi, H. (1996). *Wirkstoffdesign*. Spektrum Akademischer Verlag.
- Carlson, H. A. & McGammon, J. A. (2000). Accomodating Protein Flexibility in Computational Drug Design. *Mol Pharmacol*, 57:213–218.
- Cascieri, M. A., Ber, E., Fong, T. M., Sadowski, S., Bansal, D., Swain, C. J., Seward, E., Frances, B., Burns, D., & Strader, C. D. (1992). Characterization of the binding of a potent, selective, radioiodinated antagonist to the human neurokinin-1 receptor. *Mol Pharmacol*, 42(3):458–63.
- Cavasotto, C. N., Orry, A. J. W., & Abagyan, R. A. (2003). Structure-Based Identification of Binding Sites, Native Ligands and Potential Inhibitors for G-Protein Coupled Receptors. *Proteins*, 51: 423–433.
- Chalmers, D. T. & Behan, P. B. (2002). The use of constitutively active GPCRs in drug discovery and functional genomics. *Nat Rev Drug Discov*, 1(8):599–608.
- Chambers, J. C. & Nichols, D. E. (2002). A homology-based model of the human 5 – HT_{2A} receptor derived from an in silico activated G-protein coupled receptor. *J Comput Aided Mol Des*, 16: 511–520.
- Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). Consensus Scoring: A method for obtaining improved hit-rates from docking databases of three-dimensional structures into proteins. *J Med Chem*, 42:5100–5109.

- Cho, S. J., Garsia, M. L. Bier, J., & Tropsha, A. (1996). Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors. *J Med Chem*, 39(26):5064–5071.
- Claessens, M., Cutsem, E. V., Lasters, I., & Wodak, S. (1989). Modelling the polypeptide backbone with 'sparse parts' from known protein structures. *Protein Eng*, 4:335–345.
- Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F., & Matthew, J. B. (2002). Consensus scoring for ligand/protein interactions. *J Mol Graphics Mod*, 20:281–295.
- Claussen, H., Buning, C., Rarey, M., & Lengauer, T. (2001). FLEXE: Efficient Molecular Docking Considering Protein Structure Variations. *J Mol Biol*, 308:377–395.
- Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986). Application of molecular dynamics with interproton distance restraints to 3D protein structure determination. *J Mol Biol*, 191:523–555.
- Clore, G. M. & Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science*, 252:1390–1399.
- Colovos, C. & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 2(9):1511–9.
- Costa, T. & Hertz, A. (1989). Antagonists with negative intrinsic activity at the δ opioid receptors coupled to GTP binding proteins. *Proc Natl Acad Sci U S A*, 86:7321–7325.
- 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 Am Chem Soc*, 110(18): 5959–5967.
- Crasto, C., Marenco, L., Miller, P. L., & Shepherd, G. S. (2002). Olfactory Receptor Database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Res*, 1:354–360.
- Davies, A., Schertler, G. F., Gowen, B. E., & Saibil, H. R. (1996). Projection structure of an invertebrate rhodopsin. *J Struct Biol*, 117:36–44.
- Davis, A. D. & Teague, S. T. (1999). Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angew Chem Int Ed Engl*, 38:736–749.
- Desai, M. C., Lefkowitz, S. L., Thadeio, P. F., Longo, K. P., & Snider, R. M. (1992). Discovery of a potent substance P antagonist: Recognition of key molecular determinant. *Bioorg Med Chem Lett*, 35:4911–4913.
- Desai, M. C., Thadeio, P. F., & Lefkowitz, S. L. (1993). Synthesis of (\pm)-CP-99,994: A highly potent substance P antagonist. *Tetrahedron Lett*, 34:5831–5834.
- DeWitte, R. S. & Shakhnovich, E. I. (1996). Smog: De Novo Design Method Based On Simple, Fast, Accurate Free Energy Estimates .1. Methodology Supporting Evidence. *J Am Chem Soc*, 118(47): 11733–11744.

- Dinner, A. R., Sali, A., & Karplus, M. (1996). The folding mechanism of larger model proteins: Role of native structure. *Proc Natl Acad Sci*, 98:8356–8361.
- Dominy, B. N. & Brooks, C. L. (2002). Identifying Native-Like Protein Structures Using Physics-Based Potentials. *J Comp Chem*, 23:147–160.
- Donnelly, D., Maudsley, S., Gent, J. P., Moser, R. N., Hurrell, C. R., & Findlay, J. B. C. (1999). Conserved polar residues in the transmembrane domain of the human tachykinin NK_2 receptor: functional roles and structural implications. *Biochem J*, 399:55–61.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography*. Springer, New York.
- Duan, Y. & Kollman, P. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740–744.
- Edvardsen, O. & Kristiansen, K. (1997). Computerization of mutant data: the tinyGRAP mutant database. *7TM journal*, 6:1–6.
- Eisenberg, D., Bowie, J. U., Luthy, R., & Choe, S. (1992). Three-dimensional profiles for analysing protein sequence-structure relationships. *Faraday Discuss*, 93:25–34.
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, 11:425–445.
- Elling, C. E., Nielsen, S. M., & Schwartz, T. W. (1995). Conversion of antagonist-binding site to metal-ion site in the tachykinin NK-1 receptor. *Nature*, 374:74–77.
- Elliott, J. M., Broughton, H., Cascieri, M. A., Chicchi, G., Huscroft, I. T., Kurtz, M., MacLeod, A. M., Sadowski, S., & Stevenson, G. I. (1998). Serine derived NK1 antagonists 2: A pharmacophore model for arylsulfonamide binding. *Bioorg Med Chem Lett*, 8:1851–1856.
- Emonds-Alt, X., Vilain, P., Goulaouic, P., Proietto, V., van Broeck, D., Advenier, C., Naline, E., Neliat, G., Le Fur, G., & Breliere, J. C. (1992). A potent and selective nonpeptide antagonist of the neurokinin A (NK_2) receptor. *Life Sci Pharmacol Lett*, 50:101–106.
- Escherich, A., Lutz, J., Escrieut, C., Fourmy, D., van Neuren, A. S., Muller, G., Schafferhans, A., Klebe, G., & Moroder, L. (2001). Peptide/benzodiazepine hybrids as ligands of CCK(A) and CCK(B) receptors. *Biopolymers*, 56(2):55–76.
- Ewing, T. J. A. & Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *J Comput Chem*, 18:1175–1189.
- Ewing, T. J. A., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*, 15:411–428.
- Fideslis, K., Stern, P. S., Bacon, D., & Moulton, J. (1994). Comparison of systematic search and database methods for constructing segments of protein structure. *Protein Eng*, 7:953–960.

- Filizola, M., Perez, J. J., & Carteni-Farina, M. (1998). BUNDLE: A program for building the transmembrane domains of G-protein-coupled receptors. *J Comput Aided Mol Des*, 12:111–118.
- Fiser, A., Do, R. K. G., & Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci*, 9: 1753–1773.
- Flohr, S., Kurz, M., Kostenis, E., Brkovich, A., Fournier, A., & Klabunde, T. (2002). Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II. *J Med Chem*, 45(9):1799–1805.
- Flower, D. R. (1999). Modelling G-protein-coupled receptors for drug design. *Biochim et Biophys Acta*, 1422:207–234.
- Fong, T. M., Cascieri, M. A., Yu, H., Bansal, A., Swain, C., & Strader, C. D. (1993). Amino-aromatic interaction between histidine 197 of the neurokinin-1 Receptor and CP 96345. *Nature*, 362:350–353.
- Fong, T. M., Yu, H., Cascieri, M. A., Underwood, D., Swain, C. J., & Strader, C. D. (1994a). Interaction of Glutamine 165 in the Fourth Transmembrane Segment of the Human Neurokinin-1 Receptor with Quinuclidine Antagonists. *J Biol Chem*, 269(21):14957–14961.
- Fong, T. M., Yu, H., Cascieri, M. A., Underwood, D., Swain, C. J., & Strader, C. D. (1994b). The Role of Histidine 265 in Antagonist Binding to the Neurokinin-1 Receptor. *J Biol Chem*, 269(4): 2728–2732.
- Fong, T. M., Yu, H., & Strader, C. D. (1992). Molecular Basis for the Species Selectivity of the Neurokinin-1 Receptor Antagonists CP-96,345 and RP67580. *J Biol Chem*, 267(36):25668–25671.
- Fradera, X., Knegtel, R. M. A., & Mestres, J. (2000). Similarity-Driven Flexible Ligand Docking. *Proteins*, 40:623–636.
- Frank, H. S. & Evans, M. W. (1945). Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molar entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J Chem Phys*, 13:507–532.
- Gamper, A. M., Winger, R. H., Liedl, K. R., Sotriffer, C. A., Varga, J. M., Kroemer, R. T., & Rode, B. M. (1996). Comparative Molecular Field Analysis of Haptens Docked to the Multispecific Antibody IgE(Lb4). *J Med Chem*, 39:3882–3888.
- Garcia-Nieto, R., Perez, C., & Gago, F. (2000). Automated docking and molecular dynamics simulations of nimesulide in the cyclooxygenase active site of human prostaglandin-endoperoxide synthase-2 (COX-2). *J Comput Aided Mol Des*, 14(2):147–60.
- Garret, C., Carruette, A., Fardin, V., Moussaoui, S., Peyronel, J. F., Blanchard, J. C., & Laduron, P. M. (1991). Pharmacological properties of a potent and selective nonpeptide substance P antagonist. *Proc Natl Acad Sci U S A*, 88(22):435–7.
- Gasteiger, J., Rudolph, C., & Sadowski, J. (1990). Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.*, 3:537–547.

