QCSPScore: a new scoring function for driving protein-ligand docking with quantitative chemical shifts perturbations

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Johann Wolfgang Goethe – Universität
in Frankfurt am Main

von
Domingo González Ruiz
aus Benejúzar

Frankfurt, 2009
(D 30)

vom Fachbereich Biowissenschaften der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Volker Müller

Erster Gutachter: Prof. Dr. Holger Gohlke

Zweiter Gutachter: Prof. Dr. Gisbert Schneider

Datum der Disputation: 29.01.2010

A mis padres.

Whether you can observe a thing or not depends on the theory which you use.

It is the theory which decides what can be observed.

Albert Einstein,
objecting to the placing of observables
at the heart of the new quantum mechanics,
during Heisenberg's 1926 lecture in Berlin.

Acknowledgements

This work would have never been possible without the help and support from many people, which at this point I would like to thank.

I thank Prof. Holger Gohlke for having given me the opportunity to be the first PhD student in his group. Thank your for trusting me and having given me enough freedom to try things, even if many times they did not make much sense. This Thesis would have never be possible without your support and guidance, full of enthusiasm for good Science. It has been a terrific experience!

I am grateful to Prof. H. Schwalbe and Dr. U. Schieborr who kindly provided me experimental CSP for the three protein-ligand complexes studied in this work. I also thank Dr. U. Schieborr for fruitful discussions.

Very special thanks to Sebastian Radestock for proof-reading the Thesis, translating the Summary into German and his readiness to help me literally from the first to the last minute. Thank you for your constant support, for your encouragement, for making my life easier through the German bureaucracy and for your friendship. It has been a privilege meeting you.

Thank you to my group colleagues, from whom I learned a lot and with whom it was a pleasure working. Thanks to Christopher, Sina and Hannes, for their quick answers regarding the status of the computing queue. Many thanks to my office mates, Simone, Christopher and Sebastian, for creating a great working atmosphere, for all the fun we had in between and for their friendship.

A Beatriz de Pascual-Teresa, quien desde mis estudios de Farmacia en Madrid ya pensó que yo debería escribir una tesis doctoral. Gracias por, desde entonces, haber seguido apoyándome, asesorándome y preocupándote por mi desarrollo profesional y personal.

Thank you to the very good friends from MPI Dortmund: Gemma, Steffen, René, Sammy, Nancy, Anouk, Jacqui and Christian. Thank you for your constant encouragement and friendship after leaving Frankfurt, so that this Thesis could be finished one day.

iv

Gracias a mis buenos amigos que durante estos años han creído que sí sería capaz de terminar la tesis y, además, me han apoyado para que lo hiciese: a Mario, a Alfonso, a Johanna, a Carlos y a Steffanie, a Beatriz, a Andreas, a Ines, a Marta, a Carmen, a Claire, a Michaela, a Iciar, a José Miguel, a Pilar, a José Francisco, a Ralf, a Kathrin, a Obdulia, a Mariano y a Andrea.

A Diego y a Sergio. Habéis sido mi familia en Frankfurt, así que deciros "gracias" quizá sea poco.

Por último, gracias a mi familia. A mi hermana y a mis padres. Por su constante apoyo y cariño y por ser la red sobre la que se puede saltar sin miedo a caerse. Por haberme dado los dos regalos más importes: la vida, y la libertad para vivirla.

Table of contents

Abbreviations

DPF	Docking parameter file
3D	Three dimensional
AD	AutoDock
AIR	Ambiguous interaction restraints
BMRB	Biological magnetic resonance bank
CS	Chemical shift
CSP	Chemical shift perturbation
DS	DrugScore
Eq	Equation
HN	Amide proton
ID	Identification
LGA	Lamarckian genetic algorithm
NMR	Nuclear magnetic resonance
PDB	Protein data bank
QCSP	Quantitative chemical shift perturbation
RMSD	Root mean squared deviation
SBLD	Structure-based ligand design
VS	Virtual screening

1 Thesis motivation and objectives

Protein-ligand docking is a widely used tool in basic and applied pharmaceutical research. The appearance of the first three-dimensional structures of proteins immediately sparked the interest in developing methodologies that theoretically predict the binding mode and the interaction energy of putative ligands of them. Since the beginning of the 80s (Kuntz, et al., 1982), docking remains a field of intense research (Kitchen, et al., 2004; Cavasotto and Orry, 2007). This has already crystallized in a significant number of success stories in both Virtual Screening (VS) applications, aimed at identifying new binders to a target (Kitchen, et al., 2004; Cavasotto and Singh, 2008), and in ligand optimization efforts, the most popular among them being the commercialized drug imatinib (Gleevec) (Capdeville, et al., 2002). Yet docking is not perfect and there is still room and need for improvement.

Docking comprises both generating good protein-ligand geometries and properly estimating of the interaction energy of the binding partners. For the first task, reported success rates reach 80% of the cases (Warren, et al., 2006). The second, more challenging, translates in variable success that range from 10 to 70% in discriminating native complex structures from decoys and almost inability to correctly rank-order series of ligands according to their binding potency (Warren, et al., 2006). These figures have led to the commonly held view that in docking, the geometry problem has been resolved to a sufficient extent and the scoring problem remains the open question. However, Velec et al. have recently shown that both problems are so intimately related, that good geometries are indeed a pre-requisite for improving scoring functions (Velec, et al., 2005). Along these lines, the authors argued that the commonly accepted geometric accuracy limit of 2 Å root-mean-squared-deviation (RMSD) to the native crystal structure is far too large for subsequently performing a reliable scoring, and thus focus must be still kept in generating better geometries. An analogous observation has been made in an experimental study of a series of matrix metalloproteinases inhibitors, where "correct" docking solutions were not "good enough" for explaining experimentally observed differences in affinity (Bertini, et al., 2007).

The need for improvement in docking (Leach, et al., 2006a) is augmented from recent developments in two fields. First, advances in molecular and systems biology have lead

to the identification of more and more biochemical pathways and signaling routes that explain the physiology of the cell and the organisms. Specifically, this has crystallized in the emerging therapeutic class of protein-protein interfaces (PPI). The structural characterization of these PPI, key for rational exploitation, represents a challenge for both, experimental (NMR and X-ray crystallography) and theoretical techniques (Pagliaro, et al., 2004; Russ and Lampel, 2005; Gonzalez-Ruiz and Gohlke, 2006; Imming, et al., 2006; Weigelt, et al., 2008). With respect to the latter, it has been argued, that success of current docking approaches has been largely facilitated by the steric constraints imposed by well-defined deep cavities that exist in enzymatic targets (Ferrara, et al., 2004), which contrasts sharply with the observation that PPI are typically flat and devoid of deep binding sites for small molecules (Pettit and Bowie, 1999) as exemplified in Figure 1.1.
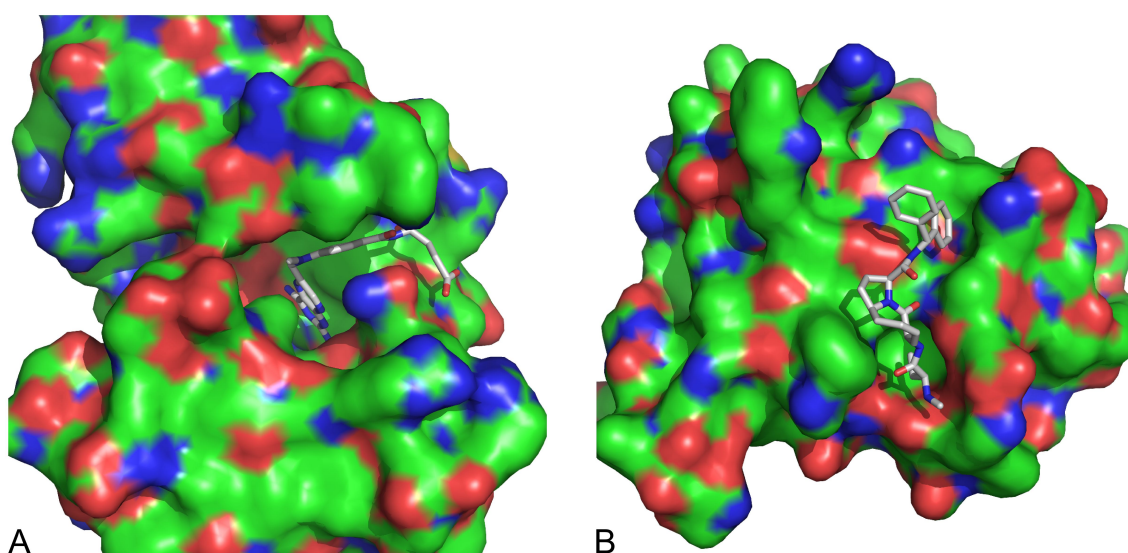
Figure 1.1 Comparison of the "deep" and well-defined binding pocket of a typical enzymatic target (A) (Dihidrofolate reductase complexed with methotrexate; PDB ID: 1df7) to the "rather flat" interacting surface involved in a protein-protein interaction (B) (X-linked inhibitor of apoptosis protein complexed with a Smac mimetic; PDB ID: 2jk7).

The second field longing for substantial improvements in protein-ligand docking is fragment-based VS. Fragment-based techniques are a rather young approach, but which has already gained deserved attention from the medicinal chemistry community (Hajduk and Greer, 2007; Congreve, et al., 2008). However, there is an open debate whether current docking tools are appropriate for dealing with fragments, due to the comparatively larger binding site regions given the size of the ligand and the supposed

"over-training" of current tools for more "drug-like" ligands (Hubbard, et al., 2007; Marcou and Rognan, 2007; Chen and Shoichet, 2009).

Together with the continuing effort to improve docking by better treatment of target flexibility and prediction of water molecules mediating in the interaction, there is an attempt at improving the physical description of the interaction by considering scoring functions with more sophisticated entropic and electrostatic treatments. Meanwhile, while progress in this "general docking" approach happens slowly, practical alternatives solutions for specific problems start to emerge. Among such alternatives, probably the most interesting are "tailored scoring functions". Here, the idea is to harness and integrate readily available information for a given target either from a collection of known binders (Fradera and Mestres, 2004; Jansen and Martin, 2004; Radestock, et al., 2005), or from directly measured experimental data for the complex under study (van Dijk, et al., 2005a) to guide/restrain docking. This present work falls in the second category.

The synergy between directly measured experimental data and theoretical docking methods made possible to study biological systems that were considered intractable or very challenging before. Remarkably, they have given a boost to the challenging field of protein-protein docking (van Dijk, et al., 2005a). Not surprisingly, now efforts are also being put in the protein-ligand docking field, specially in connection with NMR data (Carlomagno, 2005; Powers, 2007). There are many NMR observables that provide information about the structure of a biomolecular system. Chemical shifts (CS) are the most fundamental ones, and key for assigning NMR spectra, but not considered optimum for deriving structure (Szilagyi, 1995). On the one hand, the CS of a nucleus in a magnetic field reflects the chemical environment surrounding the nucleus. On the other hand, the CS results from a combination of numerous and complex effects. This means that different chemical environments can lead to similar CS which makes structure elucidation based on CS alone a difficult or unaccomplishable task. But the situation has changed recently. More data about CS in proteins have accumulated in public data bases (Ulrich, et al., 2008) thus improving our understanding and modeling accuracy. Together with the increasing availability of computer power, CS find increasingly more applicability in structural characterization of biomolecules (Wishart

and Case, 2001; Hunter, et al., 2005). I will focus on their use in for the study of protein-ligand complexes.

If the CS of a nucleus is an expression of its chemical environment, a change in that environment produced by the binding of a ligand to a protein results, consequently, in a chemical shift perturbation (CSP). CSP of isolated atoms in a protein do not provide much structural information, but taken all CSP measured on the protein side upon ligand binding it is possible to precisely define binding site regions. Such simple idea is the principle of SAR by NMR, a successful fragment-based ligand design approach established in the last decade (Shuker, et al., 1996). The qualitative approach (perturbation observed or not), nicely circumvents the aforementioned difficulties to interpret and model CSP. But at the same time, it does not provide information about the relative orientation of the ligand to the protein, a task which can be deferred to a standard docking program. The logical step forward for better exploiting structural information contained in CSP was to consider them semi-quantitatively or fully quantitatively. Among the first approaches, which consider the absolute magnitude, but not its sign (i.e. whether the CSP reflects an upfield or a downfield) are the one by Schieborr et al. (Schieborr, et al., 2005) and the one by Stark and Powers (Stark and Powers, 2008). I will discuss, that such approaches still fail to directly orient the ligand in the binding site. They keep the ligand in the binding site, but if this is large or flat enough or the ligand has internal symmetries, the way CSP information is incorporated does not contribute to properly placing it. The fully quantitative approach, by its side, can exploit the ring current effect as a unique feature to properly orient the ligand. Ring current effects leave a pattern of down- and upfields in their vicinity, depending on relative ring orientation and distance. This has been already acknowledge and prospectively explored by McCoy and Wyss (McCoy and Wyss, 2000) and later by Cioffi et al. (Cioffi, et al., 2008) In both cases CSP were, however, used for post-filtering or optimizing docking generated solutions, with the drawback that the docking algorithm itself is, in some cases, unable to generate a good solution at all. As stated above, this is a likely situation for PPI and in fragment screening attempts. My aim has, therefore been to incorporating CSP information directly into a docking simulation, by means of a new hybrid scoring function (QCSPScore) that actively drives solutions towards native-like poses.

This leads to the following specific goals:

- Evaluate different models for the prediction of CSP upon protein-ligand complex formation.

- Develop an objective function to measure the agreement between observed and back-calculated CSP, taking into account experimental and theoretical accuracy limits and efficiency considerations.

- Combine this objective function with DrugScore, resulting in a hybrid scoring scheme directly applicable to standard protein-ligand docking with AutoDock. Establish a weighting factor for both contributions in the hybrid scoring scheme.

- Extend AutoDock by implementing QCSPScore.

- Test and validate the QCSP-driven approach, comparing results and performance to standard docking and similar approaches recently published.

2 Structure-based ligand design

The discussion about aiding ligand design with information from structural biology began more than 40 years ago as the first three-dimensional (3D) structures of globins, enzymes and polypeptides were published (Beddell, et al., 1976). However, it took still 20 more years until the first success stories following the new paradigm of "structure-based ligand design" (SBLD) were reported (Erickson, et al., 1990; Roberts, et al., 1990). SBLD builds on the observation that small molecules (from now on, the ligands) can modulate the activity of bio-macromolecules (from now on the target, normally a protein) through binding to specific regions of the latter. Nowadays potent binders are rationally designed by optimizing structural and chemical complementarity of the ligand to the target. Selective binders, for those cases where several targets have similar binding sites, can be also designed if the structural nuances of the involved structures are spotted and exploited alternatively to prevent and/or enhance the interactions, respectively (Beddell, 1992; Perutz, 1992).

The value of detailed information about protein structures is widely recognized not only in SBLD applications, but also in more basic biological research areas. For example, in the study of the proteome, a newly discovered protein can be attributed a function if a structurally similar one is already known, following the paradigm "structure explains function". The success of such approach has fuelled ambition to obtain more high-resolution structures in a faster way. Policies have been adopted worldwide and a number of consortia have been established to promote research in improving the throughput (and eventually reducing the high costs) of the traditional experimental methods for solving protein structures, i.e., X-ray crystallography and NMR (Stevens, et al., 2001). The outcome of such initiatives has not only resulted in an acceleration of the number of structures deposited in public databases in the last years (Figure 2.1), but also more interestingly in the development of a substantial number of experimental and computational methods and variations available for producing structural information in a cost-effective way (Blundell, et al., 2002). SBLD benefits directly from these developments in structural genomics, because even if the particular protein-ligand complex structure of interest is not available, another structure of the target (free or bound to a different ligand) can be used for predicting the binding of a ligand of interest (Weigelt, et al., 2008).
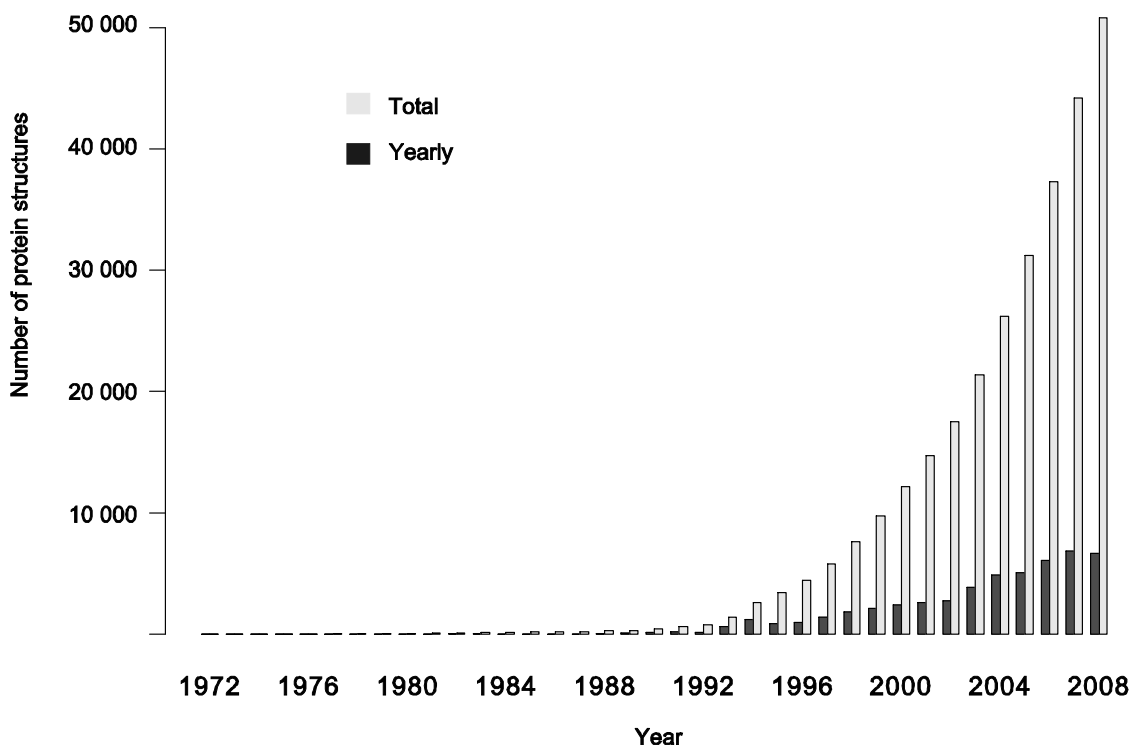
Figure 2.1 Number of biomolecular structures released from the protein databank (PDB) per year. Source: http://www.rcsb.org/

SBLD is nowadays an integral component of ligand design efforts in general. A recent review by Scapin illustrates this fact by recounting that 25-30% of the articles in a typical issue of the *Journal of Medicinal Chemistry* report a structure-based approach to inhibitor discovery and/or development (Scapin, 2006). But perhaps more compelling is the fact the increasing number of marketed drugs that have originated from SBLD, like for example, aliskiren (Rasilez) (Jensen, et al., 2008), oseltamivir (Tamiflu) (von Itzstein, 2007) and imatinib (Gleevec) (Capdeville, et al., 2002). More examples can be found in ref. (Congreve, et al., 2005). One can find SBLD in all stages of the drug design process (Figure 2.2), and research groups try to incorporate this valuable piece of information as soon as possible to every project when feasible. This provides a major strategic change, as the technique was almost exclusively used for ligand optimization.

In ligand optimization, one tries to leverage the information provided by a known protein-ligand complex. By studying such structure, one tries to find unexploited binding features. For example, should there be a lipophilic sub-pocket or a potential hydrogen-bond donor or acceptor group in the binding site not yet exploited by the known ligand this would be a clear opportunity to consider for the next generation of

molecules to be synthesized. This simple but powerful idea has opened an avenue to the development of so called "structure-based compound library design" methodologies for a specific target. Focused libraries substantially reduce the number of compounds that need to be experimentally tested in search of potent and selective binders, thus saving resources (Orry, et al., 2006).

Another application of SBLD has been to understand drug resistances as a consequence of mutations (Wong and Witte, 2004). It is possible to obtain protein-ligand complexes of an inhibitor with wild-type proteins, for which the inhibitor was active, and mutated ones. The comparison of the binding modes can reveal what are the specific protein-ligand interactions responsible for the loss of binding affinity. A new generation of binders should be designed by targeting conserved interactions only, which are less likely to mutate.

Finally, progress in high-throughput NMR and X-ray crystallography, along with increasingly available computer power and new algorithmic developments promoted SBLD to the front of the ligand design pipeline: hit finding, which applies to those cases where no known binders exist or totally new compound classes are sought. Soaking protein crystals in cocktails of chemical fragments (Hartshorn, et al., 2005), multiple NMR methodologies for screening (Coles, et al., 2003), and computer approaches such as *de novo* design approaches (Schneider and Fechner, 2005) and structure-based virtual screening techniques (Shoichet, 2004) are all recent and sophisticated developments that illustrate the willingness to harnessing structural information for aiding in the design of new ligands. Including structural information is certainly rewarding: in a recent review on fragment-based ligand design, Hajduk and Greer reported that "the ability to produce potent inhibitors (IC_{50} < 100 nM) after initiating lead optimization on fragment leads nearly triples with the aid of structure-based design, increasing from 33% to 93%" (Hajduk and Greer, 2007).
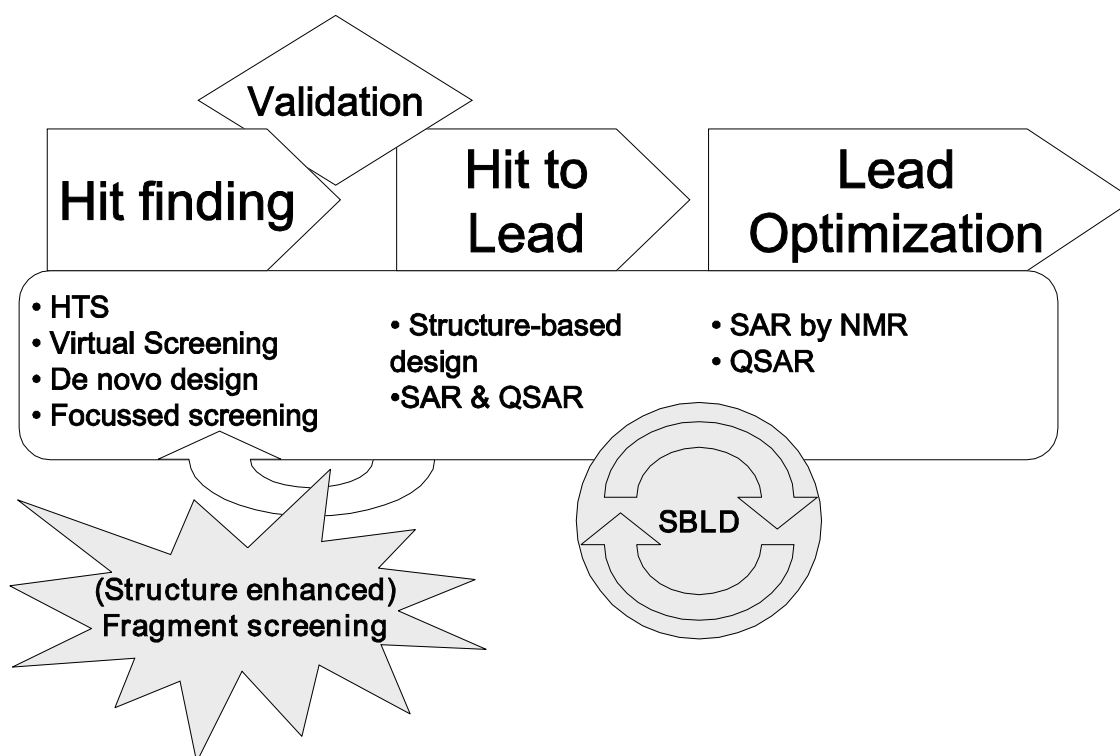
Figure 2.2 Schematic representation of the ligand design process. Structural information of the target has been traditionally used at the final lead optimization phase; nowadays, thanks to developments in X-ray crystallography, NMR, and high-throughput computational docking, structural information of protein-ligand complexes plays a role as early as in hit finding. Bringing rational design to the front of the pipeline has translated in increased success rates in obtaining potent and selective ligands.

Although knowing the free structure of the target is advantageous, the full power of SBLD comes from the study of complex structures. This is particularly true for those cases where, for example, the binding site is only revealed by a conformational change of the protein to accommodate the bound ligand (Arkin, et al., 2003). Binding of a ligand to its target protein results from a very delicate balance between attractive and repulsive forces, and the bound conformation is not straightforwardly inferable by just observing both unbound structures (Gohlke and Klebe, 2002). In spite of that, the amount of publicly available data concerning "free" structures outweighs by large that of complexes. In addition, if one analyzes the available complex structures, an immediate realization is that many targets are "overrepresented". The "Binding MOAD" data base (Hu, et al., 2005), a curated subset of the PDB focused on protein-ligand complexes, contains a total of ~10000 protein-ligand complexes relating to ~1800 non-redundant targets (http://www.bindingmoad.org/). This underlines once

more, first, how difficult it is to obtain protein-ligand complex structures and, second, how target-dependently current X-ray and NMR methodologies perform.

Particularly challenging for current techniques are proteins involved in molecular signaling through direct interaction with other proteins, the modulation of which is seen as a great opportunity to expand therapeutic possibilities (Arkin and Wells, 2004; Gonzalez-Ruiz and Gohlke, 2006). All this translates in an urgent need for methodological developments that can expand the applicability and the throughput of techniques aimed at solving protein-ligand complexes. Developing one of such methods is the focus of this thesis. Below an overview of the state-of-the-art and current trends and developments in the field is given, in order to properly frame the niche and relevance of the approach herein developed.

2.1 Protein-ligand complex structure solved by experimental methods

Despite rapid and impressive progress in other techniques such as cryo-electron microscopy (Jiang, et al., 2008), X-ray crystallography and NMR remain reference experimental techniques for characterizing macromolecules and their complexes at atomic resolution (Figure 2.3).

In X-ray crystallography, crystals of proteins are irradiated with an X-ray beam, which due to the different electron densities it encounters in its way, scatters after crossing the crystal. By rotating the crystal, multiple scatters from different perspectives can be collected and assembled so that a 3D picture of the internal structure of the crystal is inferred. The aminoacidic sequence of the protein must be known, so that the structure of the protein can be elucidated by fitting the protein atoms to the template generated from the observed electron densities (Ilari and Savino, 2008). X-ray crystallography is commonly acknowledged as a superior technique for obtaining high-quality structures, however, finding the optimal conditions for crystallization is typically challenging. By its side, NMR is a more versatile technique (Jahnke, 2007), since it can produce a larger number of observables from which to infer the structure of the macromolecule, among which J-couplings, NOEs, chemical shifts, relaxation rates, residual dipolar couplings are the most relevant (Clore and Gronenborn, 1998). The clear advantage is that finding

the right experimental conditions is considerably less demanding in NMR (Wider, 2000). However, two critical limitations of NMR are first, the requirement of relatively high sample concentrations at which the protein must not aggregate. And second, due to the high number of signals, traditional NMR is limited to the study of proteins smaller than 10 kDa; larger ones (30-40 kDa) require different isotope labeling schemes to simplify the spectra.

But despite the aforementioned current limitations, NMR for structural characterization of biomolecules is still a young technique, subject to intense research in its different components. In order to simplify the crowded spectra that typically big molecules produce, selective labeling strategies have been developed to alternatively switch on or switch off specific nuclei (Goto and Kay, 2000; Wider, 2005). A better physical understanding of nuclear spin dynamics has permitted the development of new pulse sequences, increasing the control on the generated NMR signal. A remarkable breakthrough in this sense has been the advent of TROSY spectroscopy (transverse relaxation-optimized spectroscopy (Fernandez and Wider, 2003)) which has significantly surpassed the protein size limitations (Fiaux, et al., 2002). Also, experimental developments have permitted the practical exploitation of residual dipolar couplings, which are useful for the study of relative orientations of large domains of macromolecules (Bax and Grzesiek, 1993). Additionally, NMR not only has profited from experimental advances, but also from algorithmic developments: for example, in the tedious stage of sequential assignment of the observed signals (Guntert, 2003) or at the final stage of structural characterization, where the sometimes limited experimental data can be supplemented with refinement procedures based on modern force-fields and advanced sampling techniques (Chen, et al., 2004).

This picture renders both techniques complementary rather than competing: both yield high quality structures, but each one observes different aspects of the structures. Most importantly, different experimental conditions requirements leave only one of the two techniques applicable, as will be described below.
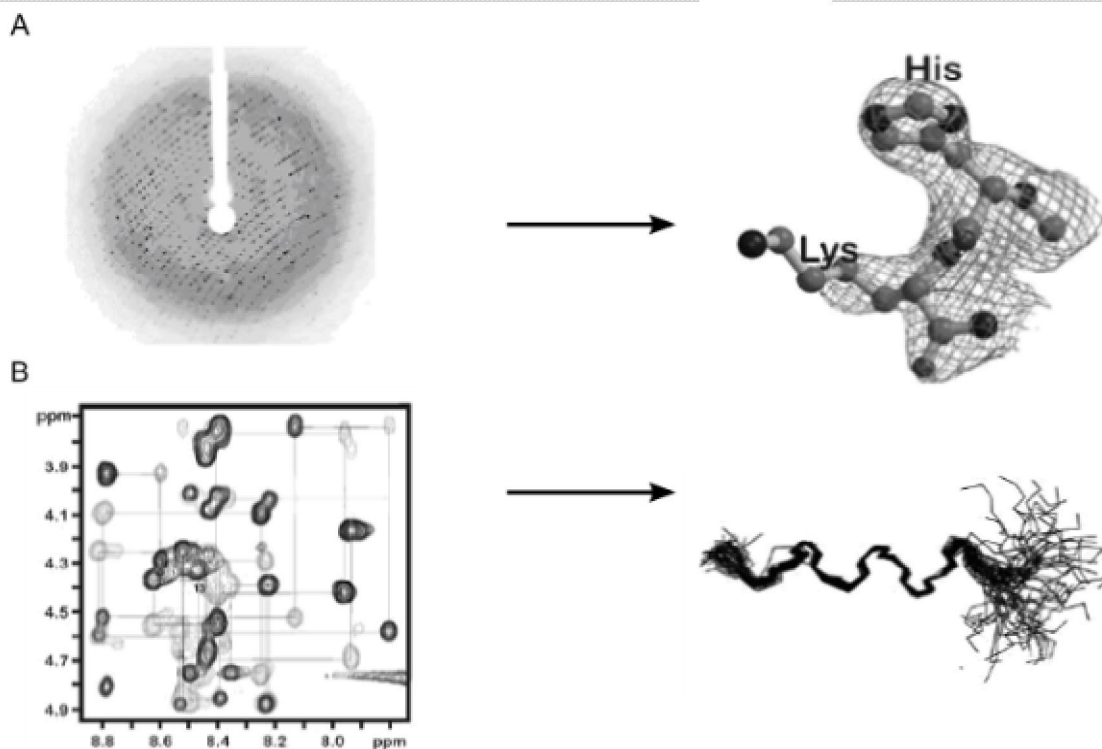
Figure 2.3 X-ray crystallography (A) and NMR (B) are the main tools for solving biomolecular structures. X-ray methods fit the structure of the molecule to the observed diffraction pattern. NMR relies on a large collection of observed NOE, which subsequently is used with distance-geometry methods.

2.1.1 Protein-ligand X-ray crystallography

X-ray crystallography is the predominant and preferred method for solving structures of proteins and protein-ligand complexes. Approximately 85% of the structures deposited in the PDB at the time of writing have been solved by this technique (Berman, et al., 2000). X-ray crystallography can be applied to macromolecules in a large range of sizes, providing high structural resolution and even the position of ordered water molecules that in many occasions mediate protein-ligand interactions (Scapin, 2006). Traditionally, protein-ligand complexes have been solved by co-crystallization, which typically yields a good-quality picture of the complex, but requires producing robust crystals with good diffracting properties (Leach, et al., 2006b). This is the most challenging step of the whole process, particularly if the experimental conditions to obtain the protein-ligand crystals differ from the ones in which the crystal of the protein alone crystallizes, or the ligand is not soluble in the crystallization medium (Jhoti, et al., 2007). In summary, a tedious, not always successful, series of trial-and-error attempts

results for each new protein-ligand system to be studied. Challenged by these limitations, the worldwide structural genomics community has pushed research towards the development of robotics and liquid handling procedures, which eventually permitted an increased throughput in the crystal production (Kuhn, et al., 2002; Sharff and Jhoti, 2003; Weigelt, et al., 2008) by parallel testing of large numbers of conditions. The synergy of these technological advances with a technique called "soaking" constitutes the big success in bringing protein-ligand crystallography to a high-throughput level.

In the soaking technique compounds (as single species or in mixtures) are incubated with pre-formed crystals of the protein, free or in complex with another ligand to be displaced by the binding of the new one. This permits that, once the crystallization conditions for a given protein have been optimized, large compound libraries can be tested in parallel, in principle. High-throughput crystallography by soaking has been key in the development and adoption of X-ray aided fragment-based techniques for hit finding and ligand design. This is a very young approach to ligand design but already counts with a remarkable number of successes reported in the literature (Jhoti, et al., 2007).

One limitation though, is that the crystal form must be compatible with the binding of the ligand, i.e., the binding of the ligand should not break the crystal. Also, the binding site must be accessible for the ligand. As a further concern, it has been argued that in some cases the binding mode of the soaked ligands does not accurately correspond to the native binding mode, as the constraints of the pre-formed crystal prevent it (Zhu, et al., 1999; Hiller, et al., 2006).

Apart from the above described experimental limitations, X-ray crystallography in general is not free from being affected by artifacts. Crystal preparation, data collection, or coordinate placement during refinement can affect and condition the final results (Shoichet and Bussiere, 2000; Davis, et al., 2003; Acharya and Lloyd, 2005). If the system under study happens to be crystallization-resistant, NMR or theoretical methods are appealing alternatives for structural investigations.

2.1.2 Traditional (NOE-based) NMR for studying protein-ligand complexes

As discussed in the introduction to this section, the study of protein-ligand complexes by NMR methods is nowadays seen as complementary to X-ray, rather than competing. "Traditional NMR" for structural characterization of biomolecules relies on the geometric information that can be derived from through-space proton-proton correlations as measured via the nuclear Overhauser effects (NOE) (Wider, 1998) The NOE is the phenomenon by which spin polarization is transmitted from one spin population to another via cross-relaxation (Levitt, 2001). It has been established that for molecules with a molecular weight of more than 5 kDa, the intensity of a NOE between two protons is proportional to the inverse of the sixth power of the distance separating them (Neuhaus and Williamson, 2000). Accordingly, NOEs are usually observed between protons which are not further than ~5 Å from one another. Observing and assigning an adequate number of NOEs with their relative intensities, suffices to generate highly determining number of distance restraints as to accurately define the structure of the protein (Wuthrich, 2001). In the case of protein-ligand complexes, despite the examples reported (Gargaro, et al., 1998; Moy, et al., 1999; Polshakov, et al., 1999a; Moy, et al., 2000), two limitations appear: first, observing and assigning NOEs in protein-ligand complexes does not provide the throughput required in drug-design efforts; second, intermolecular protein-ligand NOE are not observed for ligands that bind weakly to the protein (Reibarkh, et al., 2006; Cavanagh, et al., 2007).

The limitation of the throughput yield early attempts to simplify the standard method (Polshakov, et al., 1999b), but the real improvement in this respect comes when the structure of the target protein (unbound, or in complex with other ligands) is known. For example, in the NMR-DOC protocol (Pellecchia, et al., 2002), the idea is to apply the knowledge of the aminoacidic content of the binding site region to selectively label it for the NMR experiment. Assignment of residues in the binding site is subsequently is achieved by monitoring chemical shifts perturbations due to ligand binding or by directly measuring observing protein-ligand NOEs. Another approach is based on generating a large number of candidate ligand poses, then back-calculating the theoretical intermolecular NOE they would produce on protein amide groups (simple distance-dependence) and compare these with the experimentally observed ones

(Bertini, et al., 2005). A further refinement of this idea overcomes the necessity of assigning all observed NOEs (Constantine, et al., 2006). From the generated ligand poses the corresponding theoretical 3D ^{13}C-edited, $^{13}C/^{15}N$-filtered HSQC-NOESY spectrum is calculated and by using a fast, deterministic bipartite graph matching algorithm the prediction is compared and scored with respect to the experimental spectrum.

With respect to complexes in the weak binding regime, the most interesting in lead finding and optimization stages, instead of NOE, attention has gone to other NMR observables, including chemical shifts, line broadening, transferred NOE, intermolecular magnetization transfer or change in relaxation properties (Carlomagno, 2005) These observables play the most important role in the study of protein-ligand complexes by NMR (Jahnke, 2007). However, since structures cannot be inferred as straightforwardly from them as in the case of complete NOE collections, they must be substantially supplemented with theoretical methods that automatically generate plausible poses. Most typically, such a conformation generator is a *docking* program. Since the approach described in this Thesis falls in exactly that category, it is convenient to first present what docking is and what current docking approaches are capable of. It is from that perspective that this work has been done. Consequently, NMR-based methods for structural studies of protein-ligand complexes binding in the weak regime are reviewed afterwards, under the appellative of "hybrid NMR-driven docking approaches".

2.2 Theoretical methods for predicting protein-ligand complex structures: docking and scoring

2.2.1 Definition of protein-ligand docking

While solving the structure of a protein-ligand complex fully by experimental method is the preferred way, it comes together with high costs in time and resources. This is particularly true when a relatively large number of ligands needs being studied. In addition, it is not always possible to for example obtain good diffracting crystals of a protein-ligand complex. However, if the structure of the target protein (or a very close homologue) is known, solving the structure of a protein-ligand complex using

computational methods is nowadays a widely accepted alternative and, in the case of large libraries, definitely a more efficient one. A very popular technique for such a task is the "docking and scoring" methodology. Given the structure of the protein and the putative ligand, the docking program samples millions of likely interacting modes between the two partners and scores them according to a defined objective function aimed at predicting the interaction energy. Eventually, the best scored pose corresponds to the best theoretical binding energy and, if the simulation went right, to the native structure of the complex in biological conditions. Reported docking programs amount to more than 60, combined with more than 30 scoring functions, according to a very recent review (Moitessier, et al., 2008). On average, they achieve up to 80% success rate in the task of placing a ligand back into its binding site (re-docking) when dealing with typical enzymatic targets, i.e., proteins with deep and well-defined binding sites. Such a rate deteriorates sharply in the case or flatter binding sites (Ferrara, et al., 2004; Warren, et al., 2006).

Despite imperfections, docking has been applied with remarkable success in the past in the context of structure-based virtual screening (SBVS) approaches. There, large virtual compound libraries are docked to a given target, and the energy of the interaction is estimated. According to the interaction energies, compounds can be prioritized, thus orienting and concentrating the efforts of medicinal chemists on the most interesting compounds for actual synthesis and biochemical/biological testing. Precious time an resources are saved, this way, as illustrated, for example in a SBVS campaign in search of protein tyrosine phosphatase 1B. There, the prioritized set of ligands resulted in a 1700-fold enhancement of the hit rate over random screening (Doman, et al., 2002). Some other success docking stories have been collected in a number of reviews (Kitchen, et al., 2004; Mohan, et al., 2005; Klebe, 2006). Together with hit-finding, docking has another relevant application in the ligand optimization step. Once an interesting binder is found, efforts concentrate on proposing new one with larger potency or better selectivity profile or simply easier chemistry. By analyzing the interaction predicted by docking of proposed virtual compounds, informed decisions can be made with respect to which compounds must be synthesized first. It can become apparent that some candidates do not fit in the binding site or that others do not fulfill required specific interactions (Choong, et al., 2002).

A docking experiment requires four elements: the structure of the protein, the structure of the ligand, an automatic generator of conformations and an evaluator of those conformations (Figure 2.2).
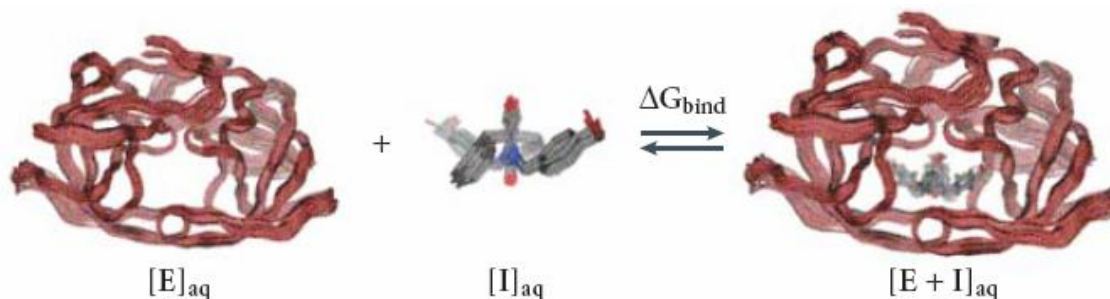
Figure 2.4 Docking consists of predicting how a ligand binds to its target protein and the energy of such interaction. Here, the binding of inhibitor Dmp323 to HIV-1 protease is illustrated (PDB code: 1BVE). Adapted from ref (Kitchen, et al., 2004)

The first docking programs considered both the protein and the ligand as rigid bodies (Kuntz, et al., 1982), which permitted a complete systematic search of ligand orientations with respect to the protein, or even manual docking if one had sufficient experimental data supporting the proposed solution. Now, ligands are always considered flexible and even some implementations allow introducing flexibility considerations of the protein. For example, version 4 of AutoDock (Morris, et al., 2009) as well as GOLD (Verdonk, et al., 2003) allow for explicit side chain flexibility. Other typical implementation of flexibility involves considering ensembles of protein conformations, experimental or theoretically simulated, as is the case of DOCK 3.5.54 (Wei, et al., 2004). For a recent review on dealing with target flexibility see reference (Cozzini, et al., 2008).

Considering both interacting partners as flexible generates a combinatorial explosion of possible mutual orientations that renders the exhaustive search as impossible. These methods include stochastic search as Montecarlo sampling or simulated annealing and heuristics such as genetic algorithms or swarm-based optimization algorithms. The optimization typically occurs on the hypersurface that the guiding objective function describes: the energy landscape. To every relative orientation, corresponds a binding energy and the task of the algorithm is to find the global minimum. The aforementioned stochastic methods have proven useful in surmounting likely energetic barriers that surround the global minimum (Morris, et al., 1998).

Objective/scoring functions are simplified attempts to efficiently evaluate the interaction energy of a given protein-ligand configuration. Rigorous energetic considerations of the interaction such as free-energy simulation techniques (Simonson, et al., 2002) are not practical for the purposes of docking. This results in docking solutions which are plausible, in terms of steric and chemical complementarity, but not native-like since important effects such as desolvation or translational and rotational entropy are neglected. And here is precisely where the subtle difference between docking and scoring emerges. The difficulties in properly modeling the binding energy of interaction is reflected in the discrepancy between the relatively high success rates in docking and the limited ability of docking programs to correctly rank-order series of inhibitors according to their potency (Warren, et al., 2006). But both problems are not disconnected, and it is accepted that better docked structures provide better results in terms of binding energy prediction (Velec, et al., 2005; Bertini, et al., 2007). Consequently it is worth revisiting the idea that the docking problem is *solved* and that most efforts must be put in improving scoring: even small further improvements in docking, will translate in a benefit to the scoring problem.

2.2.2 Scoring functions for protein-ligand docking

Scoring functions fall typically into three categories: force-field-based, empirical and knowledge-based.

In a *force-field based function*, the interaction energy between the ligand and the protein is decomposed as a sum of physics-based terms such as van der Waals and electrostatic contributions, at a molecular mechanical level, that is, simulating atoms as single particles that interact pairwise.

$$E_{vdW} + E_{ekectrostatic} = \sum_{protein} \sum_{ligand} \left[\left(\frac{A_{ij}}{d_{ij}^6} + \frac{B_{ij}}{d_{ij}^{12}} \right) + 332.0 \frac{q_i q_j}{\varepsilon(d_{ij}) d_{ij}} \right] \quad \text{Eq. (1)}$$

Where *Aij* and *Bij* are van der Waals parameters for the combination of atom type of *i* and atom type of *j*, atoms respectively from the protein and the ligand. *d* is the interatomic distance and $\varepsilon(d_{ij})$ is the distance-dependent dielectric function.

Variations of such scoring functions are implemented in popular docking packages such as AutoDock (Morris, et al., 1998), ICM (Abagyan, et al., 1994) and DOCK (Meng, et al., 1992). The parameters describing each atom type are typically obtained from fitting to experimental quantum mechanics data. Unfortunately, these parameters were originally formulated to model the systems in gas-phase, neglecting solvation and entropic effects. Finally only one protein conformation is considered, so that the score of a given ligand pose boils down to considering the internal energy of the ligand (energy of the conformation) and the intermolecular energy of interaction, neglecting the internal energy of the protein. All these simplifications, although rendering a quite efficient and to some extent accurate description of the protein-ligand interaction, translates into a limited applicability. The challenge to describe intermolecular interactions efficiently only with physics-based terms has been underlined by one simple estimation: the free energy difference between the best ligand that one might reasonably expect to identify using virtual screening (potency, ~50nM) and the experimental detection limit (potency, ~100 μM) is only about 4.5 kcal/mol. The free energy contribution due to conformational factors alone for typical druglike ligands can be as large as this (Tirado-Rives and Jorgensen, 2006). In spite of that, this remains an active field of research (Huang and Jacobson, 2007), and has found interesting applicability not for docking itself but in re-scoring schemes which attempt to more finely understand protein-ligand interactions, once the structure of the complex is in hand (Wang, et al., 2001).