- Geladi, P. & Kowalski, B. R. (1986). Partial least squares regression: a tutorial. *Anal Chim Acta*, 185:1–17.
- Gerber, P. R. (1998). Charge Distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force field MAB. *J Comput Aided Mol Des*, 12:37–51.
- Gerber, P. R. & Müller, K. (1995). MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J Comput Aided Mol Des*, 9:251–268.
- Gershengorn, M. C. & Osman, R. (2001). Minireview: Insight into G Protein-Coupled Receptor Function Using Molecular Models. *Endocrinology*, 142:2–10.
- Gerspacher, M., Vecchia, L. L., Mah, R., von Sprecher, A., Anderson, G. P., Subramanian, N., Hauser, K., Bammerlin, H., Kimmel, S., Pawelzik, V., Ryffel, K., & Ball, H. A. (2001). Dual Neurokinin NK_1/NK_2 Antagonists: N-[(R,R)-(E)-1-arylmethyl-3-(2-oxo-azepan-3-yl)carbamoyl]allyl-N-methyl-3,5-bis(trifluormethyl)benzamides and 3-[N'-3,5-bis(trifluormethyl)benzoyl-N-arylmethyl-N'-methylhydrazino-N-[(R)-2-oxo-azepan-3-yl]propionamides. *Bioorg Med Chem Lett*, 11: 3081–3084.
- Gether, U. (2000). Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocrine Reviews*, 21:90–113.
- Gether, U., Emonds-Alt, X., Breliere, J. C., Fujii, T., Hagiwara, D., Pradier, L., Garret, C., Johansen, T. E., & Schwartz, T. W. (1994a). Evidence for a common molecular mode of action for chemically distinct nonpeptide antagonists at the neurokinin-1 (substance P) receptor. *Mol Pharmacol*, 45(3): 500–508.
- Gether, U., Johansen, T. E., Snider, R. M., Lowe III, J. A., Nakanishi, S., & Schwartz, T. W. (1993a). Different binding epitopes on the NK1 receptor for substance P and non-peptide antagonist. *Nature*, 362(6418):345–348.
- Gether, U., Nilsson, L., Lowe III, J. A., & Schwartz, T. W. (1994b). 2 Specific Residues at the Top of Transmembrane Segment V and VI of the Neurokinin-1 Receptor Involved in Binding of the Nonpeptide Antagonist CP 96,345. *J Biol Chem*, 269(39):23959–23964.
- Gether, U., Yokata, Y., Emonds-Alt, X., Breliere, J.-C., Lowe III, J. A., Snider, R. M., Nakanishi, S., & Schwartz, T. W. (1993b). Two nonpeptide tachykinin antagonists act through epitopes of corresponding segments of the NK1 and NK2 receptors. *Proc Natl Acad Sci U S A*, 90:6194–6198.
- Giardina, A. M. G., Raveglia, L. F., & Grugni, M. (1997). Lead generation and lead optimization processes in the discovery of selective nonpeptide neurokinin receptor antagonists. *Drugs Future*, 22(11):1235–1257.
- Gieldon, A., Kazmierkiewicz, R., Slusarz, R., & Ciarkowski, J. (2001). Molecular modeling of interactions of the non-peptide antagonist YM087 with the human vasopressin V1a, V2 receptors and with oxytocin receptors. *J Comput Aided Mol Des*, 15(12):1085–104.

- Giolitti, A., Cucchi, P., Renzetti, A. R., Rotondaro, L., Zappitelli, S., & Maggi, C. A. (2000). Molecular determinants of peptide and nonpeptide NK-2 receptor antagonists binding sites of the human tachykinin NK-2 receptor by site-directed mutagenesis. *Neuropharmacology*, 39:1422–1429.
- Glusker, J. P., Lewis, M., & Rossi, M. (1994). *Crystal Structure Analysis for Chemists and Biologists*. VCH, Weinheim.
- Godzik, A., Kolinski, A., & Skolnick, J. (1993). De novo and inverse folding predictions of protein structure and dynamics. *J Comput Aided Mol Des*, 7:397–438.
- Gohlke, H., Hendlich, M., & Klebe, G. (2000a). Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*, 295(2):337–356.
- Gohlke, H., Hendlich, M., & Klebe, G. (2000b). Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Perspect Drug Discov Des*, 20:115–144.
- Gohlke, H. & Klebe, G. (2002a). Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew Chem Int Ed*, 41:2644–2676.
- Gohlke, H. & Klebe, G. (2002b). DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J Med Chem*, 45(19):4153–70.
- Goodsell, D. S. & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8(3):195–202. Using Smart Source Parsing.
- Gouldson, P. R., Snell, C. R., & Reynolds, C. A. (1997). A new approach to docking in the beta 2-adrenergic receptor that exploits the domain structure of G-protein-coupled receptors. *J Med Chem*, 40(24):3871–86.
- Greenfeder, S., Cheewatrakoolpong, B., Anthes, J., Billah, M., Egan, R. W., Brown, J. E., & Murgolo, N. J. (1998). Two related neurokinin-1 receptor antagonists have overlapping but different binding sites. *Bioorg Med Chem*, 6(2):189–194.
- Greenfeder, S., Cheewatrakoolpong, B., Billah, M., Egan, R. W., Keene, E., & Murgolo, N. J. (1999). The neurokinin-1 and neurokinin-2 receptor binding sites of MDL103,392 differ. *Bioorg Med Chem*, 7(12):2867–2876.
- Greer, J. (1981). Comparative model-building of the mammalian serine proteases. *J Mol Biol*, 153:1027–1042.
- Grootenhuis, P. D. J. & van Galen, P. J. M. (1995). Correlation of binding affinities with nonbonded interaction energies of thrombin inhibitor complexes. *Acta Crystallogr Sect D*, 51:560–566.
- Grüneberg, S., Stubbs, M. T., & Klebe, G. (2002). Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem*, 45(17):3588–3602.

- Guertin, K. R., Gardner, C. J., Klein, S. I., Zulli, A. L., Czekaj, M., Gong, Y., Spada, A. P., Cheney, D. L., Maignan, S., Guilloteau, J. P., Brown, K. D., Colussi, D. J., Chu, V., Heran, C. L., Morgan, S. R., Bentley, R. G., Dunwiddie, C. T., Leadley, R. J., & Pauls, H. W. (2002). Optimization of the beta-aminoester class of factor Xa inhibitors. Part 2: Identification of FXV673 as a potent and selective inhibitor with excellent In vivo anticoagulant activity. *Bioorg Med Chem Lett*, 12(12): 1671–4.
- Günther, J., Bergner, A., Hendlich, M., & Klebe, G. (2003). Utilising Structural Knowledge in Drug Design Strategies: Applications Using Relibase+. *J Mol Biol*, 326:621–636.
- Halgren, T. A. (1996). Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J Comp Chem*, 17(5 & 6):490–519.
- Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins*, 47:409–443.
- Hann, M. M., Leach, A. R., & Harper, G. (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci*, 41:856–864.
- Hao, M.-H. & Scheraga, H. A. (1999). Designing potential energy functions for protein folding. *Curr Opin Struct Biol*, 9:184–188.
- Hardin, C., Pogorelov, T. V., & Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. *Curr Opin Struct Biol*, 12:176–181.
- Havel, T. F. & Snow, M. E. (1991). A new method for building protein conformations by automatic segment matching. *J Mol Biol*, 217:1–7.
- Head, R. D., Smythe, M. L., Oprea, T. L., Waller, C. L., Green, S. M., & Marshall, G. M. (1996). VALIDATE: A New Method for the receptor-based prediction of binding-affinities of novel ligands. *J Am Chem Soc*, 118:3959–3969.
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol*, 213:899–929.
- Hendlich, M. (1998). Databases for protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, 54 (1 (Pt 6)):1178–82.
- Hendlich, M., Bergner, A., Gunther, J., & Klebe, G. (2003). Relibase: Design and Development of a Database for Comprehensive Analysis of Protein-Ligand Interactions dagger. *J Mol Biol*, 326(2): 607–20.
- Hindle, S. A., Rarey, M., Buning, C., & Lengauer, T. (2002). Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des*, 16:129–149.
- Hinds, D. & Levitt, M. (1992). A lattice model for protein structure prediction at low resolution. *Proc Natl Acad Sci USA*, 89:2536–2540.