Empirical scoring functions, pioneered by Böhm (Bohm, 1994), consist also on a weighted sum of terms, but these are not formally physically grounded.

$$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \\ + \Delta G_{hydrophobic} \sum_{hydrophobic} \left| A_{hydrophobic} \right| + \Delta G_{rotor} N_{rotor} + \Delta G_0$$

Eq. (2)

Binding energy (ΔG_{bind}) in this case is the sum of a hydrogen-bond (ionic and neutral) term, which has an angular ($\Delta\alpha$) and a distance dependence (ΔR), a hydrophobic effect term, as a function of the molecular surface area ($A_{hydrophobic}$) and an account for ligand rotational entropy, by means of considering the number of rotatable bonds in the ligand

(N_{rotor}). All terms are weighted ($\Delta G_{H\text{-}bond}$, ΔG_{ionic}, $\Delta G_{hdrophobic}$, ΔG_{rotor}) empirically, from regression analysis against experimental data. ΔG_0 is a regression constant.

These functions typically appear in popular docking packages such as GLIDE (Friesner, et al., 2004) and GOLD (Verdonk, et al., 2003). Although empirical scoring functions usually appear almost consistently amongst the most accurate ones at scoring, one should not forget that their accuracy is compromised by the experimental data used in the parameterization and that the training set of protein-ligand complexes is biased towards enzyme-inhibitor complexes where the ligand fits into a well-defined cavity. As discussed below, this indeed represent a serious drawback when dealing with small fragments or rather flat and large binding surfaces, as in protein-protein interfaces.

The third class of protein-ligand scoring functions are the so called *"knowledge-based" scoring functions*. They build up on the classical statistical physics idea that the observed distributions of properties can be used to infer the potential that gave rise to the distribution. Their first uses have been described in protein folding studies (Miyazawa and Jernigan, 1985), and the growing number of protein-ligand complexes deposited in public databases permitted the development of also protein-ligand binding potentials that are nowadays used in docking, being the most representative examples PMF (Muegge, 2000) and DrugScore (Gohlke, et al., 2000). The DrugScore equation is:

$$\Delta W = \gamma \sum_{protein} \sum_{ligand} \Delta W_{ij}(r) + (1-\gamma) \left[\sum_{ligand} \Delta W_i(SAS, SAS_0) + \sum_{protein} \Delta W_j(SAS, SAS_0) \right]$$

Eq. (3)

where W_{ij} is a distance dependent pairwise potential, *SAS* correspond to the surface accessible area terms and γ is an adjustable weight factor.

The main caveat for knowledge-based scoring function is that they have been developed to reproduce experimental structures, rather than binding energies, which somehow limits their applicability to that task. As empirical scoring functions, knowledge-based scoring functions circumvent the major limitations of force-field scoring functions, that is having to accurately describe the underlying physical phenomenon.

2.2.3 Current challenges and trends in protein-ligand docking

In common docking benchmarking and evaluation, programs are presented with three already anticipated challenges of increasing difficulty: (1) binding mode prediction, (2) discrimination of binders from non-binders, and (3) binding energy prediction. As discussed above, the degree of success of state-of-the-art docking programs in each of these tasks deteriorates according to the respective difficulties. Furthermore, correctly predicting binding energy (3rd goal) heavily relies on having good starting structures (1st goal), that is, all three goals are interconnected, and deficiencies are carried on along the line. In a recent comprehensive evaluation of docking programs, it has been shown that for the first goal, success can range from 0% to 90% of the cases, depending on the program/protein target combination. With respect to the second goal, results vary from recovering 80% of active compounds in a SBVS campaign to consistent worse than random performance, which raises the question for the source of success when it happens. The last objective is "beyond all of the current docking methods" (Leach, et al., 2006a; Warren, et al., 2006). With this picture in sight the most sensible decision is probably taking docking for its merits as a tool for generating plausible protein-ligand complex structures. Later, those poses are considered as starting points for further refinement with more sophisticated techniques, depending on the stage and needs of a given project. Of course, this is assuming that a 0% to 90% success in generating native-like structures is a convincing figure. It is certainly not for the cases in the low end, which appear to be challenging to all standard docking approaches.

Why docking fails, when it does? Noteworthy, the difficulty of the "challenging cases" has been attributed to characteristics of the target binding site. In particular, a rather large binding site surface, different from the typical enzymatic pockets on which most of the current docking methods have been trained (Warren, et al., 2006), seems to condition the success. In other words, success of current scoring functions appears to be highly dictated by well defined binding pockets, which effectively constrain the possibilities to place a given ligand inside. This reflects the limitations of current scoring functions and leaves docking out from what are nowadays probably the single two most promising developments in drug design: 1) modulation of protein-protein interfaces as valid therapeutic target (Arkin and Wells, 2004; Gonzalez-Ruiz and Gohlke, 2006) and 2) fragment screening (Hajduk and Greer, 2007).

As already advanced in the introduction, protein-protein interfaces are typically large and flat, compared to standard enzymatic binding pockets (Pettit and Bowie, 1999) (see Figure 1.1). By their side, fragments are in general too small for a binding site, thus providing few steric restraints for unambiguously docking them (Klebe, 2006) (Figure 2.5). This thesis is focused on improving the ability of docking programs to generate native-like solutions, in particular in sight of these two particularly challenging cases.
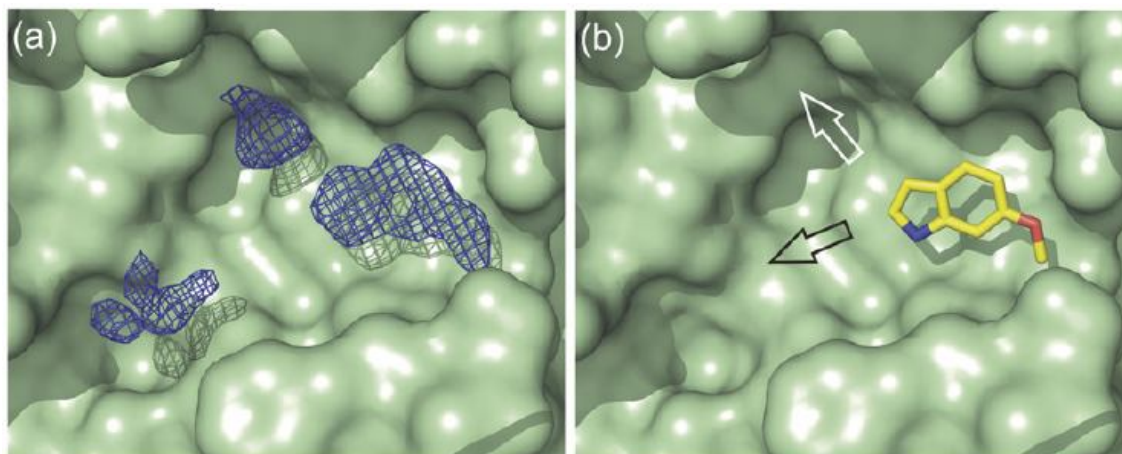
Figure 2.5. (a) Electron density showing multiple fragment-binding in an enzyme active site. (b) Core fragment chosen for synthetic elaboration showing possible growth vectors to two adjacent binding sites. Clearly, the fragment is not sterically constrained in the binding site, which for a docking program supposes a significant challenge. Adapted from (Ciulli and Abell, 2007)

The difficulty to develop scoring functions that perform equally well across many different protein families, regardless of their complexity and sophistication, is nowadays widely accepted (Kitchen, et al., 2004). For this reason, researchers tend to wisely choose the tool that works best for the system under study. If none is accepted as such, a suggested alternative has been to use several programs, pool all solutions together, and re-score by "consensus". A higher degree of refinement involves effectively tailoring the docking scoring function to the protein or protein family of interest. This can be done for example with ligand-based information. If a collection of known binders is in hand, a binding-model can be derived to bias the scoring function (Fradera and Mestres, 2004; Jansen and Martin, 2004; Radestock, et al., 2005). The peril of such approaches is that the new ligand is somehow forced to bind in a known fashion or to match known interactions, which in many cases is an acceptable assumption, but not if, for example, the aim was finding new scaffolds. Finally, if no collection of ligands is available, an appealing alternative is to incorporate directly measured experimental information from

the complex under study to restrain, guide, or postfilter docking solutions. A wealth source of such information is NMR. NMR data and docking programs make an excellent partnership, provided that the experimental observations can be efficiently and accurately modeled. Such has been the strategy pursued in this study: to combine NMR chemical shift perturbations with docking to improve the success of the latter and expand its applicability. Approaches sharing the same principles are reviewed in the next section.

2.3 Hybrid NMR-supplemented docking approaches

In a recent perspective on NMR in drug discovery published by *Nature Reviews in Drug Discovery*, experts underlined the fact that NMR is not fully exploited in industrial settings as a complement to X-ray crystallography, given that it is not rapid. Interestingly, their opinion on a possible way to improve the situation was expressed as follows: "Our view is that the combination of molecular docking supported by limited NMR experimental constraints could represent an efficient way to rapidly gather information on ligand-target complexes without full structure determination" (Pellecchia, et al., 2008). This opinion is solidly supported by the significant number of improvements and success reports that we are witnessing in the field. Nowadays, NMR can be applied in almost any stage of the ligand design process, from ligand binding detection (screening) (Coles, et al., 2003) and quantification of binding affinities (Klein, et al., 1999) to the most informative structural resolution at atomic level (Takeuchi and Wagner, 2006). See scheme in Figure 2.6.
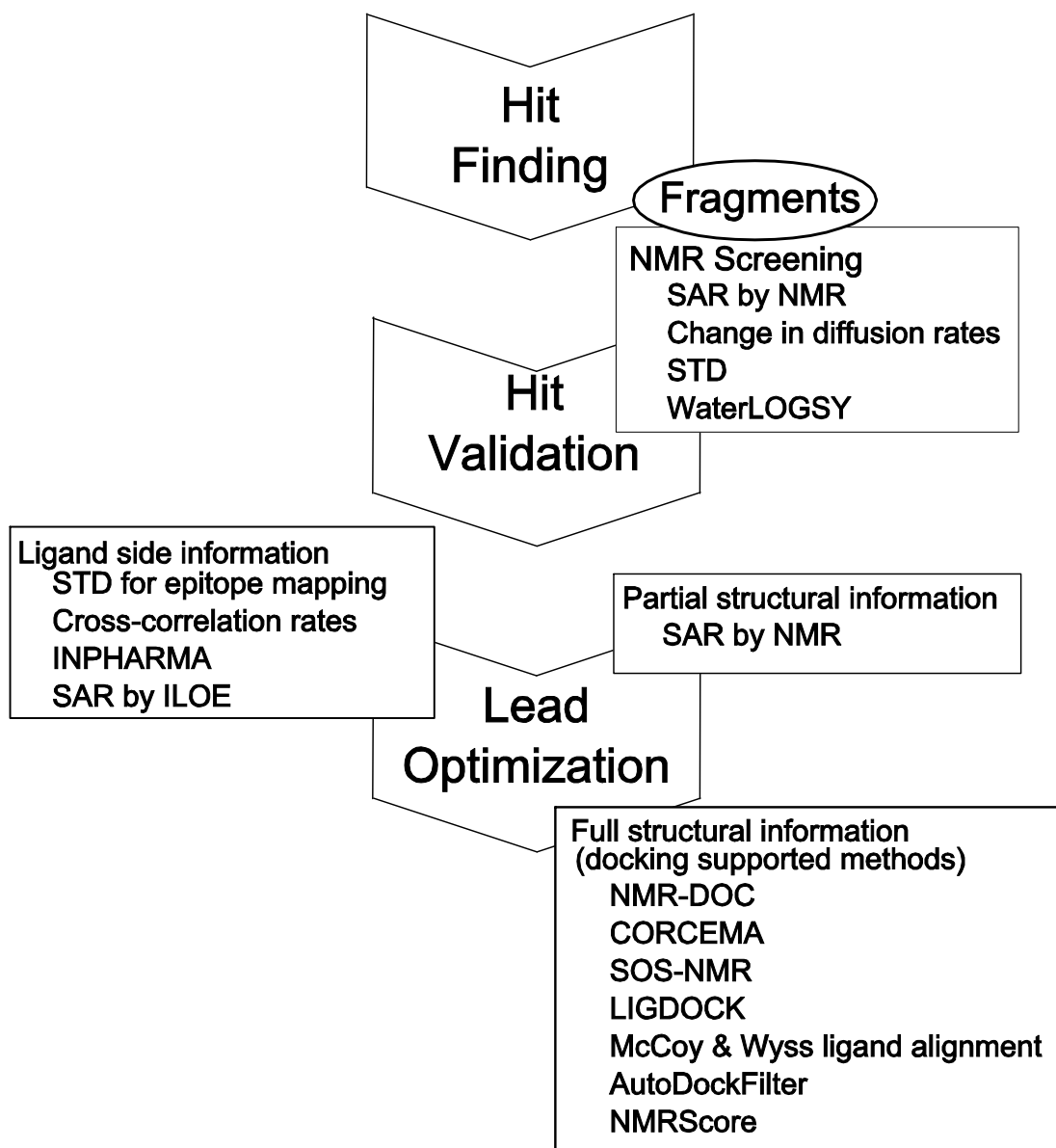
Figure 2.6 Role of NMR in the drug design process, once the target has been validated.

There are many NMR observables change upon binding and consequently are markers of the protein-ligand interaction. For example signals from the ligands, which do not have limitations on the size of the system studied nor need labeled samples. At the most basic level, changes in diffusion and relaxation rates can be used as binding reporters, as these properties are different for small ligands tumbling free compared to being bond to a large macromolecule (Hajduk, et al., 1997; Lin, et al., 1997) A more comprehensive list is collected in Table 2.1. In principle, these signals are not as rich in terms of structural information as those from the protein-side, but they constitute the basis for many of NMR-based screening techniques currently available (Meyer and Peters, 2003).

Table 2.1 Comparison of NMR methods to obtaining structural clues about protein-ligand complexes

	Observable	**Main limitation**	**Methods**	**Provides**
Protein / ligand side	NOE	- Tight binders - Complex size limit - NOE observation	- "Standard" NMR -NMR-DOC	Complex structure
Ligand side	Intermolecular magnetization transfer	- Fast exchange regime	- Transferred NOE - STD - WaterLOGSY - SAR by ILOE - CORCEMA	Ligand binding epitope
	Cross-correlated relaxation rate	- Fast exchange regime	- CCR rate	Complex structure Conformation of the ligand
	Chemical Shifts	- Vast computational resources	NMRScore	Complex structure
Protein side	Chemical shift perturbations	- Complex size limited - Assignments required	- SAR by NMR - LIGDOCK - McCoy & Wyss' approach -NMRScore	Binding site location / Complex structure

Noteworthy, NMR-based screening compares favorably to biochemical screening, since no specific assay needs being developed and knowledge of the function of the protein is not needed. Even though NMR-screening is not the main focus of this work, it happens to be the case that some of the observables exploited there, have lately found application in structural characterization, when combined with computational tools (Meyer and Peters, 2003; Carlomagno, 2005; Takeuchi and Wagner, 2006). In particular, those based in the detection of intermolecular transfer of magnetization (transferred NOE) and chemical shifts perturbations.

2.3.1 Uses of transferred intermolecular magnetization for structural characterization of protein-ligand complexes

Molecules with low molecular weight have short correlation time and, consequently, show NOE values that range from very small negative to positive. Large molecules instead show strong negative NOEs. However, when the low-molecular-weight ligand binds to its target protein, it behaves as part of it and adopts the corresponding NOE behavior, thus showing strong negative NOEs. This phenomenon is called the transferred NOE (trNOE) and relies therefore on the differences in tumbling times between free and bound small molecules. In principle it is possible to observe both

inter- and intramolecular trNOEs: the former ones are useful for defining the conformation of the ligand in the bound state and the latter for determining its orientation in the binding site (Meyer and Peters, 2003).

Magnetization transfer is the basis of STD (saturation transfer difference) and can be observed between target and ligand in short-lived complexes through the NOE (Meyer and Peters, 2003). In these experiments, appropriate pulses are used to selectively saturate signals from the target, so that ligand spins in close-contact with the target experience the transference of magnetization from those saturated spins. This is consequently reflected in a decrease of their line intensities in a 1D-spectrum, as depicted in Figure 2.7. Qualitatively, the technique has an unquestionable value for ligand screening. Additionally, some applications have been described where it has been shown that mixtures of ligands could be use likewise (Mayer and Meyer, 1999).

Figure 2.7 Schematic representation of the STD experiment. Ligand spins in close contact with the target receive part of the receptor spin saturation. Thus, ligand resonances that experience a decrease in intensity, following saturation of the target spins, identify the binding epitope of the ligand. Additionally, if the target has been selectively labelled, it is possible to establish distance restraints between specific ligand protons and residue types in the protein, which is the basis of the SOS-NMR approach (Hajduk, et al., 2004).

The quantitative use of STD, that is, considering which ligand's protons are more affected by the binding, reveals the binding epitope: that is the region of the ligand that is in closer contact to the protein (Maaheimo, et al., 2000; Haselhorst, et al., 2001). This information can be quite precious when the ligand is rather large, as in the case of polysaccharides or peptides. One can define distance restrains between specific atoms in the ligand and the surface of the binding site to restrain a docking simulation. Alternatively, those restrains can be employed in filtering out generated ligand poses

that violate them. Intramolecular ligand NOE, possibly combined with cross-correlation rates, serve also in the case of large ligands to fix it conformation (Carlomagno, 2005). Subsequently, rigid-body docking can be performed, which is a considerably less demanding task than fully flexible one.

STD, combined with specific labeling schemes, resulted in the SOS-NMR approach (Hajduk, et al., 2004). There, different samples of the target protein are prepared, each perdeuterated in a specific residue. The magnetization transfer from each residue type to the ligand protons is revealed by the trNOE, which can be easily assigned to a given residue type. If the number of investigated residue types is enough, one can, first, define the binding site and, second, impose restraints between specific protons in the ligand and certain residue-types in the protein. This restrains are finally used for post-filtering docking-generated ligand poses to unambiguously define the structure of the complex.

The most sophisticated use of STD data is presented by the CORCEMA (complete relaxation and conformational exchange matrix) approach (Jayalakshmi and Krishna, 2004) STD data is used here to refine the structure of the complex. In brief, the method is capable of predicting the STD of a protein-ligand complex. Differences between experimentally measured intensities and predicted ones for the proposed complexes are minimized during the search of the structure. An additional advantage claimed by the authors is that protein flexibility could be easily incorporated into the refinement procedure. The development was done with theoretical data and concerns were raised, that experimental data would introduce more noise and thus additional degrees of freedom which ultimately would hamper the convergence to a single structure (Carlomagno, 2005). However, successful applications have already been reported in the literature (Jayalakshmi, et al., 2004; Wen, et al., 2005).

If two ligands bind concomitantly to a protein in near-by regions, transference of magnetization can also be observed between them directly, or mediated through target protons. This is called the inter-ligand Overhauser effect (ILOE) (London, 1999). This information cannot only be used for detecting binding, but also to establish structure-activity-relationships, since both ligands could be linked together in a new molecule, thus producing a more potent binder. The technique aiming at such end has been called "SAR by ILOE" (Becattini and Pellecchia, 2006). A quantitative interpretation of the ILOE can be also used to derive distance restrains between both ligands on the receptor.

Time-sequential protein-mediated ILOEs between competitive ligands that bind in the same binding site are exploited in the INPHARMA (interligand NOE for phamacophore mapping) approach. By analyzing the ligand areas that happen to be in contact with the binding site, it is possible to derive a pharmacophoric model (Sanchez-Pedregal, et al., 2005), which in turn can be used to tailor a scoring function within a docking program.

2.3.2 Methods using CSP for characterizing protein-ligand complexes

Transferred magnetization discussed above can be applied to deduce structure provided a sufficient collection of observations from which to derive restraints. Thus they can be described as indicators of pairwise relative positions. By their side, chemical shifts perturbations (CSP) are very fine indicators of changes in the chemical environment of a nucleus as a whole, and therefore they are a more "genuine" parameter to be related to structure, despite difficulties to interpret them (Szilagyi, 1995).

 The most popular application of CSP in SBLD is probably the SAR by NMR approach (Shuker, et al., 1996). Here, small molecular fragments that bind to the protein produce different patterns of CSP on this protein, depending on the place where they bind. This information is used to map the binding site: fragment A binds in subpocket S1, fragment B, chemically different, in subpocket S2 and so on. From here, fragments can be rationally combined through linkers into new compounds, which will be better binders by virtue of simultaneously targeting at least two favorable neighbor binding regions. Unfortunately, protein nuclei not directly involved in the binding do also experience CSP due to typical protein re-arrangements or flexibility, which introduces some noise in the detection. This, indeed, supposes a serious restriction to SAR by NMR: proteins must not undergo significant rearrangement and assignments must be sufficiently comprehensive. A possibility to overcome this limitation would be monitoring differences in CSP for closely related ligands, expected to produce similar re-arrangements (Medek, et al., 2000). What the SAR by NMR lacks is the ability to provide specific structural information about the complex.

Probably the most prominent example where CSP are used in complex structure elucidation is one coming from the more challenging protein-protein docking field.

Protein-protein docking is particularly challenging because the number of possible relative orientations between interacting counterparts is several orders of magnitude larger than in the case of protein-ligand docking. For this reason pre-filtering or a bias integrated into the scoring function are the best options to obtain good results (and not postfiltering). In version 2.0 (May 2007), the HADDOCK program (Dominguez, et al., 2003) is able to incorporate various experimental data from NMR (chemical shifts, residual dipolar couplings (van Dijk, et al., 2005c), diffusion anisotropy (van Dijk, et al., 2006)) or from mutagenesis experiments. The approach has qualified successively amongst the most successful ones in the CAPRI competitions that evaluate the state of the art of protein-protein docking techniques (van Dijk, et al., 2005b). Technically, the restraints used by HADDOCK are ambiguous by definition. That is: it is known that a given residue must be in the interface (because their CS appear perturbed) but the specific counterpart is not known. These qualitative observations are modelled by means of ambiguous interaction restraints (AIR) (Nilges and O'Donoghue, 1998). An AIR is defined as an upper-bounded intermolecular distance that must be fulfilled upon complex formation. However, it does not require a particular residue pair to fulfill it, but a subset of pre-selected possible pairs. This definition effectively restricts the search of possible orientations to a more tractable number, increasing the chances of success.

$$AIR_{i(A),B} = \left(\sum_{m=1}^{N_{atoms},i(A)} \sum_{k=1}^{N_{res},B} \sum_{n=1}^{N_{atoms},k(B)} \frac{1}{d^{6}_{m,n}} \right)^{-1/6} \qquad \text{Eq. (4)}$$

Residues defined as "involved in the interaction" are taken as pairs (i, k), one from each counterpart A and B, respectively. The distance is computed for every atom m in residue i from the first protein to every atom n of residue k in the second protein. In this way, as soon as two atoms are in contact the restraint is satisfied.

Coming back to the protein-ligand field, inspired by the ideas developed in HADDOCK, the group of Schwalbe has developed LIGDOCK (Schieborr, et al., 2005), which exploits CSP in a both ambiguous (Nilges and O'Donoghue, 1998) and semi-quantitative fashion (weak, medium, strong, with only strong CSP being used) as restraints for docking, in combination with a force field. The assumption is that largest CSP must correspond to direct protein-ligand contacts, and this way, it is possible to define a scoring function that accounts for this situation. LIGDOCK relies on the

concept of AIRs as HADDOCK does, but in this case AIRs are defined unidirectionally from the ligand to the protein: every observed (large) CSP constitutes an AIR. Calculations of the structure are done using CNS (Brunger, et al., 1998), which flexibly permits defining the aforementioned AIRs, in the same fashion as in HADDOCK. Generated solutions are evaluated subsequently in terms of the AIR energy and van der Waals energy by means of a "selection plot". Solutions that have either bad AIR energy, i.e. high number of distance violations, or bad interacting van der Waals energy are discarded. The approach manages to effectively keep the ligand in the binding site, although it does not incorporate any information for discriminating between ligand orientations that occur within that binding site region...

The pioneering quantitative use of CSP for this problem, to the best of my knowledge, has been reported by McCoy and Wyss (McCoy and Wyss, 2000) (also implemented in SDILICON (Moyna, 2003)). Their method exploits the ring-current effects that ligand with aromatic rings produce on the protons of the protein. Aromatic rings not only constitute the main source of the contribution to the total CSP when present, but also produce a well-defined pattern of up- and down-fields on the surrounding spins, depending on the relative position and orientation. Profiting from developments in the theoretical prediction of CS in proteins, such as the program SHIFTS (Osapay and Case, 1991), the authors devised a protocol to align a ligand chemically similar to tryptophane to calmodulin. Consequently, they considered a probe tryptophane and generated a large number of bound poses of it to calmodulin. Using SHIFTS, they predicted the contribution to CS of the α-carbon protons of the protein that each pose would produce, and compared those values with experimental ones. The largest agreement corresponded to the orientations closest to the native of the original complex. This demonstrated the usefulness of such an approach for aligning ligands to their target proteins on the basis of CSP. The method has been already successfully applied in the context of SBLD programs (Wyss, et al., 2004; Gorczynski, et al., 2007). A further development on the method has been presented recently, where the task of generating ligand conformations is passed to the protein-ligand docking program GOLD (Verdonk, et al., 2003). Subsequently these starting poses are optimized according to their agreement to CSP (Cioffi, et al., 2008).

Very recently, a new method using a crude distance-dependent model for CSP called AutoDockFilter (Stark and Powers, 2008) has been reported. The authors assume that large amide CSP, negative or positive, must correspond to residues in close contact to the ligand. In a first step, large CSP are used to define the binding site area on which AutoDock (Morris, et al., 1998) docking is be performed. Ligand poses so generated are post-filtered according to the distance-dependent score in an AIR fashion: residues with largest CSP must have ligand atoms in closer contact than residues with smallest CSP. As in the case of LIGDOCK, such an approach ensures that the ligand remains in the binding site area, but does not provide information about how it orients once in there.

Upon binding, both ligand and protein proton resonance signals are affected, resulting in CSP. Arguably, ligand signals are easier to follow and assign, and do not present limitations on system size. This has been the kind of information exploited in the NMRScore approach (Wang, et al., 2004a; Wang, et al., 2004b; Wang, et al., 2007). There, selected CS of the ligand are computed at the semi-empirical MNDO level, speeded up by using a "divide and conquer" methodology for diagonalizing the Fock matrix. The authors have shown that such an approach is superior in recognizing native-like conformations of protein-ligand complexes compared to standard docking and scoring approaches. Care must be taken in that symmetries in the ligand structure can be misleading, and for that additional experimental information from the protein side is required. All tests of this approach have been reported so far for one complex and, despite being promising, its applicability at the scale required in a SBLD program remains uncertain. Computational times are rather lengthy. Probably, ligand refinement and/or postfiltering would be, for the time being, the most appropriate application for this approach.

The method to which this thesis is devoted shares many of the ideas of the previously presented, but taking them a step further in the approach. CSP are used quantitatively, so that not only protein binding regions are detected but also the actual orientation of the ligand with respect to the protein is directly obtained. The methods by McCoy and Wyss (McCoy and Wyss, 2000) or by Cioffi et al. (Cioffi, et al., 2008) do also use CSP quantitatively, but fail to acknowledge that the different CSP for different HN are predicted with uneven accuracy by empirical methods (Moon and Case, 2007). I will show that this has a relevant impact on the final result and requires an outlier-resistant

functional form for scoring. Additionally, it has been assumed that current docking engines are good enough at generating docking solutions, but as discussed above there are still significant gaps, particularly when dealing with large binding regions as in protein-protein interfaces or in fragment screening. This is a relevant issue to look at, considering that the large majority of the exposed methods use CSP information as post-filtering, i.e. relying on the ability of the docking engine to generate good docking solutions beforehand. Here I will propose an implementation that incorporates CSP information *at* docking time, in order to enhance the probability of generating native-like poses with respect to standard docking. I anticipate that some aspects of the problem are not going to be directly addressed in this work, as is the case of protein flexibility. However, the impact of not taking it into account will be evaluated.

3 Theory and methods

In what follows, the theoretical fundaments supporting the use of CSP in a protein-ligand context will be exposed, accompanied by a description of the methods and the dataset of protein-ligand complexes used in deriving our CSP-driven docking scheme. The semi-empirical model applied by Case et al. in the program SHIFTS for predicting CSP in proteins constitutes the foundation of our formulation (sections 3.1 and 3.2). Molecular visualization lies at the core of molecular modeling, and plays also a crucial role for the understanding of the problem on which this thesis is focused. In section 3.3 I discuss different visualization solutions proposed so far for facilitating comprehension of CSP effects of a ligand on a protein. I will explain the specific solution suggested in this work and justify its election. Concerning the actual development of the scoring function, I will explore in a first step several candidate mathematical formulations for a CSP-based scoring function in a post-filtering scenario. The statistical tools used for this evaluation (ANOVA analysis and coverage-error-plots) are described in section 3.4. In a second step, I will target the appropriate formulation of the hybrid scoring function, for which brief account on current docking failures will be exposed in section 3.5. The theoretical reasoning on how to combine the knowledge-based scoring function DrugScore with CSP information is described in sections 3.6 to 3.8. Finally, the reference docking method, the different data sets used along this work and their preparation are described (chapter 4).

3.1 Chemical shifts, chemical shifts perturbations, and their use in structure elucidation

The chemical shift (CS) of an atom is the most fundamental NMR observable (Levitt, 2001). Because atomic nuclei have spin, when an external magnetic field B_0 is applied on them, their precession frequency w_0 is affected, proportionally to that magnetic field and as a function of the atom's gyromagnetic ratio γ.

$$\omega_0 = \gamma B_0 \qquad\qquad \text{Eq. (5)}$$

Because the external field also induces currents in the electron cloud surrounding the nuclei, these currents generate in turn also an induced magnetic field which effectively

modulates the originally applied field at a local level. Thus, nuclei at two sites in the same molecule provided they are surrounded by different electronic environments experience different changes in their respective precession frequencies, as is the case of protons in –CH_3 group of ethanol and protons in –CH_2 group of the same molecule. Since this effect is caused by the induced field that valence and bonding electron produce, precisely those implicated in the chemical properties of the molecule, thus the name of *chemical shifts* (Levitt, 2001).

Practically, chemical shifts are expressed in terms of the ration between the difference in precession frequency of the target nucleus and a reference and the precession frequency of that reference, in parts per million:

$$\delta = \frac{\omega_0 - \omega_{ref}}{\omega_{ref}} \cdot 10^6 \qquad\qquad \text{Eq. (6)}$$

the typical references are protons of tetramethylsilane, $Si(CH_3)_4$ for experiments performed in organic solvents and for biomolecules those in the methyl groups of 2,2-dimethyl-2-silapentane-5-sulfonic acid.

Even if we consider the nucleus of a single isotope of an atom, the different magnetic/electronic environments in which this nucleus can be found give rise to differences in frequency of resonance, which are reflected in different CS. This phenomenon renders CS as an appealing parameter for studying the structure of matter. However, the challenge comes from the fact that the CS of a nucleus is a single number, and it is rather complicated to interpret it in geometrical terms when taking in isolation or as an absolute value. Fortunately, differences in CS to a reference state are easier to interpret. The phenomenon of a ligand binding to a protein causes a perturbation in the CS of the protein (also in those of the ligand). The unbound protein can be considered the reference state, to which CS in the complex can be compared, in order to deduce the orientation of the bound ligand.

3.2 Empirical modeling of chemical shifts perturbations

A change in the chemical environment of a nucleus results in a change in the field affecting it and therefore in the observed chemical shift $\Delta\delta$. Empirical models

decompose the total observed perturbation in *through-bond*, due to covalently bound neighbor atoms, and *through-space* effects, due to close-by interacting atoms, non-covalently bound (Szilagyi, 1995; Wishart and Case, 2001).

$$\Delta\delta = \delta_{through-bond} + \delta_{through-space} \qquad \text{Eq. (7)}$$

$\delta_{through-space}$ contributions include magnetic (electron ring currents and other magnetic anisotropies) and electrostatic (electric field, hydrogen bonds, solvation) effects. Nuclei most affected by these contributions are ^{1}H and ^{19}F atoms, since they are only bound to one other atom, while observed $\Delta\delta$ of nuclei such as ^{13}C or ^{15}N are mostly determined by the covalent structure. This particularity, makes these nuclei very interesting for deriving conformational restraints in biomolecular structure determination (Cornilescu, et al., 1999). In Table 3.1, adapted from reference (Wishart and Case, 2001), the different contributions that give rise to the observed CS for active nuclei in proteins are collected.

Table 3.1 Determinants of chemical shifts in proteins*.

Attribute	^{1}HN	^{15}N	^{1}Hα	^{13}Cα	^{13}Cβ	^{13}CO
Random coil	0	50	25	50	75	25
Torsions (Φ/Ψ)	0	0	50	25	10	50
Torsions (Φ/Ψ$_{i-1}$)	25	25	0	0	0	0
Side chain (χ)	5	0	0	5	5	5
Side chain (χ$_{i-1}$)	5	5	0	0	0	0
Hydrogen bonds	25	5	5	5	0	5
Ring currents	10	0	10	0	5	5
Local charges	10	0	0	0	0	0
Miscellaneous	20	15	10	5	5	10

*Data given as % of the total effect. The decomposition is done for average values. For example it is clear that not all atoms/groups are involved in hydrogen bonds or are in the vicinity of aromatic rings.

In the case of a ligand binding to a protein, the $\Delta\delta$ experienced by any nucleus on the protein side can be calculated as the difference between the δ in the complex and the free state:

$$\Delta\delta = \delta^{cplx} - \delta^{free} = (\delta^L_{t-bond} + \delta^L_{t-space} + \delta^{Pcplx}_{t-bond} + \delta^{Pcplx}_{t-space}) - (\delta^{Pfree}_{t-space} + \delta^{Pfree}_{t-bond}) \qquad \text{Eq. (8)}$$

where L denotes effects coming from the ligand and P effects coming from the protein, either free and in complex. It comes down to the direct effect of the ligand in the

binding site as a source of perturbation plus effects coming from the protein itself, which arise due to conformational rearrangements. If we assume no significant protein rearrangement upon ligand binding then δ^{Pcplx} equals δ^{Pfree}, which leaves only the effect of the ligand. Additionally, as I am considering the case where the ligand binds non-covalently, there is no $\delta^{L}_{t\text{-}bond}$ contribution, yielding:

$$\Delta\delta = \delta^{L}_{t-space} \qquad \qquad \text{Eq. (9)}$$

This equation can now, as explained before, be decomposed into magnetic effects, coming from electron ring currents of aromatic rings and other magnetic effects originated by chemical groups containing double/triple bonds or amide groups and electrostatic contributions:

$$\Delta\delta = \delta_{rc} + \delta_{ele} + \delta_{m} \qquad \qquad \text{Eq. (10)}$$

Where δ_{rc} is the ring current contribution, δ_{ele} is the electrostatic contribution, and δ_{m} accounts for other smaller magnetic contributions.

3.2.1 Ring current effects

The ring current contribution is typically segregated from other magnetic contributions since when it is present, it accounts for the largest part of the observed chemical shift perturbation (CSP). Indeed, there are approaches described in the literature that rely only on this contribution for orienting ligands with respect to their target proteins (McCoy and Wyss, 2000).
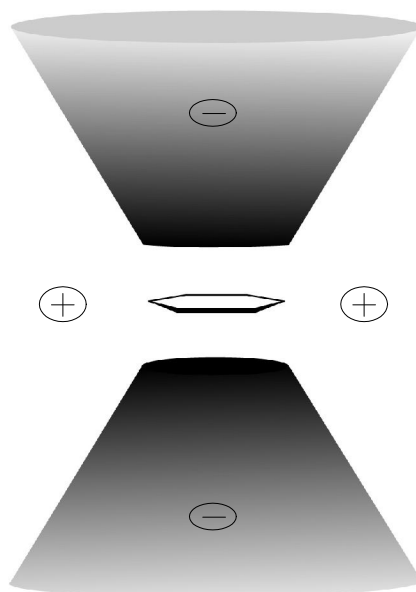
Figure 3.1 Ring currents around aromatic rings in solution induced by an external, static magnetic field. The shape of the ring current field is indicated by the red double-cone and by broken magnetic-field lines. The minus sign indicates that the NMR lines of hydrogen atoms located inside the cone in the three-dimensional protein structure are shifted "upfield", whereas for atoms outside of the cone the shifts are "downfield".

The sign of the chemical shift affected by ring currents changes as a function of ring orientation, and the intensity of the shift does as a function of distance from the ring. A single ring on the surface of a protein produces significant perturbations (~0.1 ppm) for protons 7-10 Å from the ring (Case, 1995). Out-of-plane orientations with respect to the aromatic ring give rise to negative upfield shifts and in-plane orientations give positive downfield shifts (Figure 3.1). The change of the sign of the aromatic ring current as a function of ring orientation typically produces a pattern of positive and negative proton shifts when aromatic ligands interact with a protein surface. These patterns are very sensitive to translation and rotation of the ring. Formulas for computing each of these terms have been implemented in programs for CS prediction in biomolecules such as SHIFTS (Osapay and Case, 1991) or SHIFTX (Neal, et al., 2006), recently used in predicting full protein structures from chemical shifts (Cavalli, et al., 2007), or in small molecules (Abraham, et al., 2000). The ring current contribution is generally modeled as:

$$\delta_{rc} = IBG(\vec{r}) \qquad\qquad \text{Eq. (11)}$$

where $\vec{r}$ is the vector from the observed proton to the aromatic ring, $G(\vec{r})$ is a geometric factor B collects constants that would yield the expected contribution from a benzene ring and I is the "ring current intensity" factor. The latter represents the ratio of the intensity expected for the ring in question relative to that of a benzene ring. Intensity parameters for the aromatic systems for this work were taken from Abraham *et al.* (Abraham, et al., 2000; Abraham and Reid, 2002). The geometric factor, from the Haigh-Mallion theory (Haigh and Mallion, 1980), is expressed as :

$$G(\vec{r}) = \sum_{ij} S_{ij} \left(\frac{1}{r_i^3} + \frac{1}{r_j^3} \right) \qquad \text{Eq. (12)}$$

Where r_i and r_j are the distances from ring atoms i and j to the proton and S_{ij} is the area of the triangle formed by atoms i and j and the proton projected onto the plane of the aromatic ring, as depicted in Figure 3.2(A). The Haigh-Mallion model has proven its superiority to other empirical methods for calculating ring current effects, such as the Johnson-Bovey or the point-dipole approximation (Moyna, et al., 1998).
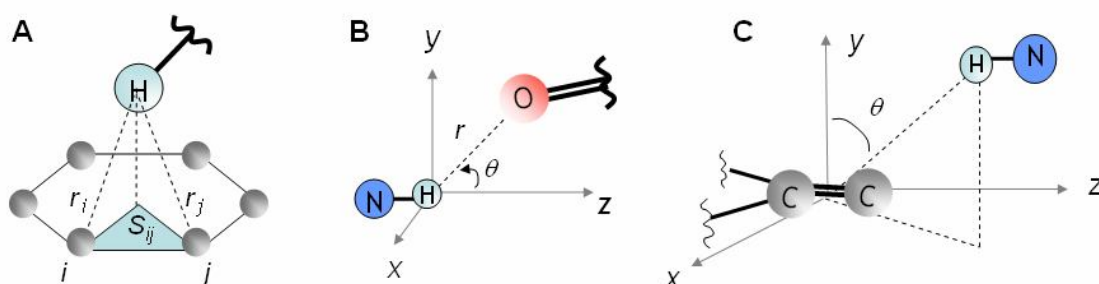
Figure 3.2 Geometric definitions used for the CSP calculations. (A) Ring current effects, (B) electrostatic effects, and (C) magnetic effects from anisotropy generating groups.

3.2.2 Electrostatic effects

The electrostatic contribution is typically computed as the projection of the electric field $\vec{E}$ calculated in vacuum onto the N-H bond, $\vec{E}(N\text{-}H)$, as seen in Figure 3.2(B), where solvent effects are neglected.

$$\delta_{ele} = A\vec{E}(N\text{-}H) = A\sum (q_i / r_i^2)\cos\theta_i \qquad \text{Eq. (13)}$$

A is an empirical constant, set it to $-1.2 \cdot 10^{-12}$ esu^{-1}, as in reference (Osapay and Case, 1991), q_i is the partial charge of atom *i*. Distance r_i and angle θ_i between the proton and each source of charge are specified in Figure 3.2(B).

3.2.3 Magnetic effects from other anisotropy-generating chemical groups

The magnetic anisotropy effect (Figure 3.2(C)) from systems with π-delocalized electrons (double / triple bonds in our case) is computed using McConnell's equation:

$$\delta_m = \frac{\sum_{i=x,y,z} \Delta\chi_{ii}(3\cos^2\theta_i - 1)}{3L_0 R^3} \qquad \text{Eq. (14)}$$

where χ_{ii} represents the axial component of the magnetic susceptibility tensor of the anisotropic group, *R* is the distance from the proton to the anisotropic group, L_0 is Avogadro's constant and θ_i is the angle between the *i*-axis and the vector connecting the proton with the chemical group.

3.3 Visualization of CSP

Information visualization is defined as the use of computer-supported, interactive, visual representations of abstract data to amplify cognition (Card, et al., 1999). Its usefulness in biology was acknowledged early on since the very first models of Myoglobin and Hemoglobin were made available by Kendrew and Perutz, respectively (Kendrew and Perutz, 1948; Perutz, 1949), establishing the field of structural biology. Visual examination of the structural differences between the two molecules permitted an accurate understanding and interpretation at the molecular level of the observed functional differences. Visualization is nowadays routinely used, as long as molecular structures or models are available. Visualization helps in integrating, understanding and rationalizing biological abstract data and supports rational planning of new experiments. Probably the most important application of scientific visualization in general comes from the ability of human cognition to visually and instantaneously detect patterns and features in large datasets, which outperforms any known pattern recognition algorithm. Furthermore: integration of images with previous theoretical knowledge facilitates

thinking and hypothesis generation and serves as a quality control for detecting unexpected features.

Approaches similar to ours which use structural information encoded in quantitative chemical shifts perturbations (QCSP) for structure elucidation have been already published. In all these approaches visualization has been used as supporting means to explain how QCSP serve in delimiting binding sites and/or orienting ligands in those binding sites. A summary of the most relevant of such visualization is presented in Figure 3.3. McCoy and Wyss (McCoy and Wyss, 2000) (Figure 3.3A) in describing their approach already discussed the drawbacks of the common CSP representation by surface mapping (Figure 3.3C) since SAR-by-NMR has been described. They pointed out that this kind of representation tends to overemphasize shifts of large, solvent exposed residues (Tyr, Lys, Arg, His) and under-emphasizes those of smaller residues such as Gly. Of course buried residues remained unseen. In addition, propagating the interaction to the whole surface of a residue is clearly misleading in those occasions where the interaction occurs only with the backbone of the protein. All this considered, they propose a visualization of the CSP as spheres centered on the actual atoms for which the CSP were measured. Up- and down-fields generated by the ligand, which can be structurally interpreted (e.g., relative orientation of an aromatic ring), translate in positive and negative CSP, represented by two different colors. Finally, differences in magnitude which reflect the distance from the source of the perturbation to the target atom are represented by proportional radii of the spheres. This kind of representation is also used by (Gorczynski, et al., 2007) (Figure 3.3B). The approach by Schieborr et al. (Schieborr, et al., 2005), focuses on a qualitative use of CSP, and, hence, only residues with the largest CSP are marked in red as in typical surface mapping. Surface mapping, despite the aforementioned inconvenience, has the advantage to show the steric constraints that the protein structure has and effectively defines the binding site. This information is not negligible, and combined with knowledge of the magnitude of the CSP can suffice for an expert to visually suggest or evaluate plausible binding modes of the ligand.
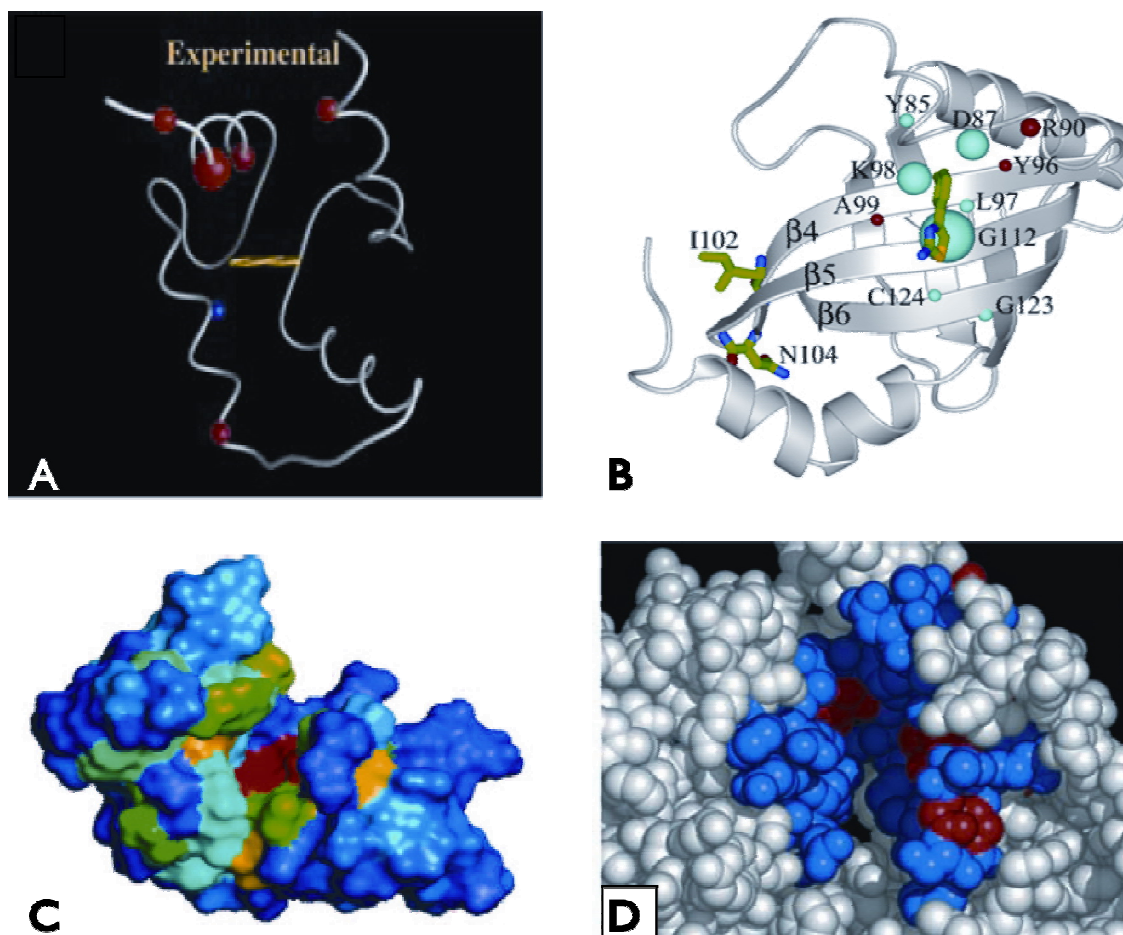
Figure 3.3 Different molecular visualization styles seen in methods published that use QCSP for docking ligands into their partner proteins. In A (McCoy and Wyss, 2000) and B (Gorczynski, et al., 2007), positive and negative QCSP are distinguished with blue and red colors while the radii of the spheres are proportional to the respective QCSP magnitudes. In C (Cioffi, et al., 2008) protein surfaced is colored following a red-green-blue scale on a per-residue basis, mapping the absolute intensity (largest to lowest) of amide CS; D (Schieborr, et al., 2005) residues colored in red are those corresponding to the largest shifts.