- Hinds, D. & Levitt, M. (1994). Exploring conformational space with a simple lattice model for protein structure. *J Mol Biol*, 243:668–682.
- Hodgkin, E. E. & Richards, W. G. (1987). Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int J Quant Chem: Quant Biol Symp*, 14:105–110.
- Holloway, M. K., Wai, J. M., Halgren, T. A., Fitzgerald, P. M., Vacca, J. P., Dorsey, B. D., Levin, R. B., Thompson, W. J., Chen, L. J., & deSolms, S. J. (1995). A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J Med Chem*, 38:305–317.
- Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, 273(5275):595–603.
- Holst, B., Zoffmann, S., Elling, C. E., Hjorth, S. A., & Schwartz, T. W. (1998). Steric Hindrance Mutagenesis versus Alanine Scan in Mapping of Ligand Binding Sites in the Tachykinin NK1 Receptor. *Mol Pharmacol*, 53:166–175.
- Hooft, R. W. W., Vriend, G., Sander, C., & Abola, E. E. (1996). Errors in protein structures. *Nature*, 381:272–272.
- Hopkins, A. L. & Groon, C. R. (2002). The druggable genome. *Nat Rev Drug Discov*, 1(9):727–730.
- Horn, F., Weare, J., Beukers, M. W., Hörsch, S., Bairoch, A., Chen, W., Edvardsen, O., Campagne, F., & Vriend, G. (1998). GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res*, 26:275–279.
- Huang, R.-R. C., Vicario, P. P., Strader, C. D., & Fong, T. M. (1995). Identification of Residues Involved in Ligand Binding to the Neurokinin-2 Receptor. *Biochemistry*, 34:10048–10055.
- Huang, R.-R. C., Yu, H., Strader, C. D., & Fong, T. M. (1994). Interaction of Substance P with the Second and Seventh Transmembrane Domains of the Neurokinin-1 Receptor. *Biochemistry*, 33:3007–3013.
- Jackson, R. M., Gabb, H. A., & Sternberg, M. J. E. (1998). Rapid refinement of protein interfaces incorporating solvation. *J Mol Biol*, 276:265–285.
- Jacoby, E. (2001). A Novel Chemogenomics Knowledge-Based Ligand Design Strategy — Application to G Protein-Coupled Receptors. *Quant Struct-Act Relat*, 20:115–123.
- Jacoby, E., Boudon, A., Kucharczyk, N., Michel, A., & Fauchere, J. L. (1997). A structural Rationale for the Design of Water Soluble Peptide-derived Neurokinin-1 antagonists. *J Recep Signal Transduction Res*, 17(6):855–873.
- Jacoby, E., Fauchere, J.-L., Raimbaud, E., Ollivier, S., Michel, A., & Spedding, M. (1999). A Three Binding Site Hypothesis for the Interaction of Ligands with Monoamine G Protein-coupled Receptors: Implications for Combinatorial Ligand Design. *Quant Struct-Act Relat*, 18:561–572.
- Jain, A. (1996). Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des*, 10:427–440.

- Jalaie, M. & Erickson, J. A. (2000). Homology model directed alignment selection for comparative molecular field analysis: application to photosystem II inhibitors. *J Comput Aided Mol Des*, 14(2): 181–97.
- Jansen, J. M., Koehler, K. F., Hedberg, M. H., Johansson, A. M., Hacksell, U., Nordvall, G., & Snyder, J. P. (1997). Molecular design using the minireceptor concept. *J Chem Inf Comput Sci*, 37(4):812–8.
- Johnson, M. A., Hoog, C., & Pinto, B. M. (2003). A novel modeling protocol for protein receptors guided by bound-ligand conformation. *Biochemistry*, 42(7):1842–53.
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). A New Approach to Protein Fold Recognition. *Nature*, 358:86–89.
- Jones, D. T. & Thornton, J. M. (1993). Protein Fold Recognition. *J Comput Aided Mol Des*, 7: 439–456.
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and Validation of a Genetic Algorithm for Flexible Docking. *J Mol Biol*, 267:727–748.
- Jones, T. H. & Thirup, S. (1986). Using known substructures in protein model building and crystallography. *EMBO J*, 5:819–822.
- Jöhren, K. & Höltje, H.-D. (2002). A model of the human M_2 muscarinic acetylcholine receptor. *J Comput Aided Mol Des*, 16:795–801.
- Kamata, K., Kawamoto, H., Honma, T., Iwama, T., & Kim, S. H. (1998). Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proc Natl Acad Sci U S A*, 95(12):6630–5.
- Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem*, 14:1–63.
- Kearsley, S. K. & Smith, M. S. (1990). An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comp Methodol*, 3(6):615–633.
- Kim, K. H. *Building a Bridge Between G-Protein-Coupled Receptor Modelling, Protein Crystallography, and 3D-QSAR Studies for Ligand Design*, chapter Recent Advances in 3D QSAR in Drug Design. Kluwer Academic Publisher, Dordrecht, The Netherlands, 1998.
- Kiyama, R., Tamura, Y., Watanabe, F., Tsuzuki, H., Ohtani, M., & Yodo, M. (1999). Homology modeling of gelatinase catalytic domains and docking simulations of novel sulfonamide inhibitors. *J Med Chem*, 42(10):1723–38.
- Klabunde, T. & Hessler, G. (2002). Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem*, 3(10):928–44.
- Klapper, M. H. (1971). On the nature of the protein interior. *Biochim et Biophys Acta*, 229:557–566.

- Klebe, G. (1994). The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J Mol Biol*, 237(2):212–35.
- Klebe, G., Abraham, U., & Mietzner, T. (1994a). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem*, 37(24):4130–46.
- Klebe, G. & Mietzner, T. (1994). A fast and efficient method to generate biologically relevant conformations. *J Comput Aided Mol Des*, 8:583–606.
- Klebe, G., Mietzner, T., & Weber, F. (1994b). Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. *J Comput Aided Mol Des*, 8(6):751–78.
- Klebe, G., Mietzner, T., & Weber, F. (1999). Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J Comput Aided Mol Des*, 13(1):35–49.
- Knegtel, R. M. A., Kuntz, I. D., & Oshiro, C. M. (1997). Molecular Docking to Ensembles of Protein Structures. *J Mol Biol*, 266:424–440.
- Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chain conformations and estimate their conformational entropy. *J Mol Biol*, 239:249–275.
- Koehl, P. & Delarue, M. (1996). Mean-field minimization methods for biological macromolecules. *Curr Opin Struct Biol*, 6:222–226.
- Koehl, P. & Levitt, M. (1999). A brighter future for protein structure prediction. *Nature Struct Biol*, 6:108–111.
- Kolakowski Jr, L. F. (1994). a G-protein-coupled receptor database. *Receptors Channels*, 2:1–7.
- Kollman, P. (1993). Free energy calculations – applications to chemical and biological phenomena. *Chem Rev*, 7:2395–2417.
- Kollman, P. (1996). Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc Chem Res*, 29:461–469.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, T., Duan, Y., Wang, W., Domini, O., Cieplak, P., Srinivasan, J., Case, D., & Cheatham III, T. E. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum methods. *Acc Chem Res*, 33:889–897.
- Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX Incremental Construction Algorithm for Protein-Ligand Docking. *Proteins*, 37:228–241.
- Kristiansen, K., Dahl, S. G., & Edvardsen, O. (1996). A database of mutants and effects of site-directed mutagenesis experiments on G-protein coupled receptors. *Proteins*, 26:81–94.

- Kubinyi, H., Folkers, G., & Martin, Y. C. E. (1997). *3D QSAR in Drug Design. Recent Advances*. Kluwer/ESCOM, Leiden.
- Kubinyi, H. E. (1993). *3D QSAR in Drug Design. Theory, Methods and Applications*. ESCOM, Leiden.
- Kulkarni, S. S. & Kulkarni, V. M. (1999). Structure Based Prediction of Binding Affinity of Human Immunodeficiency Virus-1 Protease Inhibitors. *J Chem Inf Comput Sci*, 39:1128–1140.
- Kurinov, I. V. & Harrison, R. W. (1994). Prediction of new serine proteinase inhibitors. *Nat Struct Biol*, 1:735–743.
- Labrou, N. E., Bhogal, N., Hurrell, C. R., & Findlay, J. B. C. (2001). Interaction of *Met*²⁹⁷ in the Seventh Transmembrane Segment of the Tachykinin *NK*₂ Receptor with Neurokinin A. *J Biol Chem*, 276(41):37944–37949.
- Ladduwahetty, T., Baker, R., Cascieri, M. A., Chambers, M. S., Haworth, K., Keown, L. E., MacIntyre, D. E., Metzger, J. M., Owen, S., Rycroft, W., Sadowski, S., Seward, E. M., Shephard, S. L., Swain, C. J., Tattersall, F. D., Watt, A. P., Williamson, D. W., & Hargreaves, R. J. (1996). N-Heteroaryl-2-phenyl-(benzoxyl)piperidines: A novel Class of Potent Orally Active Human NK1 Antagonists. *J Med Chem*, 39:2907–2914.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissole, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*, 26:283–291.
- Le Novere, N., Grutter, T., & Changeux, J. P. (2002). Models of the extracellular domain of the nicotinic receptors and of agonist- and Ca²⁺-binding sites. *Proc Natl Acad Sci U S A*, 99(5):3210–5.
- Lefkowitz, R. J., Cotacchia, S., Samama, P., & Costa, T. (1993). Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol Sci*, 14:303–307.

- Lemmen, C. & Lengauer, T. (2000). Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des*, 14(3):215–32.
- Lessel, U. & Schomburg, D. (1999). Importance of Anchor Group Positioning in protein loop prediction. *Proteins*, 37(1):56–64.
- Levinthal, C. (1968). Are There Pathways for Protein Folding? *C J Chim Phys*, 65:44–45.
- Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol*, 226:507–533.
- Levitt, M., Gerstein, M., Huang, E., Subbiah, S., & Tsai, J. (1997). PROTEIN FOLDING: The Endgame. *Annu Rev Biochem*, 66:549–579.
- Lin, J.-H., Perryman, A. L., Schames, J. R., & McGammon, J. A. (2002). Computational Drug Design Accomodating Receptor Flexibility: The Relaxed Complex Scheme. *J Am Chem Soc*, 124: 5632–5633.
- Lin, J.-H., Perryman, A. L., Schames, J. R., & McGammon, J. A. (2003). The Relaxed Complex Method: Accomodating Receptor Flexibility for Drug Design with an Improved Scoring Scheme. *Biopolymers*, 68:47–62.
- Liwo, A., Kazmierkiewicz, R., Czaplewski, C., Groth, M., Oldziej, S., Wawak, R. J., Rackovsky, S., Pincus, M. R., & Scheraga, H. A. (1998). United residue force field for off-lattice protein-structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *J Comput Chem*, 19:259–276.
- Liwo, A., Oldziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., & Scheraga, H. A. (1997a). A united residue force field for off-lattice protein-structure simulations: I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem*, 18: 849–873.
- Liwo, A., Pincus, M. R., Wawak, R. J., Rackovsky, S., Oldziej, S., & Scheraga, H. A. (1997b). A united residue force field for off-lattice protein-structure simulations: II. Parameterization of short-range interactions and determination of the weights of energy terms by Z-score optimization. *J Comput Chem*, 18:874–887.
- Lopez-Rodriguez, M. L., Murcia, M., Benhamu, B., Olivella, M., Campillo, M., & Pardo, L. (2001). Computational model of the complex between GR113808 and the 5-HT4 receptor guided by site-directed mutagenesis and the crystal structure of rhodopsin. *J Comput Aided Mol Des*, 15(11): 1025–33.
- Lowe, J. A., I., Drozda, S. E., McLean, S., Bryce, D. K., Crawford, R. T., Snider, R. M., Longo, K. B., Nagahisa, A., & Tsuchiya, M. (1994). Aza-tricyclic substance P antagonists. *J Med Chem*, 37:2831–2840.

- Lowe, J. A., I., Drozda, S. E., Snider, R. M., Longo, K. P., Zorn, S. H., Morrone, J., Jackson, E. R., McLean, S., Bryce, D. K., Bordner, J., Nagahisa, A., Kanai, Y., Suga, O., & Tsuchiya, A. (1992). The Discovery of (2S,3S)-cis-2-(Diphenylmethyl)-N-[(2-methoxyphenyl)methyl]-1-azabicyclo[2.2.2]-octan-3-amine as a Novel, Nonpeptide Substance P Antagonist. *J Med Chem*, 35:2591–2600.
- Lozano, J. J., Lopez-de Brinas, E., Centeno, N. B., Guigo, R., & Sanz, F. (1997). Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ. *J Comput Aided Mol Des*, 11(4):395–408.
- Lozano, J. J., Pastor, M., Gruciani, G., Gaedt, K., Centeno, N. B., Gago, F., & Sanz, F. (2000). 3D-QSAR methods of the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J Comput Aided Mol Des*, 13: 341–353.
- Luecke, H., Schobert, H. T., & Richter, J. P. (1999). Structure of Bacteriorhodopsin at 1.55 ÅResolution. *J Mol Biol*, 291:899–911.
- Luthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356(6364):83–5.
- Ma, B., Kumar, S., Tsai, C.-J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Engineering*, 12(9):713–720.
- MacLeod, A., Merchant, A. M., Cascieri, M. A., Sadowski, S., Ber, E., Swain, C. J., & Baker, R. (1993). N-acyl-L-tryptophan benzyl esters: potent substance P receptor antagonists. *J Med Chem*, 36:2044–2045.
- Mah, R., Gerspacher, M., von Sprecher, A., Stutz, S., Tschinke, V., Anderson, G. P., Bertrand, C., Subramanian, N., & Ball, H. A. (2002). Biphenyl Derivatives as Novel Dual NK_1/NK_2 -Receptor Antagonists. *Bioorg Med Chem Lett*, 12:2065–2068.
- Maignan, S., Guilloteau, J. P., Pouzieux, S., Choi-Sledeski, Y. M., Becker, M. R., Klein, S. I., Ewing, W. R., Pauls, H. W., Spada, A. P., & Mikol, V. (2000). Crystal structures of human factor Xa complexed with potent inhibitors. *J Med Chem*, 43(17):3226–3232.
- Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science*, 285 (5428):751–3.
- Marhefka, C. A., Moore, B. M., n., Bishop, T. C., Kirkovsky, L., Mukherjee, A., Dalton, J. T., & Miller, D. D. (2001). Homology modeling using multiple molecular dynamics simulations and docking studies of the human androgen receptor ligand binding domain bound to testosterone and nonsteroidal ligands. *J Med Chem*, 44(11):1729–40.
- Marinissen, M. J. & Gutkind, J. S. (2001). G protein coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol Sci*, 22:368–376.

- Marriott, D. P., Dougall, I. G., Meghani, P., Liu, Y. J., & Flower, D. R. (1999). Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J Med Chem*, 42(17):3210–3216.
- Marshall, G. R., Barry, C. D., Bossard, H. E., Dammkoehler, R. A., & Dunn, D. A. *Computer-Assisted Drug Design, ACS Symp Series*, volume 112, chapter The Conformational Parameter in Drug Design: The Active Analog Approach, Seiten 205–226. Amer. Chem. Soc., Washington DC., 1979.
- Marti-Renom, M. A., Madhusudhan, M. S., Fiser, A., Rost, B., & Sali, A. (2002). Reliability of assessment of protein structure prediction methods. *Structure*, 10:435–440.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., & Sali, A. (2000). Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biomol. Struct.*, 29:291–325.
- Massova, I. & Kollman, P. A. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov Des*, 18:113–135.
- McCarthy, J., Hogle, J. M., & Karplus, M. (1997). Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins*, 29:32–58.
- McConkey, B. J., Sobolev, V., & Edelman, M. (2003). Discrimination of native protein structures using atom-atom contact scoring. *Proc Natl Acad Sci U S A*, 100(6):3215–3220.
- McGovern, S. L. & Shoichet, B. K. (2003). Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes. *J Med Chem*, 46:2895–2907.
- Melo, F. & Feytmans, E. (1997). Novel knowledge-based mean force potential at atomic level. *J Mol Biol*, 267:207–222.
- Melo, F. & Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol*, 277(5):1141–52.
- Menzaghi, F., Behan, D. P., & Chalmers, D. T. (2002). Constitutively activated G protein coupled receptors: a novel approach to CNS drug discovery. *Curr Drug Targets*, 1:105–121.
- Mestres, J., Rohrer, D. C., & Maggiora, D. M. (1998). MIMIC: A molecular-field matching program. Exploiting applicability of molecular similarity approaches. *J Comp Chem*, 18(7):934–954.
- Mitchell, J. B. O., Alex, A., & Snarey, M. (1999a). SATIS: Atom Typing from Chemical Connectivity. *J Chem Inf Comput Sci*, 39:751–757.
- Mitchell, J. B. O., Laskowski, R. A., Alex, A., & Thornton, J. M. (1999b). BLEEP — Potential of Mean Force Describing Protein-Ligand Interactions: I. Generating Potential. *J Comp Chem*, 20(11):1165–1176.
- Moro, S., Guo, D., Camaioni, E., Boyer, J. L., Harden, T. K., & Jacobson, K. A. (1998a). Human P2Y1 receptor: molecular modeling and site-directed mutagenesis as tools to identify agonist and antagonist recognition sites. *J Med Chem*, 41(9):1456–66.

- Moro, S., Li, A. H., & Jacobson, K. A. (1998b). Molecular modeling studies of human A3 adenosine antagonists: structural homology and receptor docking. *J Chem Inf Comput Sci*, 38(6):1239–48.
- Morris, G. M., Goodsell, D. S., Halliday, R., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J Comp Chem*, 19:1639–1662.
- Morris, G. M., Goodsell, D. S., Huey, R., & Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J Comput Aided Mol Des*, 10: 293–304.
- Moult, J., Fidelis, K., Zemla, A., & Hubbard, T. (2001). Critical Assessment of Methods of Protein Structure Prediction (CASP): Round IV. *Proteins*, Suppl. 5:2–7.
- Moult, J., Hubbard, T., Bryant, S. H., Fidelis, K., & Peterson, J. T. (1997). Critical Assessment of Methods of Protein Structure Prediction (CASP): round II. *Proteins*, suppl 1:2–6.
- Moult, J., Hubbard, T., Fidelis, K., & Peterson, J. T. (1999). Critical Assessment of Methods of Protein Structure Prediction (CASP): round III. *Proteins*, suppl 3:2–6.
- Moult, J., Judson, R., Fidelis, K., & Peterson, J. T. (1995). A large-scale experiment to assess protein structure prediction methods. *Proteins*, 23:ii–iv.
- Muegge, I. (2000). A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspect Drug Discov Des*, 20:99–114.
- Muegge, I. (2001). Effect of ligand volume correction on PMF scoring. *J Comp Chem*, 22:418–425.
- Muegge, I. & Martin, Y. C. (1999). A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J Med Chem*, 42:791–804.
- Müller, G. (2000). Towards 3D Structures of G Protein-Coupled Receptors: A Multidisciplinary Approach. *Curr Med Chem*, 7:861–888.
- Murray, C. W., Auton, T. R., & Eldridge, M. D. (1998). Empirical scoring functions: II. The testing of an empirical scoring function for the prediction of ligand-receptor affinities and the use of Bayesian regression to improve the quality of the model. *J Comput Aided Mol Des*, 12:503–519.
- Murray, C. W., Baxter, C. A., & Frenkel, A. D. (1999). The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *J Comput Aided Mol Des*, 13:547–562.
- Najmanovich, R., Kuttner, J., Sobolev, V., & Edelman, M. (2000). Side-chain flexibility in proteins upon ligand binding. *Proteins*, 39(3):261–8.
- Nar, H., Bauer, M., Schmid, A., Stassen, J. M., Wienen, W., Pripke, H. W., Kauffmann, I. K., Ries, U. J., & Haue, N. H. (2001). Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure (Camb)*, 9(1):29–37.