I propose to use the best ideas from both visualization styles to provide a more comprehensive view of all the information available: structural constraints and localization and magnitude of the observed perturbations at the same time. These ideas have been collected and are exposed in the following section.

3.3.1 Visualization methodology

In this study, visualization is done with PyMOL (DeLano). CSP values are mapped onto the B-factor field of the corresponding atom objects using a modified version of the

script data2bfactor.py from Dr. Robert L. Campbell (Campbell). The workflow is as follows:

1. PDBs are loaded into a new object in PyMOL.

2. Then, one creates a duplicate of the molecule (object called "exp") to hold the experimental data.

3. B-factor values and vdW radii of all atoms in the *exp* object are set to 0. Experimental data is loaded by means of the data2b_atom method, from data2bfactor.py script. I modified this script to not only assign data to the b-factor property but also adjust the vdW radius, which is set to the absolute value of the CSP. Then vdW radii are empirically scaled for optimum visualization.

4. *Exp* is then represented as spheres. Atoms with CSP larger than the pre-calculated average are colored in blue and those with CSP lower than the average in red. Atoms set to the average CSPr are colored in green and their vdW radius is set to 0.2 Å.

5. Finally, all bonds of amide protons are displayed as black lines, for better perception of the extent of the assignment. Figure 3.4 shows a screenshot of the visualization used for discarding CSP unrelated to the direct effect of the ligand binding for the case of 1ECV.
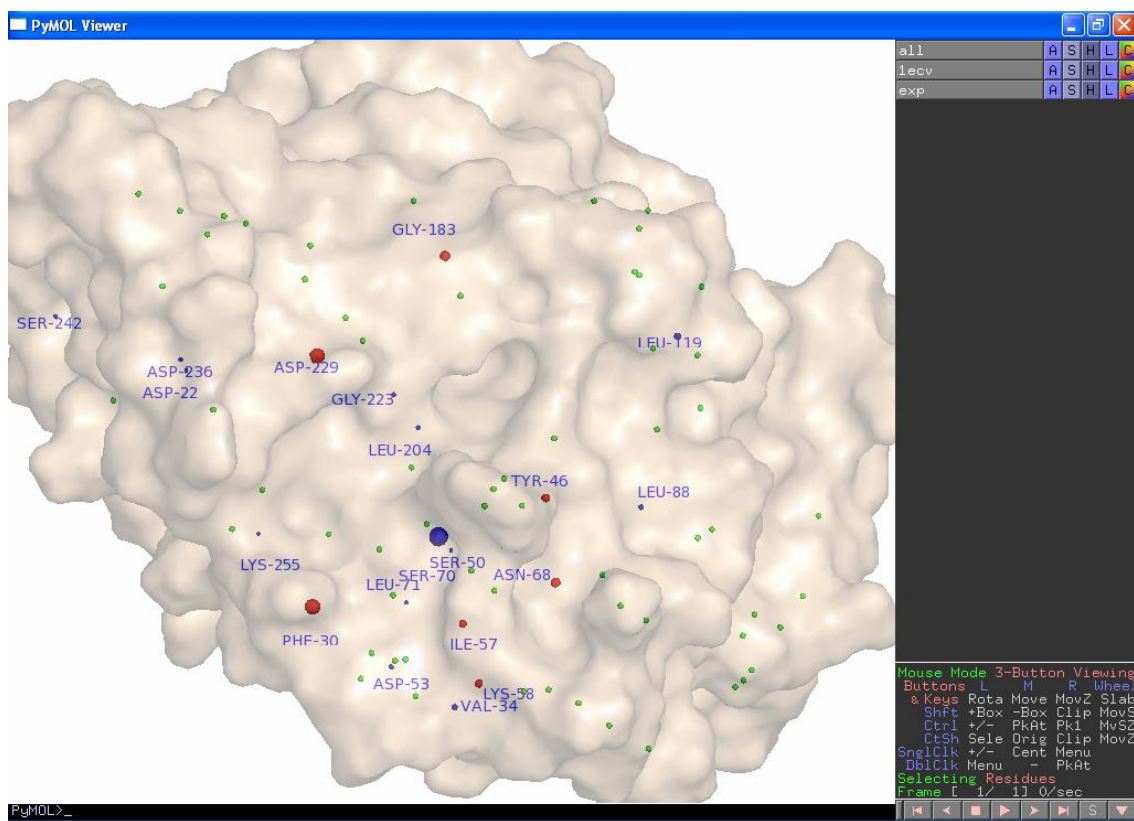
Figure 3.4 Visualization of experimentally measured QCSP for the complex 1ecv on the protein. Red spheres represent negative CSP and blue positive CSP. The size of the balls are proportional to the absolute value of the QCSP. Green balls are CS which have been assigned but do not change significantly. The accumulation of CSP around the central cavity is used to visually discard distant QCSP which are not likely to be consequence of the presence of the ligand in their direct vicinity. In this case Asp-22, Phe-30, Val-34, Asp-53, Lys-58, Ile-57, Leu-204, Asp-229, Asp-236 and Ser-242 could easily be discarded.

Representing the structure of the protein as a transparent surface makes apparent the deepest crevices (i.e., candidate binding sites) and steric constraints to which the ligand has to adapt to in order to bind the protein. CSP are mapped in the same way McCoy and Wyss suggest, that is: positive and negative CSP with two different colors and the spheres relatively scaled according to the absolute magnitude of the CSP. In addition, those CS which have been assigned but do not change upon ligand binding are also mapped as small green balls. This is useful as it helps identifying "forbidden" areas for the ligand. This visualization helps, additionally, to filter out CSP which are large in magnitude but are segregated from the "largest patch" of CSP, as they can be attributed to protein re-arrangement effects. Finally, CSP that would not fit in any reasonable binding mode hypothesis can be easily spotted and considered for further inspection as they might be indicating a potential assignment error.

Even though a human expert can visually assess potential solutions and easily decide on their plausibility, envisioning these solutions from the unbound protein and ligand structures is a much more difficult and markedly low-throughput task. Automatic computational methods such as molecular docking can, on the other hand, generate and evaluate millions of potential protein-ligand orientations, selecting the ones that fit best in terms of chemical and geometrical complementarity. It will be discussed below that despite its efficiency, molecular docking is also not a definitive answer to elucidating protein-ligand complex structures, since its accuracy can vary from case to case. This is particularly true for those cases where the ligand does not occupy the binding site area completely (see section 3.5 below).

3.4 Measuring agreement between experimental and back-calculated CSP: assessment of candidate scoring functions

There are several approaches in the literature aimed at scoring docking poses according to their agreement with experimentally observed CSP. Typically an objective function is required (E_{QCSP}) to be minimized. The straightforward functional form is based on a least-squared-minimization:

$$E_{QCSP} = \sum (\Delta\delta^{obs} - \Delta\delta^{calc})^2 \qquad \text{Eq. (15)}$$

where $\Delta\delta^{\{calc,obs\}}$ stands for observed and calculated CSP respectively. McCoy and Wyss (McCoy and Wyss, 2000) already realized that such an expression could be affected by the fact that the ligand does not always bind 100% to the protein and proposed a normalized variation, which also emphasizes the "fit of a pattern":

$$E_{QCSP} = \sum \left[\frac{\Delta\delta^{obs}}{\Delta\delta_{max}^{obs}} - \frac{\Delta\delta^{calc}}{\Delta\delta_{max}^{calc}} \right]^2 \qquad \text{Eq. (16)}$$

Later, Gorczynski et al. (Gorczynski, et al., 2007), who used the SDILICON program by Moyna (Moyna, 2003), which uses the same model as McCoy and Wyss, evaluated the agreement between experimental and back-calculated CSP using Pearson's correlation coefficient. Cioffi et al. (Cioffi, et al., 2008) suggested a similar approach,

based on the RMSD between observed and back-calculated CSP and defined their objective function as:

$$E_{QCSP} = \frac{\Delta\delta^{obs}}{\Delta\delta^{calc}} \qquad \text{Eq. (17)}$$

where $\Delta\delta^{calc}$ is the root-mean-square difference between the calculated and observed CSP values normalized by the absolute value of the observed:

$$\Delta\delta^{calc} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\frac{(\Delta\delta_i^{obs} - \Delta\delta_i^{calc})^2}{\left|\Delta\delta_i^{obs}\right|}} \qquad \text{Eq. (18)}$$

and $\Delta\delta^{obs}$ is the root mean absolute observed CSP:

$$\Delta\delta^{obs} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left|\Delta\delta_i^{obs}\right|} \qquad \text{Eq. (19)}$$

In addition, an association constant scaling factor was included as a variable to scale the experimental CSP values to allow for ambiguity in the extent to which the protein is bound in the NMR experiment. The authors remarked that this approach optimizes the structure based on the relative changes in chemical shift rather than absolute values. Either minimizing normalized least-squared-differences or Pearson's correlation coefficient has the benefit of overcoming the problem of having to predict correct absolute CSP values. This is important since it does not make the assumption that the protein binds fully to the ligand in the NMR experiment. According to this, our first candidate objective function was Pearson's correlation coefficient between observed and back-calculated CSP:

$$E_{QCSP} = \frac{\sum_{i=1}^{N}(\Delta\delta_i^{obs} - ave(\Delta\delta^{obs}))(\Delta\delta_i^{calc} - ave(\Delta\delta^{calc}))}{(N-1)s_{\Delta\delta^{obs}} \cdot s_{\Delta\delta^{calc}}} \qquad \text{Eq. (20)}$$

s stands for the standard deviations of observed and calculated CSP and N is the total number of pairs of CSP.

A Pearson's correlation based scoring scheme assumes that all CSP will be predicted on average with the same accuracy. This cannot be expected if an empirical model is used

for back-calculating CSP. Differences of exposure or localization in more rigid or flexible regions of the protein condition the individual accuracy of each predicted CSP. Most importantly, HN involved in hydrogen bonds with the ligand show a disproportionate error in prediction compared with the non-hydrogen bonded ones (Moon and Case, 2007). This renders a scoring function very sensitive to outliers, i.e., very large CSP will dominate the calculation. For this reason, I decide to also test Kendall's τ correlation (Kendall, 1938), which is a robust alternative to Pearson's:

$$\tau = \frac{4P}{n(n-1)} - 1 \qquad\qquad \text{Eq. (21)}$$

where n is the number of experimental and computed CSP pairs and P is the number of concordant pairs. A pair of experimental values (x_i, y_i) and a pair of computed values (x_j, y_j) is concordant if $\text{sign}(x_j - x_i) = \text{sign}(y_j - y_i)$. The values of τ go from -1 (perfect ranking disagreement) to +1 (perfect ranking agreement) and pass through 0, which denotes an independence of rankings.

Both Pearson's (Eq. 20) and Kendall's correlation (Eq. 21) functional forms will be considered and assessed as candidate objective function for scoring ligand poses with respect to the agreement between experimental and calculated CSP.

3.4.1 ANOVA analysis

In order to assess the discriminatory power of the proposed scoring schemes (Pearson's correlation and Kendall's correlation) I performed an analysis of the variance (ANOVA). This kind of analysis for scoring functions evaluation has been described by Seifert (Seifert, 2006). As explained by this author, ANOVA is applied to determine the proportion of the variability within the observed scores that is due to the scoring method itself as opposed to variability caused by random errors. Figure 3.5 illustrates how native-like solutions score sufficiently differently from decoys in both an ideal and a realistic scoring function.
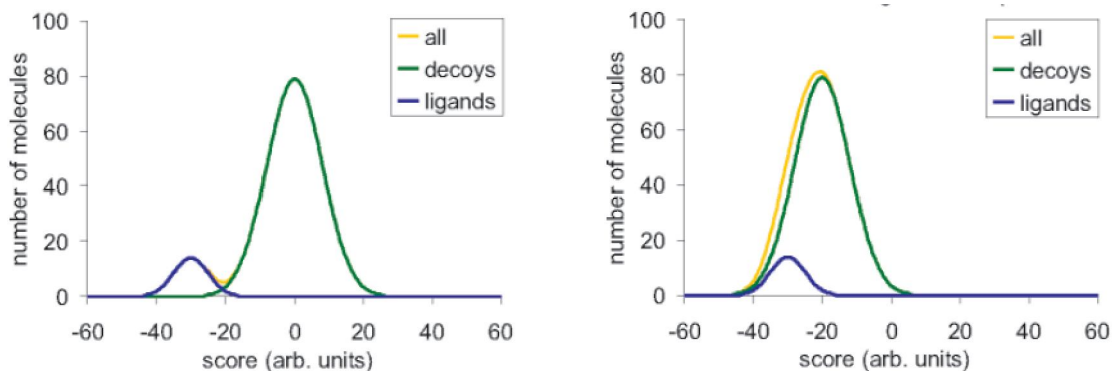
Figure 3.5 Score distributions for an 'ideal' scoring function (left) and a 'realistic' scoring function (simulated data). In the first case, native-like poses ('ligands' in the legend) score sufficiently differently from decoys. In a realistic scoring function there is some overlap in the scores, which results in false positives: identification of decoys as native-like solutions. Figure adapted from (Seifert, 2006).

The ANOVA analysis departs from the null hypothesis that there is no difference in the scores between native-like poses and geometric decoys. The rejection of this null hypothesis indicates that the scoring function is indeed able to discriminate between native-like poses and decoys, at a given significance level. The independent variable of the analysis is the assignment to the native-like pose group or the decoy group ($p=2$; $i=1…p$). N and n_i denote the total number of poses (m) per group, respectively. I perform our analysis following the seven steps described (Seifert, 2006):

1. First, I computed the following sums of scores (s):

$$S_1 = \frac{1}{N}(\sum_{i=1}^{p}\sum_{m=1}^{n_i} s_{mi})^2; S_2 = \sum_{i=1}^{p}\sum_{m=1}^{n_i} s_{mi}^2; S_3 = \sum_{i=1}^{p}\frac{1}{n_i}(\sum_{m=1}^{n_i} s_{mi})^2$$

2. Then, the total sum of squares $S_{tot}=S_2-S_1$, and $S_{treat}=S_3-S_1$, which measures the variability due to the scoring method is computed. $S_{error}=S_2-S_3$ accounts for the variability due to random errors.

3. The degrees of freedom for each sum are $df_{tot} = N-1$, *dftreat* $= p-1$, and *dferror* $= N-p$, respectively.

4. The corresponding variances are calculated dividing the sum of squares by their corresponding degrees of freedom.

5. Finally, measures for the discrimination of the two groups by a scoring method are given by the coefficients $\eta^2 = S_{treat}/S_{tot}$ and $\omega^2 = (S_{treat}-(p-1)\cdot\sigma)/(S_{tot}+\sigma_{error})$. Coefficient η^2 is the fraction of the variance within the actual data set explained by the scoring function. A weak, medium and strong effect of the scoring function in discriminating between native-like poses and decoys is characterized by η^2 values of $\leq 1\%$, ~6% and $\geq 14\%$ respectively.

3.4.2 Calculations framework

The aforementioned analysis is done on a comprehensive set of decoys and native-like solutions, described below. The CSP scores for each pose are computed in using in-house developed python scripts, which constitute an extension of the PDB libraries from the Biopython package (Hamelryck and Manderick, 2003). Statistical analysis are performed within R (Ihaka and Gentleman, 1996).

3.5 Limitations of current protein-ligand docking approaches

Protein-ligand docking, the problem of given the structure of a protein and a ligand, finding the right binding mode, has had a big impact in rational drug design, but is still an unsolved problem. There are three the main applications of docking: 1) binding mode prediction of known ligands, 2) identification of new ligands through virtual screening, and 3) predicting binding affinities of related compounds. These applications build on another, and thus constitute an increasing level of difficulty to get them right. This work is devoted to improve success with respect to the first application.

Binding mode prediction has been the most successful docking application since the inception of the method. All three main classes of scoring functions (force field-based, knowledge-based and empirical ones) achieve success rates between 70 – 90%, but this is target dependent. This means that some combinations of scoring schemes work better for some targets (Stahl and Rarey, 2001; Schulz-Gasch and Stahl, 2003; Ferrara, et al., 2004; Warren, et al., 2006).

Docking involves two problems: searching and scoring. During the first years, docking was typically done considering a rigid receptor and a rigid ligand, which permitted an exhaustive enumeration of mutual orientations. Now this is not any more feasible, given

the combinatorial explosion that occurs even if only ligand bonds are treated rotatable. This is a more realistic assumption, since the bioactive conformation doesn't need to be equal to the most stable of the free ligand. Needless to say that flexibility of the protein should also be accounted for, how to deal with it correctly remains a field of intense investigation (Cozzini, et al., 2008). For the case of a flexible ligand to be docked to a rigid target, the search problem is considered to be solved. That is, current algorithms and increasing computing power make it possible to sample relevant binding modes of a ligand to a protein, i.e., those which are geometrically and chemically complementary. But search is not absolutely decoupled from scoring: since the search is an optimization problem that occurs on the hyper-surface defined by the scoring function, a "rugged" landscape of it increases the challenge of the finding the global minimum. Therefore, the ideal scoring function should have 1) a global minimum that coincides with the native solution of the complex and 2) a funnel-like smooth shape that permits this solution to be found.

The lack of correspondence of the energy function minimum with the native conformation gives raise to so-called "decoys" (false positives) (Graves, et al., 2005). These are solutions which are physically reasonable but incorrect, with respect to the known native structure. Interestingly, the existence of decoys does not prevent docking programs from sampling the native solution: the problem is typically that this solution is not located at the minimum. Decoys are useful for understanding the weaknesses of the scoring function, since they tell about critical contributions of the scoring function that are not being taken into account or have been overweighed. Consequently, they serve as the basis for further improvements. Such an analysis is presented in the next section.

3.6 Reference method: Docking with DrugScore-only

DrugScore is one of the most reliable scoring functions for identifying (Gohlke, et al., 2000) and generating (Velec, et al., 2005) native-like docking solutions (Wang, et al., 2003; Ferrara, et al., 2004), with success rates of about 75%. Still, docking based on only DrugScore may fail due to, e.g., not considering interactions to structural waters, wrongly assigned protonation states, or missing entropy terms.

Throughout this work the conformational search engine of AutoDock 3.0.5 (Morris, et al., 1998) was used as a framework for development to which DrugScore potentials grids were plugged-in for scoring. From the several search schemes available in AutoDock I chose the Lamarckian genetic algorithm (LGA), which is considered as the most efficient alternative (Morris, et al., 1998). DrugScore grids, with 0.375 Å spacing, were calculated and scaled as previously described (Sotriffer, et al., 2002). Grids are centered in the binding site and cover a volume that extends at least 7 Å beyond any ligand atom in the native bound conformation. In the cases where an iodine atom was present in the ligand, it was substituted by bromine, as no potentials for iodine atoms were available in DrugScore.

3.6.1 Evaluation of docking success

A standard docking job consisted on 100 independent docking runs for each protein-ligand complex (initial population: 150; termination criterion: 1 million energy evaluations; mutation and cross-over rates: 0.8 and 0.02 respectively; elitism: 1; local search frequency: 0.06; maximum iterations: 300). After the 100 runs have been completed, ligand poses within 1.0 Å threshold are automatically clustered by the standard procedure implemented in AutoDock. That is, the best (lowest) scored conformation acts as a seed to construct the first cluster. All structures from the remaining 99 runs within 1.0 Å RMSD are clustered together with this one and removed from the starting pool of solutions. The process continues iteratively with the not-yet clustered until all 100 solutions have been assigned to a cluster. Solutions with RMSD smaller than or equal to 2.0Å to the native crystal conformation (reference) are classified as native-like (Cole, et al., 2005). Conformations with RMSD larger than 2.0 Å are classified as non-native-like solutions. Docking is considered successful if the largest cluster of solutions contains at least one native-like solution. Additionally, unsuccessful dockings according to this criterion are further divided into two groups: those that generated native-like solutions and those that did not.

3.7 Hybrid scoring: mixing DrugScore with CSP

Once the most appropriate scoring scheme with respect to CSP is established, our aim is to linearly combine it with DrugScore so that it can be routinely applied in a standard

docking scenario. In order to combine both contributions a reasonable solution is to proceed as in biomolecular structure elucidation protocols (Jack and Levitt, 1978; Brunger and Nilges, 1993), defining a hybrid energy function that must be minimized:

$$E_{hybrid} = E_{DS} + \omega_{QCsP} E_{QCSP}$$ Eq. (22)

In our case E_{DS} (i.e., DrugScore) accounts for "general knowledge" about protein-ligand complexes, E_{QCSP} for the specific data observed for the complex under study, and ω_{QCSP} is an empirical weight to be determined. The weighting of the E_{QCSP} contribution will reflect the user's "confidence" on the experimental data and the accuracy of the model to back-calculate it. This will vary from situation to situation. In biomolecular structure elucidation, for example, the most rigorous quantitative method to determine the optimal weight is complete cross-validation (Brunger, 1992) or Bayesian-based methods (Habeck, et al., 2006), both incurring in large computing times and relying on rather complete experimental data collections. These approaches are thus not transferable to the case presented here since real data sets are typically not that complete and efficiency must be kept.

Empirical scoring functions for docking, pioneered by Böhm (Bohm, 1994), share a similar spirit with the hybrid scoring presented here. There, the assumption is that the ligand binding energy can be decomposed as a weighted sum of uncorrelated physically-based terms (vdW term, electrostatics, hydrogen bond effects, etc.). In that case, the weights are determined from regression analysis using experimentally determined binding energies and, potentially, X-ray structural information. The resulting scoring functions depend on the molecular data sets used to perform regression analyses and fitting, which often yields different weighting factors for the various terms. This approach for finding the appropriate weighting cannot be applied to the hybrid scoring scheme because there is no experimental correlate to the pseudo-energy computed.

The strategy followed departs from the analysis of the docking "failures". I assume that the docking program is able to sample native-like solutions (Leach, et al., 2006a), but in some cases, due to a varied number of small inaccuracies, is unable to pick the right one. This yields a "likely" solution in terms of geometry and favorable chemical interactions. This solution has a better score than the native solution but is geometrically

different from it, which is why it is called a decoy (Graves, et al., 2005). This scenario is depicted in Figure 3.6, where the concept of the "inverse" energy gap, the scoring difference between a decoy and a native-like solution, is also illustrated. I anticipated that analyzing differences in E_{DS} and E_{QCSP} contributions between decoys and native-like solutions, respectively, will lead to appropriate bounds for the weighting factor.
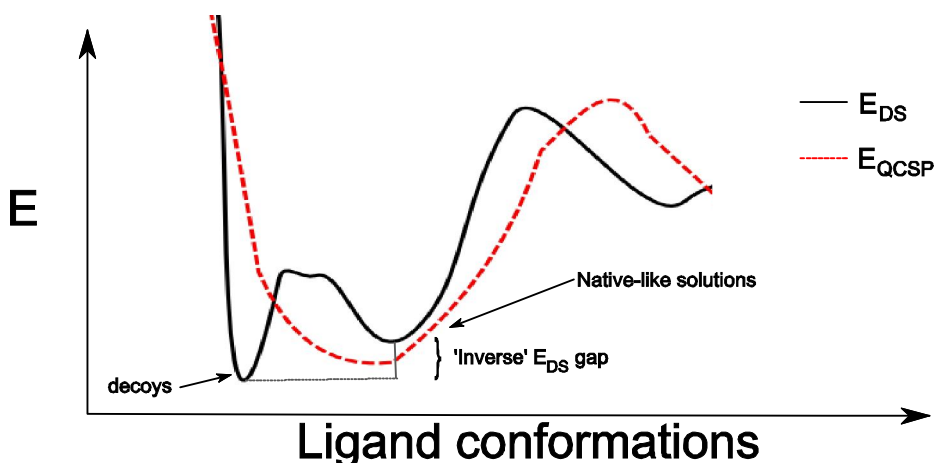
Figure 3.6 The E_{DS} inverse energy gap. Docking failures (decoys) typically occur in less accessible regions of the conformational space (continuous line). The expected effect of introducing the E_{QCSP} contribution is to "close" the access to those solutions while leaving open the access to the native-like solutions. The challenge is to appropriately weight this latter contribution because a very low weight would not discard the DrugScore decoys and too high a weight could eventually discard the native-like solutions.

I therefore studied the distribution of E_{DS} and E_{QCSP} gaps in a comprehensive data set, consisting of 70 protein-ligand complexes containing aromatic rings out of 85 complexes of the Astex diverse set (Hartshorn, et al., 2007). As detailed below, I performed standard DrugScore docking, consisting in 100 simulations for each protein-ligand complex. Out of each of these 100 simulations, according to the different results, complexes can be classified in three groups: those where the top-ranked solutions are native-like, those where native-like solutions were generated but these do not get the best score, and thus decoys are also present, and cases where only decoys and no native-like solutions were generated.

In order to analyze the respective E_{DS} gap for each complex between decoys and native-like solutions it is only possible to study those cases from the second group, where both species decoys and native-like solutions, were generated.

For analyzing the E_{QCSP} gap, first I computed CSP of HN protons induced by the bound ligand, because no experimental CSP is available for the Astex diverse set. E_{QCSP} is a correlation coefficient, and thus its values are limited between -1 and +1. This is a fundamental difference to E_{DS} where such theoretical limits do not exist and each complex behaves differently. For this reason, in the case of the analysis of the E_{QCSP} gap it is acceptable to consider all cases, including those where either no decoys or no native-like solutions were generated.

Analyzing how E_{QCSP} scores behave with respect to the RMSD to native-structures, provides and additional insight on the sensitivity of the scoring function to small conformational changes.

From the differences in both gaps, different weighting regimes will be proposed and explored.

3.8 QCSP-steered docking

In order to test the different candidate scoring functions and weighting regimes, the original AutoDock code was extended to include the contribution from the experimental CSP (E_{QCSP}) in a hybrid scoring scheme as in Eq. (22) (see Appendix I for implementation and usage). Practically, the docking proceeds as a standard AutoDock run (see below), generating and evaluating potential protein-ligand configurations. For efficiency, only if a generated configuration has a favorable E_{DS} interaction energy, the E_{QCSP} contribution is computed. The total time required for a docking with our hybrid scoring scheme increases linearly by a factor of ~ 3 with the number of aromatic rings considered. As an example, a single docking run of 1a9u (4 aromatic rings) takes 14 minutes on an Intel Pentium D 2.80 GHz CPU, whereas docking without the CSP information requires ~ 1 min.

Docking with QCSPScore was performed using the same grids as in DrugScore-only tests (see below).

4 Data set description and preparation

Two different datasets were used in this study, the first one for the theoretical development of the scoring function and the second one for validation with real experimental data.
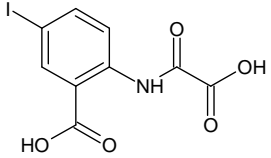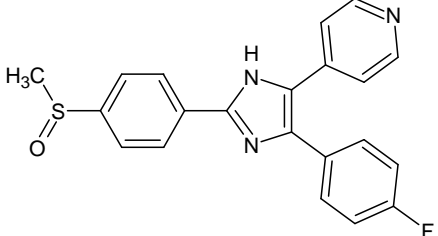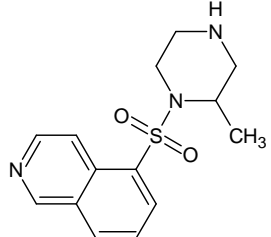
4.1 CSP Theoretical data – training set

The development data set I chose is the subset of the Astex diverse set (Hartshorn, et al., 2007) that contains ligands with aromatic rings (70 out of 85 complexes). The Astex dataset has been assembled considering high quality structures from the PDB. Ligands are all drug-like, and proteins are all drug-discovery or agrochemical targets. No target is represented more than once. PDB codes for the 70 complexes are: 1gpk, 1hnn, 1hp0, 1hq2, 1hvy, 1y6b, 1r58, 1sg0, 1sj0, 1t40, 1t46, 1tz8, 1u1c, 1unl, 1uou, 1v0p, 1v48, 2br1, 1n46, 1v4s, 1vcj, 1w2g, 1p62, 1hwi, 1j3j, 1k3u, 1ke5, 1kzk, 1lpz, 1lrh, 1nav, 1of1, 1of6, 1opk, 1oyt, 1p2y, 1q1g, 1q41, 1q4g, 1r1h, 1xoz, 1yqy, 1yv3, 1ywr, 2bsm, 1uml, 1ygc, 1x8x, 1r9o, 1gm8, 1s3v, 1z95, 1xoq, 1ia1, 1l2s, 2bm2, 1meh, 1pmn, 1ig3, 1t9b, 1xm6, 1jje, 1yvf, 1mzc, 1g9v, 1jd0, 1owe, 1tow, 1n2v, 1oq5.

4.2 CSP Experimental data – test set

The experimental set was limited to the three protein-ligand complexes already studied by Schieborr *et al.* (Schieborr, et al., 2005) (PDB: 1ecv, 1a9u and 1ydr) in the context of the development of a CSP-based method for solving protein-ligand complexes. Data was kindly provided by these authors. A crystal structure is available in the PDB for each case. The extent of CSP assignments in the binding site varies from 30 – 40% for 1ydr and 1ecv to 75% for 1a9u. All three proteins experience some re-arrangement upon ligand binding. The RMSDs from free to bound structures range from 0.73 Å in the case of PTP1b to 2.11 Å in the case of PKA, where the enzyme changes between "open" and "closed" forms (Table 4.1). Additionally, DrugScore-only docking did not succeed in any of the three cases, making possible a better evaluation of the benefits of introducing CSP for guiding docking.

Table 4.1 Data set used for validation

Ligand	PDB code[1]	Rmsd[2]
	1ecv (PTP1b)	0.73 (1pty)
	1a9u (p38)	1.01 (1p38)
	1ydr (PKA)	2.11 (1cmk)

[1] The protein name is given in parentheses.

[2] Rmsd between holo and apo structure, computed for all atoms of residues within 7 Å from the ligand. In Å. The PDB code of the apo structure is given in parentheses.

4.3 CSP data preparation

From the set of available HN CSP for each complex, values within ± 1 standard deviation from the average were considered as noise arising from the digital resolution of the spectrometer. Yet, these signals still provide some relevant information with respect to the orientation of the ligand because they indicate areas not affected by the binding. Thus, they were not discarded but re-set to the average CSP value. Non-transformed QCSP are visualized in the context of the protein structure (see below) in order to delimit the binding site area and discard isolated CSP presumably originating from protein re-arrangements. Table 4.2 summarizes the actual CSP used for each complex.

Table 4.2 Summary of HN CSP considered in the calculations with experimental CSP reference data

PDB code	Exp. CSP[1]	Signifi-cant CSP[2]	CSP set to avg.[3]	Discarded CSP[4]	CSP finally considered[5]	Binding site[6]
1ecv	82	20	62	Asp-22, Phe-30, Val-34, Asp-53, Lys-58, Ile-57, Asn-68, Leu-204, Asp-236, Ser-242	72	7/21
1a9u	193	20	173	Gln-11, Phe-42, Arg-57, Asn-100, Gly-219, Ile-297	187	17/22
1ydr	137	14	123	Glu-64, Gln-84, Lys-111, Ala-304, Glu-346	132	11/30

[1] Total number of experimental HN CSP available.

[2] Number of experimental HN CSP that deviate by more than one standard deviation from the average CSP over all HN nuclei.

[3] Number of experimental HN CSP whose value was set to the average CSP over all HN nuclei.

[4] Number of experimental HN CSP that were not considered in the calculation, although they deviate by more than one standard deviation from the average CSP over all HN nuclei. See Materials and Methods section for further explanation.

[5] Number of CSP finally considered in the docking.

[6] Number of HN within 7 Å of the ligand in the native structure that have CSP assigned and total number of HN within 7 Å of the ligand in the native structure.

4.4 Protein and ligand preparation for docking

The Astex diverse set was obtained directly from http://www.ccdc.cam.ac.uk/products/life_sciences/gold/validation/astex_diverse/.

Proteins in mol2 format were converted to PDB using Openbabel (Guha, et al., 2006). The ligand-bound protein complexes used for the experimental validation were obtained from the PDB. In both cases proteins were protonated with REDUCE (Word, et al., 1999). This step is needed to determine the position of amide protons for the evaluation of E_{QCSP}; DrugScore only considers heavy atoms. DrugScore grids were calculated using a grid spacing of 0.375 Å. The potential values were then scaled as previously described (Sotriffer, et al., 2002). Grids are centered on the binding site and cover a

volume that extends at least 7 Å beyond any ligand atom in the native bound conformation.

Ligands were converted to mol2 format with PRDRG (Schüttelkopf and Aalten, 2004). Atom types and AM1-BCC partial charges were calculated and assigned using ANTECHAMBER (Wang, et al., 2006). In the case of 1ydr, an iodine atom in the ligand was substituted by bromine, as no potentials for iodine atoms were available in DrugScore. All rotatable bonds were defined as active torsions in the AutoDock context. Ligands were visually inspected to detect aromatic rings. These groups were listed together with their corresponding intensities (see empirical model for CSP computation above) in an additional input file by referring to the atom numbers in the final AutoDock ligand file.

5   Results and discussion

5.1 General strategy

The aim of this project is to develop and validate an improved scoring function for docking through the use of experimentally measured CSP. On the one hand, state-of-the-art docking programs are able to sample native-like solutions. However, all scoring functions, on a target-dependent basis, may fail to prioritize these native-like solutions over non-native-like ones (Warren, et al., 2006). On the other hand, a fully experimental approach is too cost-ineffective, particularly in a setting where throughput is a concern. The approach presented here seeks a compromise between both scenarios of efficiency with low accuracy and accuracy with low efficiency. The goal is to improve docking success rate by supplementing it with easily obtainable sparse experimental data (QCSP).

The proposed scoring function is a linear combination of QCSP data together with DrugScore in a straightforward hybrid scoring strategy $E_{hybrid} = E_{DrugScore} + \omega_{QCSP} E_{QCSP}$, which I call QCSPScore. There the global score of a given ligand pose E_{hybrid} is the original DrugScore score $E_{DrugScore}$ plus a weighted term accounting for the agreement between experimentally measured CSP and theoretical CSP that the pose would produce on the protein. This simple formulation opened up, however, two important questions:

- how to measure the agreement between experimental and theoretical CSP?

- how to weight the experimental data?

In section 5.3 I address the first question. Using the data set of native-like and decoy geometries generated in a preliminary step (section 5.2) and the empirical model for back-calculating CSP described in the theory chapter, I tested two scoring schemes: Pearson's and Kendall's correlation.

The weighting of the experimental contribution is established through an exhaustive evaluation of what I defined as the "DrugScore inverse energy gap" between decoys and native-like solutions, together with their average scoring differences in E_{QCSP} terms. Results from this study are reported and discussed in sections 5.4 and 5.5. The resulting

candidate weighting schemes are applied in re-docking experiments performed on the Astex dataset, reported in section 5.6.

The final step of the development consisted in translating the conclusions obtained from the training set to the evaluation set, comprised of the three cases for which experimental data was available (section 5.7). This final evaluation served not only to assess the transferability of results derived from theoretical data, but also to analyze and assess the impact of limitations one faces in a real case situation, such as: the limitations of the model used to compute CSP, receptor flexibility, and extent and distribution of CSP assignment on the docking success.

An additional chapter compares results obtained with QCSPScore with the results that the method AutoDockFilter (Stark and Powers, 2008) would have produced the same data set. Finally, other methods are also considered for the discussion on the advantages that QCSPScore method represents and the aspects that require further study.

5.2 DrugScore performance on the Astex dataset: generation of native-like and decoy poses

DrugScore is one of the most reliable scoring functions for identifying (Gohlke, et al., 2000) and generating (Velec, et al., 2005) native-like docking solutions (Wang, et al., 2003; Ferrara, et al., 2004), with success rates of about 75%. Still, docking based on only DrugScore may fail due to, e.g., not considering interactions to structural waters, wrongly assigned protonation states, or missing entropy terms. As a consequence, docking decoys then obtain a more favorable DrugScore score E_{DS} than native-like solutions, leading to an "inverse" E_{DS} gap. To study this inverse E_{DS} gap one needs to generate and characterize real native-like and decoy geometries, which was the preliminary step performed. Since this involved re-docking a large set of protein-ligand complexes with DrugScore only, it served as an additional internal validation of the appropriateness of DrugScore for the tasks I am aiming at.

Developing a method as the one proposed in this study requires a sufficiently large and varied number of protein-ligand complexes with: a) known high-quality structure and b) available CSP data. Typically if the X-ray structure is known (high-quality) there is no need to attempt to observe and assign CSP. Otherwise, if CSP are available it generally

means that crystallization was not possible and thus it is unlikely that a high quality structure for reference is available. The search public databases of structures (PDB (Berman, et al., 2000)) and NMR data (BMRB (Ulrich, et al., 2008)) for such complexes was unsatisfactory, since the few cases where both CSP and high-quality structures were available were also successfully re-docked by DrugScore alone. Recently Stark and Powers, who have developed AutoDockFilter, a method in a similar spirit to ours, faced the same difficulty in assembling a data set and they resorted to "simulate" the experimental data (Stark and Powers, 2008). The group of Professor Schwalbe, who developed LIGDOCK (Schieborr, et al., 2005), have kindly provided me with experimental data for three protein-ligand complexes. With these antecedents, the strategy was to develop QCSPScore by first simulating CSP data for a large and varied data set and in a second step testing that development on the three complexes for which I had experimental data.

The training dataset consisted of 70 out of the 85 complexes from the "Astex diverse set" (Hartshorn, et al., 2007) that contain ligands with aromatic rings (see Materials section). I re-docked these protein-ligand complexes using DrugScore only, in order to generate representative native-like solutions and geometric decoys. Geometric decoys are those poses which are not native-like but receive better scoring than the native-like ones (Graves, et al., 2005). In the terminology of a classifier, they would be "false positives". The re-docking exercise provides the necessary examples and counter-examples for evaluating the candidate scoring schemes. Each docking consisted on 100 runs, thus generating 100 protein-ligand conformations for every complex. These 100 results for each complex are subsequently clustered by geometric proximity and classified as native-like if their RMSD to the native structure is < 2.0Å or non-native-like otherwise. Finally, non-native-like solutions are classified as decoys if they obtain a better score than the native-like ones. This classifications result in three groups of complexes: I) complexes that generated only native-like solutions, i.e., the first ranked cluster corresponds to a native-like solution, II) complexes that generated only decoys, i.e., only non-native-like solutions, and III) complexes that generated both decoys and native-like solutions, i.e., the native-like solutions were not first ranked. Complexes, according to this classification, are collected in Table 5.1.

Table 5.1 Classification of cases according to the solutions generated by DrugScore-only docking

No decoys[1]		Decoys + native-like[2]	Only decoys[3]
1gpk	1r58	1ia1	1g9v
1hnn	1sg0	1ig3	1gm8
1hp0	1sj0	1jje	1jd0
1hq2	1t40	1l2s	1mzc
1hvy	1t46	1meh	1n2v
1hwi	1tz8	1pmn	1oq5
1j3j	1u1c	1r9o	1owe
1k3u	1uml	1s3v	1tow
1ke5	1unl	1t9b	1yvf
1kzk	1uou	1xm6	
1lpz	1v0p	1xoq	
1lrh	1v48	1z95	
1n46	1v4s	2bm2	
1nav	1vcj		
1of1	1w2g		
1of6	1x8x		
1opk	1xoz		
1oyt	1y6b		
1p2y	1ygc		
1p62	1yqy		
1q1g	1yv3		
1q41	1ywr		
1q4g	2br1		
1r1h	2bsm		
48 of 70 (69%)		13 of 70 (19%)	9 of 70 (13%)

1 Cases where the first ranked cluster had an average RMSD < 2.0Å to the native solution.
[2] Cases where the first ranked cluster had an average RMSD > 2.0Å to the native solution but also some clusters with RMSD < 2.0Å were generated.
[3] Cases where no-native-like solution was generated

Independently of this classification, as anticipated, I also re-evaluated DrugScore success rates for this data set. As discussed in the Theory chapter, docking is considered successful if the largest cluster of solutions contains at least one pose with a RMSD < 2.0Å to the crystal native structure. Out of the 70 complexes 52 were re-docked "successfully" according to this criterion, which represents a 74% success rate, as expected for DrugScore from previous studies (Wang, et al., 2003; Ferrara, et al., 2004). Unsuccessfully re-docked complexes were: 1r58, 1yvf, 1mzc, 1g9v, 1jd0, 1owe, 1tow, 1n2v, 1oq5, 1r9o, 1gm8, 1s3v, 1z95, 1xoq, 1l2s, 2bm2, 1xm6, 1jje.

This data set was also re-docked with Goldscore within GOLD (Jones, et al., 1997) in the original work where it was described (Hartshorn, et al., 2007). It is interesting to observe that despite fundamental differences in the nature of Goldscore and DrugScore scoring functions, the former empirical and the latter knowledge-based some of the unsuccessful re-docking cases are shared, namely those affected by the presence of structural waters mediating protein-ligand interactions (1g9v, 1gm8, 1xm6 and 1r9o).

5.3 Measuring agreement between experimental and back-calculated CSP: Pearson's vs. Kendall's correlation as candidates schemes for scoring according to QCSP

Pearson's correlation has been used before (Gorczynski, et al., 2007) for measuring agreement between experimentally observed CSP and theoretically back-calculated ones. Others relied on minimization of root mean squared errors for the same purpose (McCoy and Wyss, 2000; Cioffi, et al., 2008). Both are intuitive methods for such task. However, they both are inappropriate in those cases where outliers are present. Measured CSP on the protein side due to ligand binding do not distribute homogeneously: only the protons very close to the perturbation, i.e., those in the binding site, experience a very significant shift. In particular, given the rapid diffusion of the CSP effect with distance ($\sim r^{-3}$ for the ring current effect), sets of observed CSP typically comprise a large majority of "unaffected" CS and a few which are, comparatively, largely affected. Thus, if all data must be included into the calculation, largely perturbed CS and unperturbed ones, Pearson's correlation or sum of squared differences is inappropriate, since they are largely dominated by those large shifts, whilst the non-changing CS provide structural information which is equally accurate and valuable. For this reason, I set to explore and compare the behaviour a rank-correlation based scoring would have, in this case Kendall's rank correlation (Kendall, 1938).

In order to compare Pearson's correlation and Kendall's correlation scoring schemes I performed a double analysis on the generated native-like and decoys: first an ANOVA analysis to assess the discriminatory power of both schemes; second an analysis of the energy landscapes generated by both schemes.

5.3.1 ANOVA analysis

Recently, Seifert has proposed the coefficients η^2 and F from ANOVA analysis as suitable parameters for guiding scoring function development efforts (Seifert, 2006). These coefficients measure respectively "explained variance" of the scores by the scoring method -discriminatory power- and signal-to-noise ratio of scoring functions with respect to native-like and decoy geometries. Following those ideas, I analyzed both Pearson's and Kendall's correlation based scoring schemes in order to find evidence that would favour one scheme over the other.

The ANOVA analysis was performed as described in the Methods chapter. η^2 and F values from the ANOVA analysis for Pearson and Kendall's correlation based scoring schemes resulted, respectively, in 44%, 205.30 and 59%, 369.95. This results show the superiority of Kendall's correlation based scheme over the Pearson's correlation based one. Still, in absolute terms, both figures are good enough as to affirm, according from this analysis, that either scheme would be suitable for scoring ligand poses. This is according to the cut-offs described by Seifert, where a $\eta^2 > 14\%$ represents a "strong effect" of the scoring function for explaining the variance in scores in both groups, native-like and decoys. The ANOVA analysis gave support to introducing Kendall's correlation based scoring, but was inconclusive with respect to determining its superiority over Pearson's correlation. Preliminary tests, however, showed that one obtains consistently better results when using Kendall's based scoring. The next section is devoted to investigate the reasons explaining it.