- Natsugari, H., Ikeura, Y., & Kiyota, Y. (1995). Novel potent, and orally active substance P antagonists: Synthesis and antagonist activity of N-benzylcarboxamide derivatives of pyrido[3,4-b]pyridine. *J Med Chem*, 38:3106–3120.
- Nederkoorn, P. H., Timmerman, H., & Donne-Op den Kleder, G. M. (1995). Does the ternary complex act as a secondary proton pump and a GTP synthase? *Trends Pharmacol Sci*, 16:156–161.
- Nishi, T., Fukazawa, T., Ishibashi, K., Nakajima, K., Sugioka, Y., Iio, Y., Kurata, H., Itoh, K., Mukaiyama, O., Satoh, Y., & Yamaguchi, T. (1999). Combined NK_1 and NK_2 Tachykinin Antagonists: Synthesis and Structure-Activity Relationships of Novel Analogues. *Bioorg Med Chem Lett*, 9:875–880.
- Ofner, S., Hauser, K., Schilling, W., Vassout, A., & Veenstra, S. J. (1996). SAR of 2-benzyl-4-aminopiperidines: CGP-49823, an orally and centrally active non-peptide NK-1 antagonist. *Bioorg Med Chem Lett*, 6:1623–1628.
- Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., & Shichida, Y. (2002). Functional role of internal water molecules in rhodopsin revealed by x ray crystallography. *Proc Natl Acad Sci USA*, 99:5982–5987.
- Oliveira, L., Hulsen, T., Hulsik, D. L., Paiva, A. C. M., & Vriend, G. (2002). Modelling G protein-coupled receptors. *in preparation*.
- Oliveira, L., Paiva, A. C., & Vriend, G. (1993). A common motif in G protein-coupled seven transmembrane helix receptors. *J Comput Aided Mol Des*, 7:649–658.
- Oliveira, L., Paiva, A. C. M., & Vriend, G. (1999). A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng*, 12(12):1087–1095.
- Oprea, T. I., Davis, A. M., Teague, S. J., & Leeson, P. D. (2001). Is There a Difference between Leads and Drugs? A Historical Perspective. *J Chem Inf Comput Sci*, 41:1308–1315.
- Ortiz, A. R., Kolinski, A., & Skolnick, L. (1998). Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments. *J Mol Biol*, 277:419–448.
- Ortiz, A. R., Pastor, M., Palomer, A., Cruciani, G., Gago, F., & Wade, R. C. (1997). Reliability of Comparative Molecular Field Analysis Models: Effects of Data Scaling and Variable Selection Using a Set of Human Synovial Fluid Phospholipase A_2 Inhibitors. *J Med Chem*, 40:1136–1148.
- Ortiz, A. R., Pisabarro, M. T., Gago, F., & Wade, R. C. (1995). Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J Med Chem*, 38:2681–2691.
- Osguthorpe, D. J. (1997). Analysis of the predicted structures of domain 1 of protein G3 (T0030) and NK-Lysin (T0042). *Proteins*, suppl 1:172–178.
- Osguthorpe, D. J. (1999). Improved Ab Initio Predictions With a Simplified, Flexible Geometry Model. *Proteins*, suppl 3:186–193.

- Osguthorpe, D. J. (2000). Ab initio protein folding. *Curr Opin Struct Biol*, 10:146–152.
- Österberg, F., Morris, G. M., Sanner, M. F., Olson, A. J., & Goodsell, D. S. (2002). Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*, 46(1):34–40.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., H., M., Fox, B. A., Le Trong, I., Teller, D. C., T., O., Stenkamp, R. E., Yamamoto, M., & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, 289(5480):739–45.
- Pan, Y., Huang, N., Cho, S., & MacKerell Jr., A. D. (2002). Consideration of Molecular Weight during Compound Selection in Virtual Target-Based Database Screening. *J Chem Inf Comput Sci*, 43(1):267–272.
- Patacchini, R. & Maggi, C. A. (2001). Peripheral tachykinin receptors as targets for new drugs. *Eur J Pharmacol*, 429(1-3):13–21.
- Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., & Landau, E. M. (1997). X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, 277: 1676–1681.
- Petrella, R. J., Lazirdis, T., & Karplus, M. (1998). Protein sidechain conformer prediction: a test of the energy function. *Folding & Design*, 3:353–377.
- Pieper, U., Narayanan, E., Ashley, C. S., Ilyin, V. A., & Sali, A. (2002). MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res*, 30(1):255–259.
- Pillard, J., Czaplewski, C., Liwo, A., Lee, J., Ripoll, D. R., Kazmierkiewicz, R., Oldziej, S., Wedemeyer, W. J., Gibson, K. D., Arnautova, Y. A., Saunders, J., J., Y. Y., & Scheraga, H. A. (2001). Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc Natl Acad Sci U S A*, 98(5):2329–2333.
- Pontius, J., Richelle, J., & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*, 264(1):121–36.
- Predix. URL: http://www.predixpharm.com/comp_drug.htm.
- Probst, W. C., Snyder, L. A., Schuster, D. J., Brosius, J., & Sealfon, S. C. (1992). Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol*, 11:1–20.
- Qi, H., Shah, S. K., Cascieri, M. A., Sadowski, S. J., & MacCoss, M. (1998). L-Tryptophan urea amides as NK_1/NK_2 dual antagonists. *Bioorg Med Chem Lett*, 8:2259–2262.
- Rarey, M., Kramer, B., & Lengauer, T. (1997). Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. *J Comput Aided Mol Des*, 11:369–384.
- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996a). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3):470–89.

- Rarey, M., Wefing, S., & Lengauer, T. (1996b). Placement of medium-sized molecular fragments into active sites of proteins. *J Comput Aided Mol Des*, 10:41–54.
- Reichard, G. A., Grice, C. A., Shih, N.-Y., Spitler, J., Majmunder, S., Wang, S. D., Paliwal, S., Anthes, J. C., & Piwinski, J. J. (2002). Preparation of Oxime Dual NK_1/NK_2 Antagonist with Reduced NK_3 Affinity. *Bioorg Med Chem Lett*, 12:2355–2358.
- Renzetti, A. R., Catalioto, R. M., Carloni, C., Criscuoli, M., Cucchi, P., Giolitti, A., Zappitelli, S., Rotondaro, L., & Maggi, C. A. (1999a). Effects of Tyrosine289Phenylalanine Mutation of Binding and Functional Properties of the Human Tachykinin NK_2 Receptor Stably Expressed in Chinese Hamster Ovary Cells. *Biochem Pharmacol*, 57:899–906.
- Renzetti, A. R., Catalioto, R. M., Criscuoli, M., Cucchi, P., Ferrer, C., Giolitti, A., Guelfi, M., Rotondaro, L., Warner, F. J., & Maggi, C. A. (1999b). Relevance of Aromatic Residues in Transmembrane Segments V to VII for Binding of Peptide and Nonpeptide Antagonists to the Human Tachykinin NK_2 Receptor. *J Pharmacol Exp Ther*, 290(2):487–495.
- Rognan, D., Lauemoller, S. L., Holm, A., Buus, S., & Tschinke, V. (1999). Predicting Binding Affinities of Protein Ligands from Three-Dimensional Models: Application to Peptide Binding to Class I Major Histocompatibility Proteins. *J Med Chem*, 42:4650–4658.
- Rong, S. B., Zhang, J., Neale, J. H., Wroblewski, J. T., Wang, S., & Kozikowski, A. P. (2002). Molecular modeling of the interactions of glutamate carboxypeptidase II with its potent NAAG-based inhibitors. *J Med Chem*, 45(19):4140–52.
- Roth, B. L., Kroeze, W. K., Patel, S., & Lopez, E. (2000). The Multiplicity of Serotonin Receptors: Uselessly diverse molecules or an embarrassment of riches? *The Neuroscientist*, 6:252–262.
- Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vossall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., & Lewis, S. (2000). Comparative genomics of the eukaryotes. *Science*, 287(5461):2204–15.
- Russell, R. B. & Eggleston, D. S. (2000). New roles for structure in biology and drug discovery. *Nature Struct Biol*, Seiten 928–930.
- Sadowski, J. & Gasteiger, J. (1993). From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Atomic Model Builders. *Chem. Reviews*, 93:2567–2581.
- Sadowski, J., Gasteiger, J., & Klebe, G. (1994). Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J Chem Inf Comp Sci*, 34:1000–108.
- Sakmar, T. P. (2002). Structure of rhodopsin and the superfamily of seven helical receptors: the same and not the same. *Curr Opin Cell Biol*, 14:189–195.

- Sali, A. (1995a). Comparative protein modeling by satisfaction of spatial restraints. *Molecular Medicine Today*, 1:270–277.
- Sali, A. (1995b). Modelling mutations and homologous proteins. *Curr Opin Biotech*, 6:437–451.
- Sali, A. & Blundell, T. L. (1990). Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J Mol Biol*, 212(2):403–428.
- Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234:779–815.
- Sali, A., Shakhnovich, E. I., & Karplus, M. (1994a). How does a protein fold? *Nature*, 369:248–251.
- Sali, A., Shakhnovich, E. I., & Karplus, M. (1994b). Kinetics of Protein Folding. A Lattice Model Study of the Requirements for Folding to the Native State. *J Mol Biol*, 235:1614–1636.
- Samama, P., Cotecchia, S., Costa, T., & Lefkowitz, R. J. (1993). A mutation induced active state of the β_2 adrenergic receptor. Extending the ternary complex model. *J Biol Chem*, 268:4625–4636.
- Samudrala, R. & Moult, J. (1998). An All-Atom Distance-dependent Conditional Probability Discriminatory Function for Protein Structure Prediction. *J Mol Biol*, 275:895–916.
- Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marti-Renom, M. A., Madhusudhan, M. S., Mirkovis, N., & Sali, A. (2000). Protein Structure modeling for structural genomics. *Nature Struct Biol*, Seiten 986–990.
- Sanchez, R. & Sali, A. (1998). Advances in comparative protein-structure modelling. *Curr Opin Struct Biol*, 7:206–214.
- Schaffer, A. A., Wolf, Y. I., Ponting, C. P., Koonin, E. V., Aravind, L., & Altschul, S. F. (1999). IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. *Bioinformatics*, 15:1000–1011.
- Schafferhans, A. & Klebe, G. (2001). Docking ligands onto binding site representations derived from proteins built by homology modelling. *J Mol Biol*, 307(1):407–427.
- Schertler, G. F. & Hargrave, P. A. (1995). Projection structure of frog rhodopsin in two crystal forms. *Proc Natl Acad Sci*, 92:11578–11582.
- Schnecke, V. & Kuhn, L. A. (2000). Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Discov Des*, 20:171–190.
- Schnecke, V., Swanson, C. A., Getzoff, E. D., Tainer, J. A., & Kuhn, L. A. (1998). Screening a Peptidyl Database for Potential Ligands to Proteins With Side-Chain Flexibility. *Proteins*, 33: 74–87.
- Schonbrun, J., Wedemeyer, W. J., & Baker, D. (2002). Protein structure prediction in 2002. *Curr Opin Struct Biol*, 12(3):348–54.

- Schulz-Gasch, T. & Stahl, M. (2003). Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Mod*, 9(1):47–57.
- Scussa, F. (2002). World's best selling drugs. *Med Ad News*, 21(5):1–46.
- Selbie, L. A. & Hill, S. J. (1998). G protein-coupled-receptor cross-talk: the fine-tuning of multiple receptor-signalling pathways. *Trends Pharmacol Sci*, 19:87–93.
- Selector (1996). Tripos Inc., St. Louis MO 63144.
- Seward, E. M., Owen, S. N., Sabin, V., Swain, C. J., Cascieri, M. A., Sadowski, S., Strader, C., & Baker, R. (1993). Quinuclidine-Based NK-1 Antagonists I: 3-Benzoyloxy-1-Azabicyclo[2.2.2]octanes. *Bioorg Med Chem Lett*, 3:1361–1366.
- Shacham, S., Topf, M., Avisar, N., Glaser, F., Marantz, Y., Bar-Haim, S., Noiman, S., & Becker, O. M. (2001). Modeling the 3D Structure of GPCRs from Sequence. *Med Res Rev*, 21:472–483.
- Shih, N.-Y., Albanese, M., Anthes, J. C., Carruthers, N. I., Grice, C. A., Lin, L., Mangiaracina, P., Reichard, G. A., Schwerdt, J., Seidl, V., Wong, S.-C., & Piwinski, J. J. (2002). Synthesis and Structure-Activity Relationships of Oxime Neurokinin Antagonists: Discovery of Potent Arylamides. *Bioorg Med Chem Lett*, 12:141–145.
- Shim, J.-Y., Welsh, W. J., & Howlett, A. C. (2003). Homology Model of the CB_1 Cannabinoid Receptor: Sites Critical for Nonclassical Cannabinoid Agonist Interaction. *Biopolymers*, 71:169–189.
- Siegal, G., van Duynhofen, J., & Baldus, M. (1999). Biomolecular NMR: recent advances in liquids, solids and screening. *Curr Opin Chem Biol*, 3:530–536.
- Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C., & Baker, D. (1999). Improved Recognition of Native-Like Protein Structures Using a Combination of Sequence-Dependent and Sequence-Independent Features of Proteins. *Proteins*, 34:82–95.
- Sippl, M. J. (1993). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Comput Aided Mol Des*, 7: 473–501.
- Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr Opin Struct Biol*, 5(2):229–35.
- Sippl, W. (2003). Development of biologically active compounds by combining 3D QSAR and structure based design methods. *J Comput Aided Mol Des*, 16:825–830.
- Sisto, A., Bonelli, F., Centini, F., Fincham, C. I., Potier, E., Monteagudo, E., Lombardi, P., Arcamone, F., Goso, C., Manzini, S., Giolitti, A., Maggi, C. A., Venanzi, M., & Pispisa, B. (1995). Synthesis and biological evaluation of novel neurokinin-1 tachikinin receptor antagonists: the use of cycloalkyl amino acids as a template. *Biopolymers*, 36:511–524.
- Skolnick, J., Fetrow, J. S., & Kolinski, A. (2000). Structural genomics and its importance for gene function analysis. *Nat Biotechnol*, 18(3):283–7.

- Skolnick, J., Kolinsky, A., & Ortiz, A. R. (1997). MONSSTER: A Method for Folding Globular Proteins with a Small Number of Distance Restraints. *J Mol Biol*, 265:217–241.
- Snider, R. M., Constantine, J. W., Lowe, J. A. r., Longo, K. P., Lebel, W. S., Woody, H. A., Drozda, S. E., Desai, M. C., Vinick, F. J., & Spencer, R. W., e. a. (1991). A potent nonpeptide antagonist of the substance P (NK1) receptor. *Science*, 251:435–7.
- Sotriffer, C., Klebe, G., Stahl, M., & Böhm, H.-J. *Burgers Handbook of Medicinal Chemistry*, chapter Docking and Scoring Functions and Virtual Screening. Burger's Medicinal Chemistry, 6th Edition, 2002a.
- Sotriffer, C., Kr"amer, O., & Klebe, G. (2003). Protein flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *in preparation*.
- Sotriffer, C. A., Gohlke, H., & Klebe, G. (2002b). Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J Med Chem*, 45(10):1967–70.
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J Am Chem Soc*, 120:9401–9409.
- Srinivasan, S., March, C. J., & Sudarsanam, S. (1993). An automated method for modeling proteins on known templates using distance geometry. *Protein Sci*, 2:227–289.
- Stahl, M. & Rarey, M. (2001). Detailed Analysis of Scoring Functions for Virtual Screening. *J Med Chem*, 44:1035–1042.
- Stevenson, G. I., MacLeod, A., Huscroft, I., Cascieri, M. A., Sadowski, S., & Baker, R. (1995). 4,4-Disubstituted piperidines: A new class of NK-1 antagonists. *J Med Chem*, 38:1264–1266.
- Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. J. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*, 112:6127–6129.
- Stubbs, M. T., Reyda, S., Dullweber, F., Moller, M., Klebe, G., Dorsch, D., Mederski, W. W., & Wurziger, H. (2002). pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *Chembiochem*, 3(2-3):246–9.
- Swain, C. J., Seward, E. M., Cascieri, M. A., Fong, T. M., Herbert, R., MacIntyre, D. E., Merchant, K. J., Owen, S. N., Owens, A. D., Sabin, V., Teall, M., VanNiel, M. B., Williams, B. J., Sadowski, S., Strader, C., Ball, R. G., & Baker, R. (1995). Identification of a Series of 3-(Benzyloxy)-1-azabicyclo[2.2.2]octane Human NK1 Antagonists. *J Med Chem*, 38:4793–4805.
- Swain, C. J., Seward, E. M., & Sabin, V. (1993). Quinuclidine-based NK-1 antagonists 2: Determination of the absolute stereochemical requirements. *Bioorg Med Chem Lett*, 3:1703–1706.
- Takeuchi, Y., Shands, E. F. B., Beusen, D. D., & Marshall, G. R. (1998). Derivation of a Three-Dimensional Pharmacophore Model of Substance P Antagonists Bound to the Neurokinin-1 Receptor. *J Med Chem*, 41:3609–3623.

- Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov*, 2: 527–541.
- Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., & Stenkamp, R. E. (2001). Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, a Model of G-Protein-Coupled Receptors (GPCRs). *Biochemistry*, 40(26):7761–7772.
- Thibaut, U., Folkers, G., Klebe, G., Kubinyi, H., Merz, A., & Rognan, D. (1994). Recommendations for CoMFA Studies and 3D QSAR Publications. *Quant Struct-Act Relat*, 13:1–3.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22:4673–4680.
- Thornton, J. M., Orengo, C. A., Todd, A. E., & Pearl, F. M. G. (1999). Protein Folds, Functions and Evolution. *J Mol Biol*, 293:333–342.
- Thornton, J. M., Todd, A. E., Milburn, D., Borkakoti, N., & Orengo, C. A. (2000). From structure to function: Approaches and limitations. *Nature Struct Biol*, Seiten 991–994.
- Ting, P. C., Lee, J. F., Anthes, J. C., Shih, N.-Y., & Piwinski, J. J. (2000). Synthesis and NK_1/NK_2 Receptor Activity of Substituted-4(Z)-(methoxyimino)pentyl-1-piperazines. *Bioorg Med Chem Lett*, 10:2333–2335.
- Ting, P. C., Lee, J. F., Anthes, J. C., Shih, N.-Y., & Piwinski, J. J. (2001). Synthesis of Substituted 4(Z)-(Methoxyimino)pentyl-1-piperidines as Dual NK_1/NK_2 Inhibitors. *Bioorg Med Chem Lett*, 11:491–494.
- Ting, P. C., Lee, J. F., Shih, N.-Y., Piwinski, J. J., Anthes, J. C., Chapman, R. W., Rizzo, C. A., Hey, J. A., Ng, K., & Nomeir, A. A. (2002). Identification of a Novel 1'-[5-((3,5-Dichlorobenzoyl)methylamino)-3-(3,4-dichlorophenyl)-4-methoxyimino)pentyl]-2-oxo-(1,4'-bipiperidine) as a Dual NK_1/NK_2 Antagonist. *Bioorg Med Chem Lett*, 12:2125–2128.
- Tiraboschi, G., Jullian, N., Thery, V., Antonczak, S., Fournie-Zaluski, M. C., & Roques, B. P. (1999). A three-dimensional construction of the active site (region 507-749) of human neutral endopeptidase (EC.3.4.24.11). *Protein Eng*, 12(2):141–9.
- Tokarski, J. S. & Hopfinger, A. J. (1997). Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: application to a set of peptidomimetic renin inhibitors. *J Chem Inf Comput Sci*, 4:792–811.
- Totrov, M. & Abagyan, R. *Protein-ligand docking as an energy optimization problem*, chapter Drug-receptor thermodynamics: introduction and applications. Wiley, Chichester, 2001.
- Tramontano, A., Leplae, R., & Morea, V. (2001). Analysis and Assessment of Comparative Modeling Predictions in CASP4. *Proteins*, Suppl 5:22–38.

- Trosset, J. Y. & Scheraga, H. A. (1998). Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc Natl Acad Sci U S A*, 95 (14):8011–5.
- Tsai, J., Taylor, R., Chothia, C., & Gerstein, M. (1999). The Packing Density in Proteins: Standard Radii and Volumes. *J Mol Biol*, 290:253–266.
- Tsai, J. Gerstein, M. & Levitt, M. (1997). Simulating the minimum core for hydrophobic collapse in globular proteins. *Protein Sci*, 6:2606–2616.
- Unger, R., Harel, D., Wherland, W., & Sussman, J. L. (1989). A 3-D building blocks approach to analyzing and predicting structure of proteins. *Proteins*, 5:355–373.
- Unger, V. M. & Schertler, G. F. (1995). Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. *Biophys J*, 68:1776–1886.
- UNITY (2001). Chemical Information Software, 4.2.1 edit., Tripos Inc., St. Louis MO.
- Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., & Goddard, W. A., r. (2002). Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A*, 99(20):12622–7.
- Vakser, I. A. (1996). Long-distance potentials: an approach to the multiple-minima problem in ligand-receptor interaction. *Protein Eng*, 9(1):37–41.
- van Gunsteren, W. (1998). Validation of molecular dynamics simulation. *J Comput Phys*, 108:6109–6116.
- van Rhee, A. M. & Jacobson, K. A. (1996). Molecular Architecture of G protein-Coupled Receptors. *Drug Des Discov*, 37:1–38.
- Vasquez, M. (1996). Modeling side-chain conformation. *Curr Opin Struct Biol*, 6:217–221.
- Vaz, R. J., McLean, L. R., & Pelton, J. T. (1998). Evaluation of proposed modes of binding of (2S)-2-[4-[[[(3S)-1-acetimido-3-pyrrolidinyl]oxy]phenyl]-3-(7-amidino-2-naphthyl)propanoic acid hydrochloride and some analogs to Factor Xa using a comparative molecular field analysis. *J Comput Aided Mol Des*, 12:99–110.
- Vedani, A., Briem, H., Dobler, M., Dollinger, H., & McMasters, R. (2000). Multiple-Conformation and Protonation-State Representation in 4D-QSAR: The Neurokinin-1 Receptor System. *J Med Chem*, 43:4416–4427.
- Vedani, A., Zbinden, P., & Snyder, J. P. (1993). Pseudo-receptor modeling: a new concept for the three-dimensional construction of receptor binding sites. *J Recept Res*, 13(1-4):163–77.
- Venclovas, C., Zemla, A., Fidelis, K., & Moult, J. (2001). Comparison of Performance in Successive CASP Experiments. *Proteins*, Suppl. 5:163–170.

- Verkhivker, G., Appelt, K., Freer, S. T., & Villafranca, J. E. (1995). Empirical free energy calculations of protein-ligand crystallographic complexes: I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng*, 8:677–691.
- Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Luty, B. A., Marrone, T., & Rose, P. W. (2000). Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des*, 14:731–751.
- Wallner, B. & Elofsson, A. (2003). Can correct protein models be identified? *Protein Sci*, 12: 1073–1086.
- Wallquist, A., Jernigan, R. L., & Covell, D. G. (1995). A preference-based free energy parametrization of enzyme-inhibitor binding. Application to HIV-1-protease inhibitor design. *Protein Sci*, 4:1881–1903.
- Wang, R., Gao, Y., & Lai, L. (1998). SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *J Mol Model*, 4:379–394.
- Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des*, 16(2):379–394.
- Wang, R., Lu, Y., & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem*, 46(12):2287–2303.
- Wang, R. & Wang, S. (2001). How does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J Chem Inf Comput Sci*, 41:1422–1426.
- Wang, T. & Wade, R. C. (2001). Comparative binding energy analysis. *J Med Chem*, 44:961–971.
- Wang, Y., Zhang, H., Li, W., & Scott, R. A. (1995a). Discriminating compact nonnative structures from the native structure of globular proteins. *Proc Natl Acad Sci U S A*, 92(3):709–13.
- Wang, Y., Zhang, H., & Scott, R. A. (1995b). A new computational model for protein folding based on atomic solvation. *Protein Sci*, 4(7):1402–11.
- Ward, P., Armour, D. R., & Bays, D. E. (1995). Discovery of an oral bioavailable NK1 receptor antagonist, (2S,3S)-(2-methoxy5-tetrazol-1-ylbenzyl)(2-phenylpiperidin-3-yl)amine (GR-203040), with potent antimetic activity. *J Med Chem*, 38:4985–4992.
- Wei, L., Huang, E. S., & Altman, R. B. (1999). Are predicted structures good enough to preserve functional sites? *Structure*, 7:643–650.
- Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. (1986). An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem*, 7(2):230–252.
- Williams, B. J., Teall, M., & McKenna, J. (1994). Acyclic NK-1 antagonists: 2-Benzhydryl-2-aminoethylethers. *Bioorg Med Chem Lett*, 4:1903–1908.

- Wilmot, C. M. & Thornton, J. M. (1988). Analysis and Prediction of the Different Types of β -turn in Proteins. *J Mol Biol*, 203:221–232.
- Wilson, C. A., Kreychman, J., & Gerstein, M. (2000). Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. *J Mol Biol*, 297(1):233–249.
- Wold, S., Johansson, E., & Cocchi, M. *PLS - Partial Least-Squares Projections to Latent Structures*, chapter 3D QSAR in Drug Design. Theory, Methods and Applications. Volume 1, Seiten 523–550. Escom, Leiden, 1993.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. Wiley, New York.
- Xiang, Z. & Honig, B. (2001). Extending the Accuracy Limits of Prediction for Side-chain Conformations. *J Mol Biol*, 311(2):421–430.
- Yang, A. S. & Honig, B. (2000a). An integrated approach to the analysis and modelling of protein sequences and structures. II. On the relationship between sequence and structural similarity for proteins that are not obviously related in sequence. *J Mol Biol*, 301:679–689.
- Yang, A. S. & Honig, B. (2000b). An integrated approach to the analysis and modelling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J Mol Biol*, 301:691–711.
- Yang, L., Berk, S. C., Rohrer, S. P., Mosley, R. T., Guo, L., Underwood, D. J., Arison, B. H., Birzin, E. T., Hayes, E. C., Mitra, S. W., Parmar, R. M., Cheng, K., Wu, T. J., Butler, B. S., Foor, F., Pasternak, A., Pan, Y., Silva, M., Freidinger, R. M., Smith, R. G., Chapman, K., Schaeffer, J. M., & Patchett, A. A. (1998). Synthesis and biological activities of potent peptidomimetics selective for somatostatin receptor subtype 2. *Proc Natl Acad Sci U S A*, 95(18):10836–10841.
- Zar, J. H. (1999). *Biostatistical Analysis*, 4th Ed. Prentice Hall.
- Zhang, X. P., Sjoling, S., Tanudji, M., Somogyi, L., Andreu, D., Eriksson, L. E., Graslund, A., Whelan, J., & Glaser, E. (2001). Mutagenesis and computer modelling approach to study determinants for recognition of signal peptides by the mitochondrial processing peptidase. *Plant J*, 27(5):427–38.
- Zoffmann, S., Gether, U., & Schwartz, T. W. (1993). Conserved HisVI-17 of the NK-1 receptor is involved in binding of non-peptide antagonists but not substance P. *FEBS Lett*, 336(3):506–510.