5.3.2 Advantages of robust correlation over non-robust for driving docking

Preliminary tests not shown here indicated that Kendall's based correlation scoring performs consistently better than Pearson's correlation based. The little difference in performance showed by the ANOVA analysis between both schemes must then have a large influence when moving from a re-scoring exercise to a re-docking one, which I try to rationalize below.

A first attempt to analyze the different impact of both scoring schemes can be done by looking at the number of false-positives that they generate. It is important to remember

that a decoy is indeed a false positive. If CSP contribution is expected to rescue native-like solutions that score worst than decoys in DrugScore terms, then this contribution must introduce itself a low number of those false positives. In Figure 5.1 I have plotted all DrugScore native-like and decoy poses generated against their respective Pearson's and Kendall's correlation coefficient. The plots are divided in four quadrants by dotted lines. The top left quadrant comprises ligand geometries which are native-like and, accordingly, attain a high correlation. Bottom-right geometries are decoys, which correctly obtain also a low correlation. Bottom-left, despite of being geometrically "correct", obtain correlations comparable to the ones obtained by decoys, i.e. they are false negatives. Finally, the top-right quadrant includes false positives. Clearly, in the case of Pearson's correlation the number of false positives is larger than in the case of Kendall's correlation (36 vs. 4). Already in a re-scoring setting, this test is telling us that bad solutions are much more likely to get a good score when evaluated in terms of Pearson's correlation than when evaluated in terms of Kendall's correlation. The next step is to explore why it is happening.
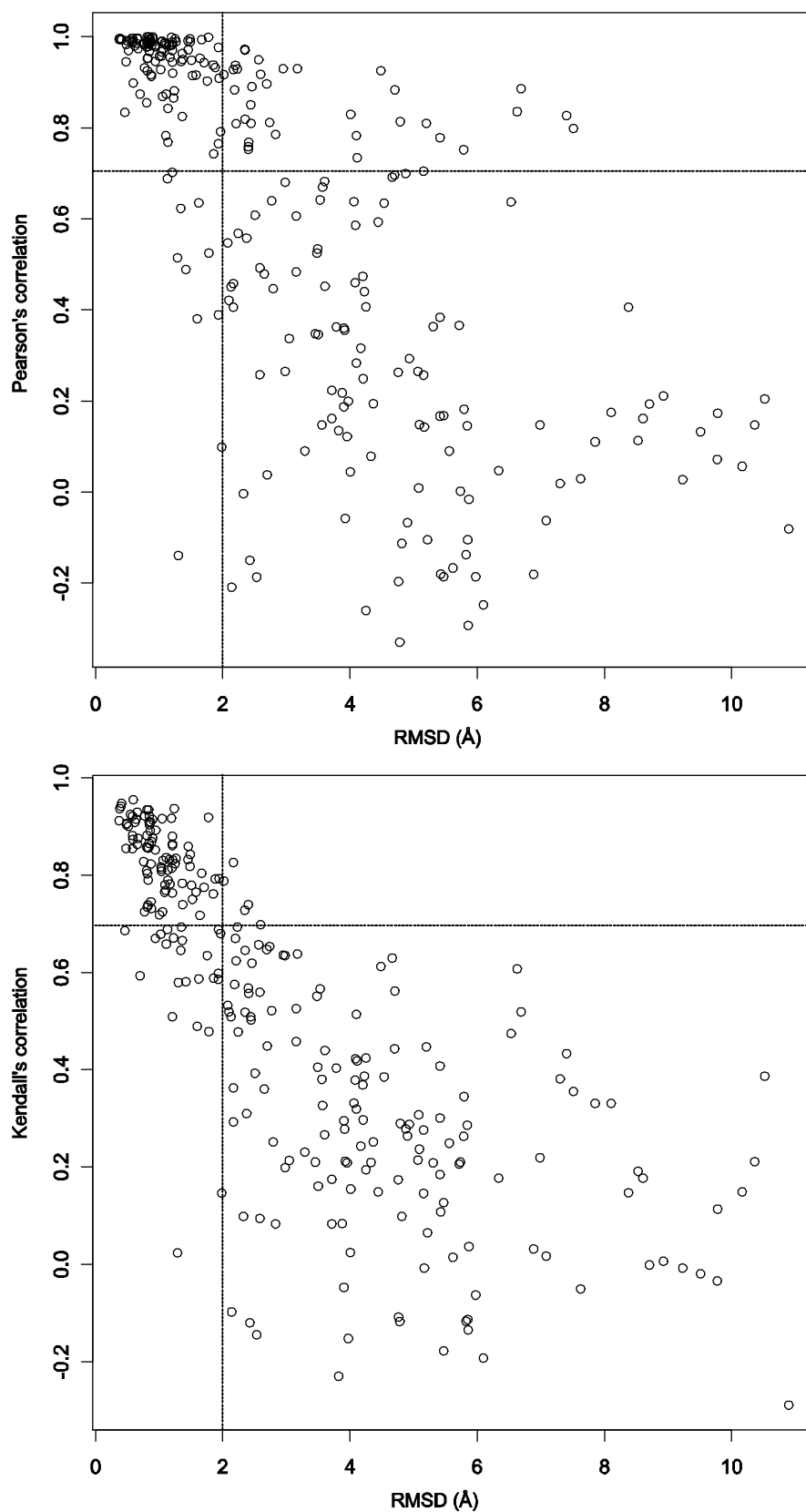
Figure 5.1 Pearson's and Kendall's correlation vs. RMSD for each ligand orientation considered in the training data set. Dotted vertical line divides native-like and decoy geometries. Horizontal dotted line divide "high"-score geometries from "low" score ones.

If a decoy pose gets a higher Pearson's correlation than expected, the reason is the presence of outliers. It is worth considering that from the total pool of observed amide CSP of a protein-ligand complex, a few have considerably larger values than the majority. In this regard, non-robust statistics are strongly influenced by those large values. To exemplify and illustrate this situation I will study in the case of the complex 1ygc. In Figure 5.2 I have plotted theoretical CSP for the native complex structure vs. the back-calculated ones corresponding to a decoy and a native-like solution generated by DrugScore, together with their respective Pearson's and Kendall's correlation coefficients.
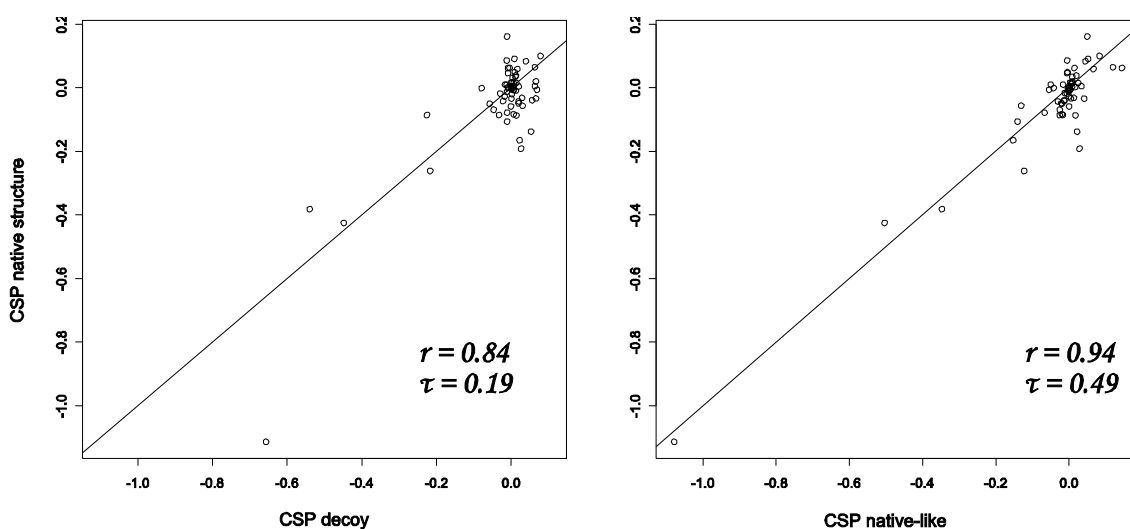
Figure 5.2 Scatterplot of theoretical CSP for the native conformation and back-calculated CSP for one of the decoys generated by DrugScore for 1ygc (left-hand side) and a native-like solution (right-hand side), together with their corresponding Pearson's and Kendall's correlation coefficients.

By observing both scatter plots in Figure 5.2 one can easily identify a substantially better correlation for the native-like pose than for the decoy, even though respective Pearson's correlations differ only by 0.1 (0.94 vs. 0.84). Only one CSP, substantially larger than the others, is disproportionately influencing the correlation coefficient. On the contrary, Kendall's correlation does not give a very high score to the native-like solution (0.49) but it is sufficiently different from the correlation in the non native-like case. In fact, in relative values, the difference between the score of the decoy and the native-like solution in Kendall's correlation case is three times larger than in the case of Pearson's correlation based. This makes Kendall's correlation based scoring a more

appealing scheme, particularly for these cases where a single large CSP dominates the whole calculation.

To study this phenomenon in a structural context, I have mapped the CSP onto the protein in Figure 5.3.
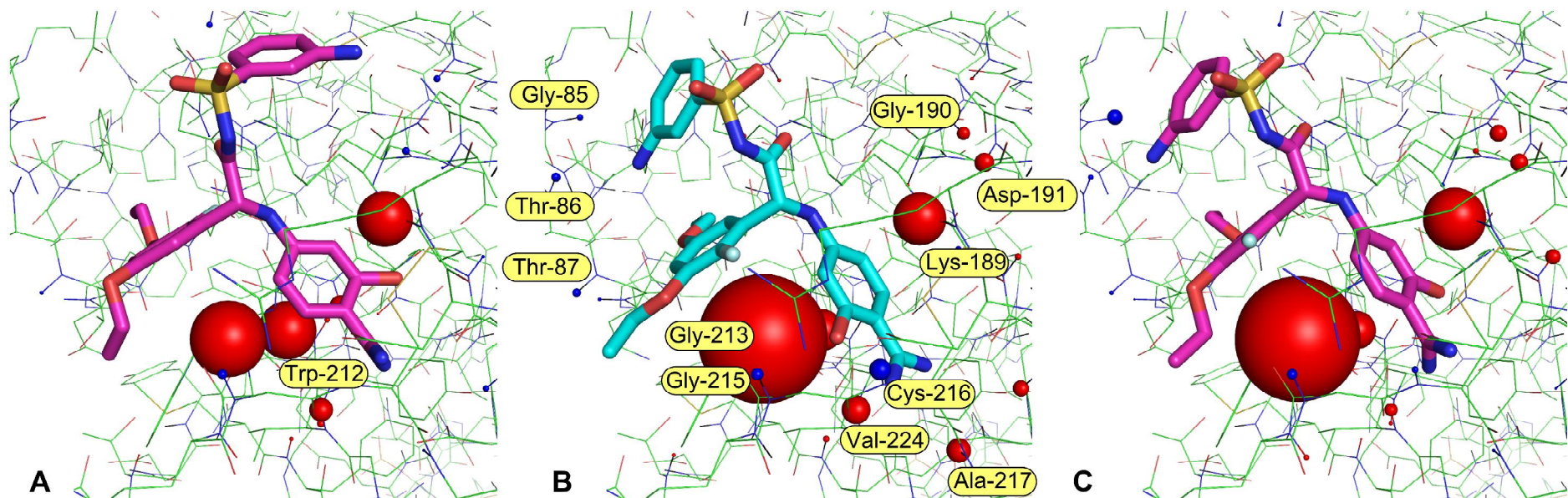
Figure 5.3 Comparison of a decoy (A) and a native-like solution (C) to the native structure of 1ygc (B). Amide protons of the protein are represented as balls, scaled according to the magnitude of their theoretical CSP generated by the presence of the ligand. Red balls represent negative CSP (downfields) and blue positive (upfields). The CSP pattern generated by the native-like pose in C resembles more the one generated by the native structure in B than the decoy in A does, even though the geometry of both decoy and native-like pose does not differ much.

Figure 5.3A shows one of the decoys for 1ygc. The corresponding CSP it would have produced on the protein are depicted as balls, centred on the amide protons and scaled according to the absolute magnitude of the CSP. Compared to the native structure (Figure 5.3B), only the phenyl ring in the top is flipped, and the other two ones slightly tilted. With respect to the CSP pattern, two main differences between the decoy and the native structure are worth noting: 1) the difference in the magnitude of the CSP of Gly 213, induced by the different tilting angles of two lower phenyl rings of the ligand and 2) the different patterns in upfield CSP (blue balls) involving the residues in the upper part of the figure, caused by the different orientation of the phenyl ring at the top. A subjectively evaluation these two differences in the patterns suggests that the first one, a reduced magnitude of a single CSP, is not that critical than the second one, where some CSP disappear and new ones appear. However, the CSP of Gly 213 plays a greater role in computing Pearson's correlation, just because of its absolute magnitude.

Collected evidence from ANOVA analysis and the analysis of the sensitivity to outliers for both schemes points to Kendall's correlation as superior scoring scheme for the purposes I am aiming at. I set thus $-E_{QCSP}$ to be Kendall's τ.

It is interesting to note that these results emerged even in the ideal setting where only theoretical data has been used. It is not unreasonable to expect new sources and classes of outliers in a scenario where calculations are done with real experimental data. These new outliers will appear from imperfections of the theory and will compromise even more the convenience of a Pearson's correlation based scheme.

In what follows, I study the energy gaps between decoys and native-like solutions, both in E_{DS} and E_{QCSP} terms, in order to establish the most appropriate weighting strategy for mixing both contributions.

5.4 Energy gap analysis between native-like configurations and decoys

From the QCSPScore equation, $E_{hybrid} = E_{DrugScore} + \omega_{QCSP} E_{QCSP}$, the weight for the experimental contribution ω_{QCSP} must be established. To that end I started by studying the respective "energy gaps", both in E_{DS} and E_{QCSP} terms, between native-like solutions

and decoys (see Theory). The rationale of mixing DrugScore with CSP has been that, while decoys score better than native-like solutions in E_{DS} terms, they do otherwise in E_{QCSP} terms. As a consequence, an appropriate weighting of E_{QCSP} will compensate the "negative E_{DS} gap". And thus, finding out whether a general weighting scheme can be applied becomes a key point of this investigation. The strategy has been to start by analyzing differences in E_{DS} and E_{QCSP} contributions between decoys and native-like solutions, respectively. While DrugScore values are different for each protein-ligand system, E_{QCSP} values, being a correlation coefficient are constrained between -1 and +1. The question is double: is there a typical DrugScore energy gap between decoys and native-like solutions, possibly reflecting known missing terms in the scoring function such as vibrational entropy loss? And do native-like solutions have a typical value for E_{QCSP}, sufficiently different from non-native-like ones?

For analyzing the E_{DS} gap between decoys and native-like solutions it is only possible to study the group of 13 cases (group II, Table 5.1) for which both species were generated. Figure 5.4 shows the distribution of energy differences between the average energy of the best-scored cluster in each case containing decoys and the average energy of the cluster with the best average RMSD in the whole run. Interestingly, 46% of the cases have a low energy gap (< 0.5 DrugScore units), with an average energy gap of 0.61 units and a maximum gap of 1.31 units.
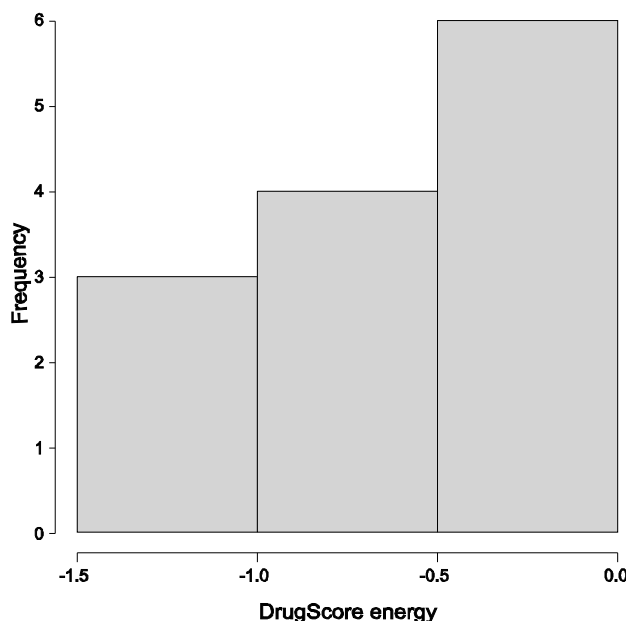
Figure 5.4 Distribution of DrugScore energy gaps between decoys and native-like solutions for the 13 cases where DrugScore-only docking generated both native-like solutions and decoys. In 46% of the cases, the energy gap is < 0.5 DrugScore units, with a maximum gap of 1.31 units.

For analyzing the E_{QCSP} gap, I used the computed CSP for HN protons described in section 5.2. Given that the E_{QCSP} contribution is represented by the negative Kendall's rank correlation coefficient, native-like solutions are expected to have high correlation values and thus, score values close to -1 whereas decoys should score zero or more. The analysis presented in section 5.3 has already disproved this hypothesis. Due to the sensibility of CSP to small geometric variations, very high scores are almost never obtained. In particular, the analysis of the complexes in the group II (both native-like solutions and decoys were generated) revealed an average gap of only 0.47, with large standard deviations for both decoys and native-like solutions of 0.22 and 0.23, respectively (Table 5.2).

Table 5.2 E_{QCSP} **scores for decoys and native-like solutions and differences for those cases of complexes of the Astex data set where both decoys and native-like solutions were generated when using DrugScore-only docking.**

PDB	$-E_{QCSP}$ decoy[1]	$-E_{QCSP}$ native-like	Difference
1ia1	0.60	0.56	-0.04
1ig3	0.23	0.82	0.59
1jje	0.01	0.79	0.78
1l2s	0.17	0.72	0.55
1meh	0.10	0.54	0.44
1pmn	0.41	0.85	0.44

PDB	$-E_{QCSP}$ decoy[1]	$-E_{QCSP}$ native-like	Difference
1r9o	-0.07	0.85	0.92
1s3v	0.21	0.63	0.42
1t9b	-0.10	0.85	0.95
1xm6	0.19	0.86	0.67
1xoq	0.45	0.47	0.02
1z95	0.09	0.09	0.00
2bm2	0.54	0.90	0.36
Average:	0.22	0.69	0.47
Standard Deviation:	0.22	0.23	0.33

[1] E_{QCSP} is the negative Kendall's rank correlation coefficient between computed CSP for a given ligand pose and CSP reference data computed for the native state.

A more representative analysis can be performed by expanding this sample to those cases where either no decoys or no native-like solutions were generated (groups I and III from Table 5.1). The distributions of the $-E_{QCSP}$ scores for these cases are shown in Figure 5.5: 79% of the native-like cases score over 0.7 while only 4% of the decoys obtain E_{QCSP} contributions worse than that figure. This confirms an intuitive expectation: I) native-like solutions show poorer scores than expected; II) E_{QCSP} scoring is more specific than sensitive, that is, it is better at rejecting decoys than favoring native-like solutions, which in line with the original goal for QCSPScore. As already stated, the fact that some native-like poses show poor E_{QCSP} scores reflects a "hard" scoring term character (Schulz-Gasch and Stahl, 2003), where slight changes in the orientation of a ligand with respect to the native state translate into large changes of the CSP pattern.
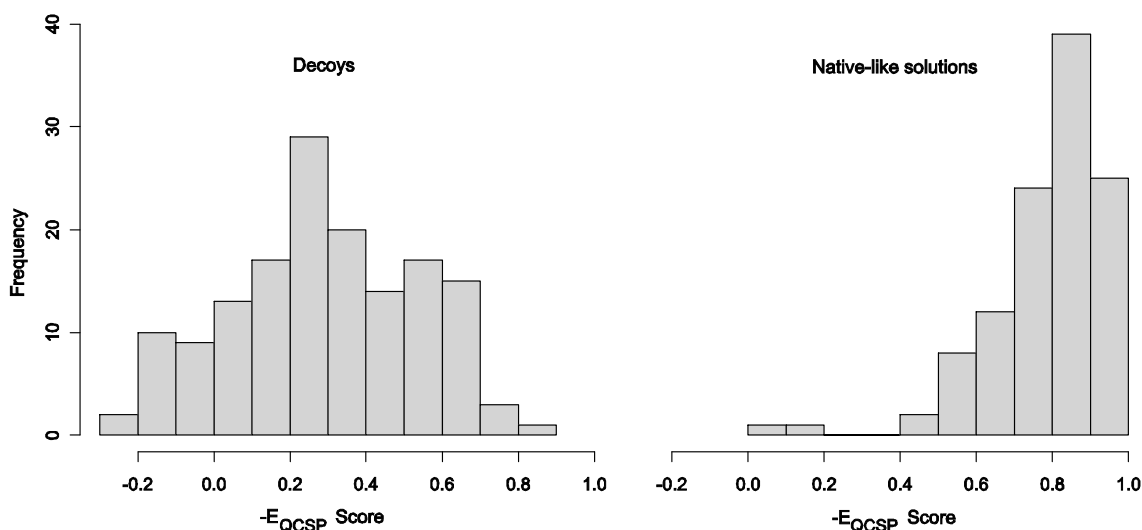
Figure 5.5 Distribution of the $-E_{QCSP}$ contributions for decoys and native-like poses, considering the subset of the Astex data set that contained ligands with aromatic rings. The CSP data for the bound ligand configuration was computed.

5.5 Weighting of the E_{QCSP} contribution

The analysis of both the E_{DS} and E_{QCSP} contributions for decoys and native-like solutions leads us to suggest three alternative weighting factors:

I)   As E_{QCSP} gap, I take the difference of the median values of decoys (0.29) and native-like solutions (0.82), which results in $\Delta E_{QCSP} \approx 0.6$. Since the sample for establishing the E_{DS} gap is much smaller, I take the maximum E_{DS} gap (1.31 units) plus one standard deviation (0.45), which results in $\Delta E_{DS} \approx 1.80$. The weighting factor to compensate an "inverse" E_{DS} gap by E_{QCSP} then results in $\omega_{QCSP} = 3$.

II) I consider different cut-offs for the E_{QCSP} gap, in order to include 75% of the decoys (lowest scores; 3rd quartile: 0.47) and 75% of the native-like cases (best scores; 1st quartile: 0.73), which results in $\Delta E_{QCSP} \approx 0.2$. Considering the same E_{DS} gap as in I), I then set the weighing factor to $\omega_{QCSP} = 10$.

III) For the sake of completeness, I also explore an intermediate weighting factor of $\omega_{QCSP} = 5$.