Danksagung

- Herrn Prof. Dr. G. KLEBE danke ich herzlich für die Möglichkeit, diese außerordentlich interessante Arbeit in seinem Arbeitskreis durchführen zu können. Der gewährte Spielraum für die Entwicklung eigener, kreativer Ideen, die ständige Diskussionsbereitschaft sowie die Ermutigung und das entgegengebrachte Vertrauen für die Präsentationen von Teilen dieser Arbeit auf internationalen Tagungen haben wesentlich zum Gelingen dieser Arbeit beigetragen.
- Ein ganz besonderer Dank gilt meiner Zimmergenossin KATRIN SILBER für die stets sehr angenehme Unterhaltung und Unterstützung sowohl während als auch außerhalb der Arbeit an unseren Doktorarbeiten.
- Ich danke allen (z.T. ehemaligen) Post-Doktoranden (insbesondere Dr. CHRISTOPH SOTRIFFER und Dr. HOLGER GOHLKE), Doktoranden, Diplomanden und technischen Angestellten, die wertvolle Denkanstöße lieferten, Diskussionsbereitschaft gezeigt und eine angenehme Arbeitsatmosphäre geschaffen haben. Dr. ANDREA SCHAFFERHANS danke ich für die Einarbeitung in die Methoden dieser Arbeit. Ein weiterer Dank für ihre Hilfen in (Sybyl-)Softwarefragen gilt Dr. RUTH BRENK und KATRIN SILBER. Ein extra Dank noch einmal an DANIEL KUHN für das Transportieren der Promotionstasche!
- Für das Korrekturlesen von Teilen dieser Arbeit bzw. darin enthaltener Publikationen bedanke ich mich bei Dr. CHRISTOPH SOTRIFFER, Dr. HOLGER GOHLKE und SILVIA EVERS.
- Bei Dr. RUTH BRENK und Dr. STEFAN SCHMITT bedanke ich mich für die Überlassung der LaTeX-Vorlage.
- Ich danke meinem Literaturarbeiter BERND KRAUSS für seine geleistete Arbeit, deren Ergebnisse zum Teil in diese Arbeit eingeflossen sind.
- Dr. ARNE SVENSSON und AstraZeneca (Mølndal) danke ich für die Vermessung der Affinitätswerte der NK1 Antagonisten.

- Dr. MARKUS HAUSWALD (Bayer AG, Wuppertal) danke ich für die motivierende Einführung in die Welt des computergestützten Wirkstoffdesigns und die Idee, die vorliegende Arbeit im AK Klebe durchzuführen.
- A. Šali danke ich für die vorzeitige Bereitstellung von MODELLER6 sowie für die hilfreiche Korrespondenz zur Einführung der DrugScore Potentiale in MODELLER.
- Meiner Familie (inklusive Eltern und Schwiegereltern) und meinen Freunden danke ich für die Unterstützung, die als wesentlicher Bestandteil zur Realisierung dieser Arbeit beigetragen hat.

Aus der vorliegenden Arbeit sind folgende Posterbeiträge, Vorträge und Publikationen hervorgegangen:

Posterbeiträge:

- Evers, A., Klebe, G., *Modelling Protein Binding Sites. A new approach for modelling ligands into binding-sites of proteins with unknown 3D-structure.* Poster präsentiert bei der Model(l)ing 2001-Tagung, Erlangen (2001). (Das Poster wurde mit einem Posterpreis ausgezeichnet).
- Evers, A., Klebe, G., *Modelling Protein Binding Sites. A new approach for modelling ligands into binding-sites of proteins with unknown 3D-structure.* Poster präsentiert bei der GPCR-Tagung, Wiesbaden (2001).
- Evers, A., Klebe, G., *Modelling Protein Binding Sites. A new approach for refining homology models by (Q)SAR.* Poster präsentiert beim 14th European Symposium on Quantitative Structure-Activity Relationships, Bournemouth / Großbritannien, (2002).
- Hillebrecht, A., Evers, A., Klebe, G., *Optimising Lead Structures by Subpocket-Guided Design.* Poster präsentiert bei der Jahrestagung der GDCh, Abteilung "Medizinische Chemie", Fulda (2003).
- Evers, A., Klebe, G., *Discovery of a potent antagonist for a GPCR by Virtual Screening on a homology model.* Poster präsentiert bei der Jahrestagung der GDCh, Abteilung "Medizinische Chemie", Fulda (2003).

Vorträge:

- Evers, A., Klebe, G., *Modelling protein binding-sites.* Vortrag präsentiert beim 16. Darmstädter Molecular-Modelling Workshop, Darmstadt (2002).
- Evers, A., Klebe, G., *Modelling protein binding-sites.* Vortrag präsentiert beim 14th European Symposium on Quantitative Structure-Activity Relationships, Bournemouth / Großbritannien, (2002).
- Evers, A., Klebe, G., *Modelling protein binding-sites.* Vortrag präsentiert beim International Workshop "New Approaches in Drug Design & Discovery", Rauischholzhausen, (2003).

Aufsätze:

- Grimm, C., Evers, A., Brock, M., Maerker, C., Klebe, G., Buckel, W., Reuter, K., *Crystal Structure of 2Methylisocitrate Lyase (PrpB) from Escherichia coli and Modelling of its Ligand Bound Active Centre. J Mol Biol*, 328 (2003), 609-621.
- Evers, A., Gohlke, H., Klebe, G., *Ligand-supported Homology Modelling of Protein Binding Sites using Knowledge-based potentials. J Mol Biol*, im Druck.
- Evers, A., Klebe, G., *Modelling protein binding-sites: A new approach for refining homology models by (Q)SAR. Proceedings of the 14th European Symposium on Quantitative Structure-Activity Relationships (Euro QSAR 2002), Bournemouth, September 8-13*, im Druck.
- Evers, A., Klebe, G., *Ligand-supported homology modelling of G-Protein Coupled Receptors: Models sufficient for successful Virtual Screening.* angenommen bei *Angewandte Chemie*.
- Evers, A., Klebe, G., *Successful Virtual screening for submicromolar antagonists of the neurokinin-1 receptor based on a ligand-supported homology model.* Manuskript in Vorbereitung.

Die gemodelte Struktur des Komplexes der 2Methylisocitrat Lyase (PrpB) mit dem Substrat ist mit dem Code 1MZX in der PDB-Datenbank abgelegt.

Erklärung

Ich versichere, dass ich meine Dissertation

*"A New Method for Ligand-supported Homology Modelling of Protein Binding Sites:
Development and Application to the neurokinin-1 receptor"*

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch in keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 29. September 2003

(Andreas Evers)

Lebenslauf

Andreas Evers

Geburtstag:	14. August 1973
Geburtsort:	Köln
1980 – 1981	<i>Grundschule</i> in Köln
1981 – 1984	<i>Grundschule</i> in Bergisch Gladbach
1984 – 1993	<i>Albertus-Magnus Gymnasium</i> , Bensberg
Jun. 1993	Abitur
Aug. 1993 – Okt. 1994	Zivildienst am <i>Kuratorium für Heimdialyse</i> , Köln-Merheim
1994 – 1998	Chemiestudium an der <i>Universität zu Köln</i> <i>Vordiplomprüfungen</i> Feb. 1997 <i>Diplomprüfungen</i> Dez. 1998 Schwerpunkt: <i>Biochemie</i>
Jan. 1999 – Aug. 1999	<i>Diplomarbeit</i> am Institut für Biochemie, <i>Universität zu Köln</i> , im Arbeitskreis Prof. D. Schomburg <i>Dichteberechnungen an Proteinen</i>
Aug. 1998 – Dez. 1999	<i>Industriepraktikum</i> in der Molecular Modelling Gruppe der Strukturforschung der Bayer AG in Wuppertal
Jan. 2000 – Mai 2000	<i>Auslandsaufenthalt</i> an der John Curtin School of Medical Research (JCSMR), Australien, in der "Computational Proteomics and Therapy Design group" von Dr. J. Gready
Jun. 2000	Beginn der Arbeiten zur vorliegenden Dissertation unter Betreuung von Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie der <i>Philipps-Universität Marburg</i>
Jun. 2000 – Sep. 2003	Wissenschaftlicher Angestellter am Institut für Pharmazeutische Chemie der Philipps-Universität Marburg; Betreuung des Studentenpraktikums im ersten Semester