I note, though, that this is still a theoretical framework of development. The weighting factor will be influence in a "real" case also by factors such as the quality and

completeness of the experimental data, the appropriateness of the model for back calculating CSP, etc. These and their effects on the weight I will evaluate and discuss in section 5.7

5.6 Docking with computed CSP reference data

Table 5.3 shows the percentage of docking successes for the 70 complexes from the Astex diverse data set, using the hybrid scoring function and considering each of the weighting factors proposed above. Again, CSP of HN protons computed for the native ligand configurations were used as reference to determine E_{QCSP}. I considered docking results successful when the largest cluster of poses contains at least one solution with an rsmd to the native ligand pose of < 2.0Å. Compared to DrugScore-only docking (71% successful dockings) a large improvement in the docking accuracy is already achieved when $\omega_{QCSP} = 3$ is used for weighting the E_{QCSP} contribution (87% successful dockings). Thus, for the additional 11 cases that can now be docked successfully, the "inverse" E_{DS} gap between decoys and native-like solutions is sufficiently small so that a small E_{QCSP} contributions suffices to compensate for it. All cases are recovered when $\omega_{QCSP} = 10$ is applied. Importantly, cases initially successfully docked with DrugScore-only did no suffer deterioration. Thus, the scoring optima coincide for both the E_{DS} and E_{QCSP} contributions in these cases. Finally, in the last column I have collected the number of cases that would have been successful under a more stringent RMSD cut-off of ≤ 1.0 Å. Such analysis is interesting to assess whether introducing CSP data also increases the native-likeness of the generated solutions. As discussed, such high-quality structures are a pre-requisite for subsequently performing more sophisticated binding energy calculations (Velec, et al., 2005; Bertini, et al., 2007). The impact of CSP in this respect is even higher than considering the standard success criterion. At the lowest weight $\omega_{QCSP} = 3$ the number of cases where solutions in the largest cluster had an RMSD ≤ 1.0 Å to the native structure increased a 48% with respect to the DrugScore-only case, which indirectly shows that the sampling is more concentrated around native-like structures.

Table 5.3: Docking results for the 70 complexes of the Astex diverse set containing ligands with aromatic rings as a function of the weighting factor ω_{QCSP}.

	Successful[1]	Unsuccessful, native-like[2]	Unsuccessful, no native-like[2]	L.C. ≤ 1.0 Å[3]
No QCSP	50 (71%)	13 (19%)	7 (10%)	33 (47%)
$\omega_{QCSP} = 3$	61 (87%)	8 (11%)	1 (2%)	49 (70%)
$\omega_{QCSP} = 5$	65 (93%)	5 (7%)	0	58 (83%)
$\omega_{QCSP} = 10$	70 (100%)	0	0	60 (86%)

[1]A case is considered successful when the largest cluster's average RMSD to the native structure is < 2.0Å.

[2]For the unsuccessful cases, I also distinguish those that at least generate a native-like solution from those that do not.

[3]Number of cases where the largest cluster's average RMSD to the native structure is $\leq$ 1.0Å.

5.7 Docking with experimental CSP reference data

For the above results, I used computed HN CSP from ligand-bound complexes. That way, the problem is simpler than a real-case scenario where other factors such as conformational changes of the protein upon complex formation or other experimental uncertainties play also a role. Additionally, in an experimental setting, it is not always possible to observe and assign all HN CSP in the binding site region, which imposes an additional difficulty when trying to match theoretical patterns of interaction. In a sense, having neglected all those circumstances contributed to achieving the reported high docking success rates when using "theoretical" CSP information. The purpose of this section is then to study the transferability of the weighting schemes developed above for theoretical data and analyze in finer detail the behavior of QCSPScore "real-case scenario", testing the assumptions made so far.

To that end, and as announced in section 5.1, I used three protein-ligand complex for which there is experimental CSP data available (PDB codes: 1ecv, 1a9u and 1ydr). Results of standard re-docking with DrugScore are presented in Figure 5.6.
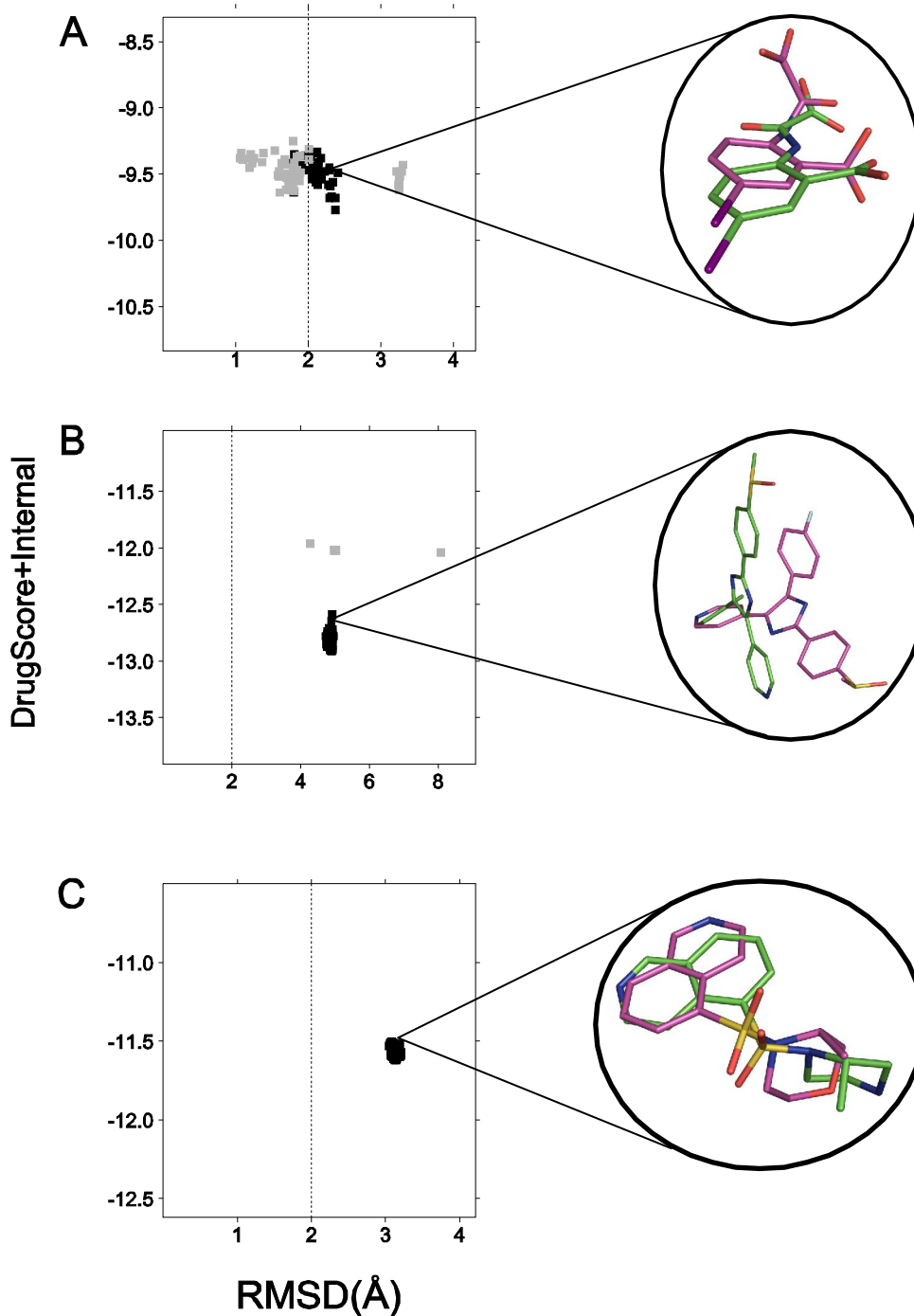
Figure 5.6 DrugScore results for 1ecv (A), 1a9u (B) and 1ydr (C). The largest cluster is represented by black dots. The dotted line divides native-like solutions from non-native-like. The reference conformation from the crystal structure is shown in green.

Out of these three cases, only for 1ecv standard DrugScore re-docking generated native-like docking solutions. Additionally, in all cases DrugScore solutions are characterized by a tendency to maximally bury the ligand in the binding site, which is in agreement with the known fact that this scoring function lacks contributions from configurational

and vibrational entropy changes upon binding. Therefore, maximizing protein-ligand contacts improves the score.

The results using QCSPScore, that is, introducing experimental CSP for each of the proposed weighting regimes, are collected in Table 5.4.

Table 5.4 Docking results using the hybrid scoring scheme, for each of the tested weighting regimes

Weight[2]	LC Size[3]	Rmsd[4]	τ[5]	% N. L.[6]
1ecv				
0	37	2.13	0.10	57
3	35	1.92	-0.07	60
5	35	3.26	0.01	50
10	53	3.27	0.01	39
1a9u				
0	96	4.86	0.00	0
3	77	4.80	0.11	0
5	73	4.79	0.11	0
10	45	4.99	0.19	5
1ydr				
0	100	3.12	-0.16	0
3	76	3.08	-0.04	18
5	48	6.67	0.18	0
10	51	10.11	0.24	0

[1] Results are presented for the proposed weighting factors $\omega_{QCSP} = \{3,5,10\}$ in addition to DrugScore-only docking.
[2] The weighting factor ω_{QCSP}. DrugScore-only docking corresponds to weight 0.
[3] Size of the largest cluster of solutions.
[4] Average RMSD of the poses of the largest cluster with respect to the native crystal structure. In Å.
[5] Average Kendall's correlation coefficient of the poses in the largest cluster
[6] Percentage of native-like solutions among the whole pool of 100 generated solutions.

The first and most striking observation is that only in one case, 1ecv with a $\omega_{QCSP} = 3$ I obtained better RMSD average of the largest cluster using CSP than when using DrugScore only. In all other cases, results were comparable to the DrugScore only case or worst, which is in sharp contrast with the results obtained for the simulated data, where eventually all complexes were correctly docked. Clearly, real data poses significantly different challenges than simulated and in what follows I will discuss the sources and interpretation of these discrepancies.

A remarkable difference between theoretical and real cases is the average Kendall's correlation values that native-like solutions obtain in both scenarios. Whereas in the theoretical cases these values concentrate around 0.8, in the real data cases studied do

not go beyond 0.2 at the highest tested $\omega_{QCSP} = 10$. Since the theoretical analysis proved that scanning up to that value of $\omega_{QCSP} = 10$ is enough for the CSP contribution to effectively drive the docking, an inappropriate weighting scheme can be ruled out as the cause for lack of success. What remains is: a) inaccuracies in the model for back-calculating CSP and b) the influence of incomplete/asymmetrically distributed set of observed CSP, since in the theoretical case all HN in the binding site were considered but these cannot be always observed/assigned in an experimental scenario.

A very illustrative example of the influence of incomplete/asymmetrical assignments is the case of 1ydr. At $\omega_{QCSP} = 10$, 1ydr resulted in solutions with an RMSD to the native structure >10Å, considerably worst than what DrugScore only, ~ 3 Å due to a change in the orientation of the quinoline ring (Figure 5.6 C). In Figure 5.7 the native structure of 1ydr is represented together with one of the wrong solutions at 10 Å RMSD and the assigned CSP. The aforementioned asymmetry in the assignment of CSP is evident from the picture. The wrong solution corresponds to the next most buried one after the original decoy and the native structure, i.e. the following in decreasing order of DrugScore terms, but the best explored in terms of E_{QCSP}. Indeed, the wrong ligand orientation, given the lack of restrains for example from Asn-171 or Glu-170 reproduces the CSP on the top right of the picture without interfering with other CSPs.

Figure 5.7 Binding site area of the native structure of 1ydr complex (ligand conformation green) and a wrong solution generated by QCSPScore at $\omega_{QCSP} = 10$ (ligand conformation in yellow). Red spheres represent negative CSP and blue spheres positive CSP, cantered at HN and scaled according to their absolute value. Green spheres are observed CS that do not change upon ligand binding. The experimental data distributes asymmetrically: most of the observed CSP correspond to the upper part of the figure and in the lower part, residues for which no experimental data was available are labelled.

With respect to the inaccuracies of the model, the effect on docking is apparent by means of an increase of indetermination. I.e., the restraints that CSP introduce, given that the errors in the calculations of specific HN are bigger than their relative differences, allow for an exchange of observations: an electrostatic effect can be compensated by a ring-current effect and vice versa. An example is what one can observe in the case of 1a9u. Assignments here are more complete than for the aforementioned 1ydr case. However, four aromatic rings give enough chance/indetermination for the molecule to accommodate better alternative orientations than the native one. Along this line, I considered alternatively to the full model for back calculating CSP, focus only on the ring current effects. Ring current effects are known to be the largest in proportion and more importantly define a specific pattern (McCoy and Wyss, 2000). The hypothesis was that the electrostatic contribution, the way is implemented here and considering how far apart from each other HNs are in a protein is more a source of noise than valuable information. For example, a very tight protein-ligand electrostatic interaction, real or sample during a docking simulation, produces a

strong CSP. This strong CSP can be equal in magnitude to the one produced by an aromatic ring, and if the aromatic ring is not deeply bound in the protein and/or no sufficient CSP around it have been observed, ring-current and electrostatic effects can be interchanged. It is known that including electrostatics improves the global correlation between experimental and back-calculated CSP (Osapay and Case, 1991). However, this limited global improvement does not negate the fact that a model consisting of the effect of ring-currents only accounts for the largest part of that correlation. In what follows, I analyze the effect of neglecting electrostatics.

5.7.1 Improved accuracy by neglecting electrostatics

Table 5.5 summarizes the docking results with the hybrid scoring function, but now considering only ring-current effects. The results confirm the hypothesis that targeting only "well-defined" interaction patterns such as the ring-current effects, despite theoretically losing accuracy in the total CSP prediction, is at this time a better strategy.

Table 5.5: Summary of docking results for three complexes where experimental CSP reference data was available

Weight[2]	LC Size[3]	Rmsd[4]	τ[5]	% N. L.[6]
1ecv				
0	37	2.13	0.10	57
3	60	1.84	0.13	70
5	37	1.80	0.14	78
10	51	1.36	0.34	87
1a9u				
0	96	4.86	0.00	0
3	76	4.80	0.06	1
5	59	4.80	0.07	1
10	42	1.13	0.25	42
1ydr				
0	100	3.12	-0.16	0
3	72	0.75	0.07	72
5	61	0.81	0.07	61
10	56	10.10	0.24	0

[1] Results are presented for the proposed weighting factors $\omega_{QCSP} = \{3,5,10\}$ in addition to DrugScore-only docking.

[2] The weighting factor ω_{QCSP}. DrugScore-only docking corresponds to weight 0.

[3] Size of the largest cluster of solutions.

[4] Average RMSD of the poses of the largest cluster with respect to the native crystal structure. In Å.

[5] Average Kendall's correlation coefficient of the poses in the largest cluster

[6] Percentage of native-like solutions among the whole pool of 100 generated solutions.

Now in all three cases native-like solutions were generated and in some of the weighting schemes these accounted for the largest cluster. Docking results with and without using CSP and the effects of them are also depicted in Figure 5.8.
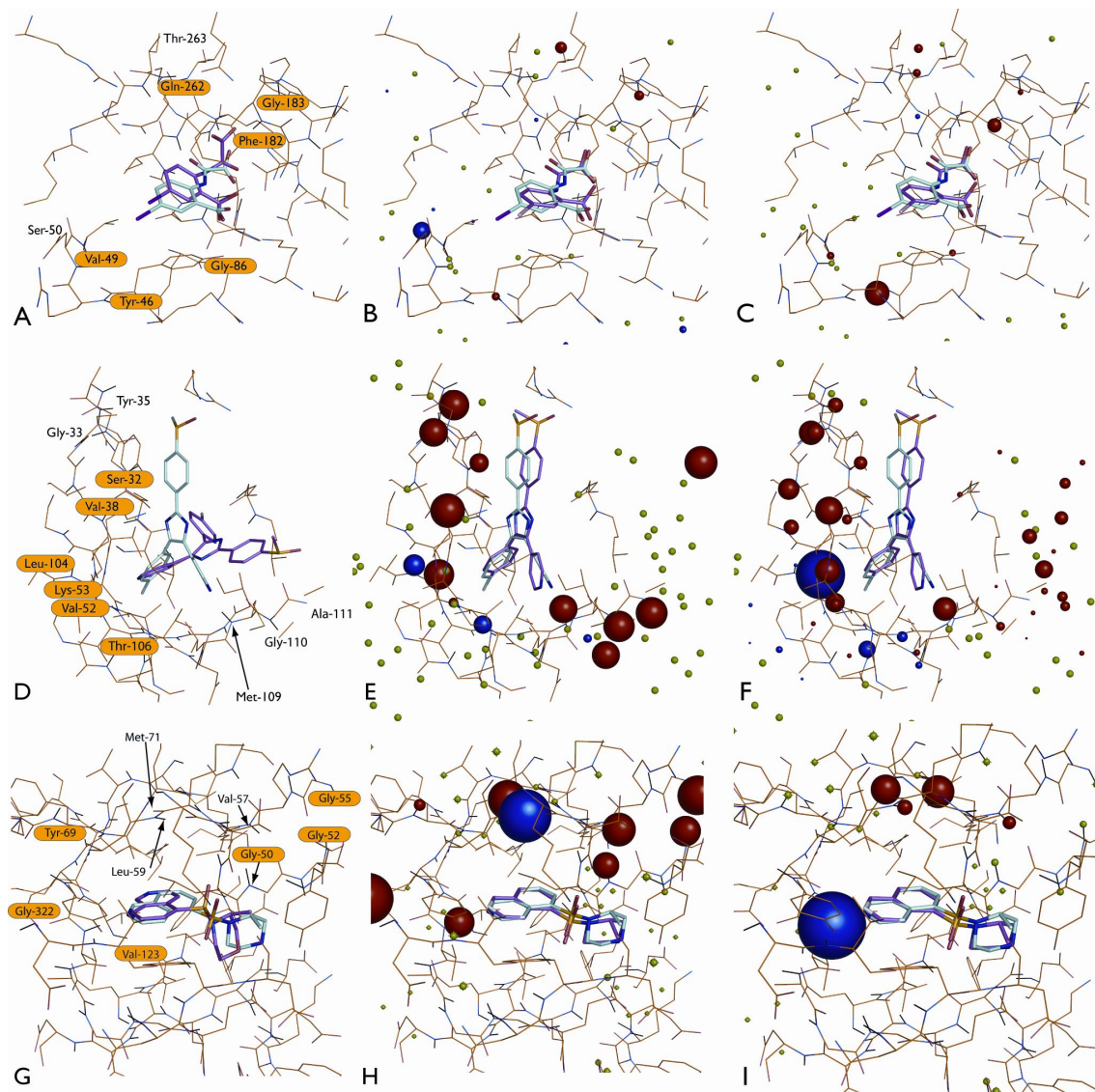
Figure 5.8 Comparison of docking solutions (magenta) with the crystal structure (cyan). In the first column (A, D, G), results using only DrugScore for scoring are presented; the second and third columns (B, C, E, F, H, I) display the best solutions obtained with QCSP-steered docking. Results for complex 1ecv are given in panels A, B, C; for 1a9u in panels D, E, F; for 1ydr in panels G, H, I. In the second column experimental CSP are mapped to the protein whereas in the third column calculated CSP corresponding to the docked pose using the ring currents-only model are displayed. Blue balls represent upfield shifts (positive CSP), red ones downfield shifts (negative CSP), and green balls are used to mark CSP that have been assigned but do not experience any perturbation. The size of the blue and red balls is proportional to the magnitude of the CSP.

With respect to the question whether an optimal weight exists, one can recognize that the docking experiments with computed CSP reference data already suggested that $\omega_{QCSP} = 10$ is an appropriate weighting factor in general. This scheme was appropriate also for 1ecv and 1a9u, which largest clusters had respectively 1.36Å and 1.13Å RMSD on average to the native structure. In contrast, for 1ydr, successful docking with RMSD ~ 0.8Å required lower ω_{QCSP} values of 3 or 5. The requirement to use lower ω_{QCSP} values points to the problem of over-emphasizing the influence of imperfect experimental CSP reference data and/or neglecting shortcomings in the computational model used for back-calculating CSP from a given ligand pose. I thus set out to investigate in more detail factors that influence the performance of the hybrid scoring scheme in the case of experimental CSP reference data. I identify as the most important determinants hydrogen bond formation, extent and distribution of the CSP assignment, and flexibility of the target. The latter also includes structural differences of the NMR and X-ray determined protein structures.

5.7.2 The hydrogen-bond effect

In both the 1a9u and 1ydr cases, the ligand forms a hydrogen bond to an HN proton in the protein for which experimental CSP were available. The effect of a hydrogen bond formation cannot be modeled accurately with current empirical methods (Osapay and Case, 1991; Wishart and Case, 2001; Parker, et al., 2006; Moon and Case, 2007). Accordingly, following preliminary tests, I decided to omit this effect when back-calculating CSP. In the case of 1a9u, the presence of the hydrogen bond formed between the pyridine N and HN of Met-109 did not deteriorate the final docking result (Table 5.5). This is because back-calculating the CSP of HN of Met-109 is successful (Figure 5.8E, F) even with the ring current-only model, if the pyridine ring of the ligand adopts a slightly tilted and displaced conformation compared to the native structure. A similar observation was already described by McCoy and Wyss (McCoy and Wyss, 2002) who pointed out that the typical downfield shift of a proton in a hydrogen bond can also be generated by a properly oriented aromatic ring in the proton's vicinity. I note, however, that such a fortuitous effect cannot occur in the case of a narrower crevice, which does not allow ring tilting and displacement. That situation compromises the docking accuracy, as demonstrated for 1ydr. Here, the ligand is forming a hydrogen bond with the HN proton of Val-123 (Figure 5.8G). At the same time it is tightly

constrained in the binding pocket. Thus, no evasive movements of the quinoline ring are possible allowing for a compensation of the missing hydrogen bond term when computing the CSP. As a result, a low E_{QCSP} score is obtained even for native-like solutions. Following the idea that no CSP contribution at all must be better than a wrong contribution, I repeated the docking without considering HN of Val-123. Results for 1a9u and 1ydr without their respective HN known to be involved in a hydrogen bond are collected in Table 5.6. With an average E_{QCSP} score improving modestly but significantly from -0.07 to -0.12 for native-like solutions, a successful docking was now achieved with $\omega_{QCSP} = 10$. Thus, all three complexes with experimental CSP reference data were successfully docked using a uniform weighting factor $\omega_{QCSP} = 10$.

Table 5.6 Docking results using CSP and neglecting HN involved in hydrogen bonds.

Weight[2]	LC Size[3]	Rmsd[4]	τ[5]	% N. L.[6]
1a9u – no Hbond				
0	96	4.86	0.00	0
3	81	4.83	0.04	0
5	70	4.82	0.04	1
10	67	1.06	0.22	67
1ydr – no Hbond				
0	100	3.12	-0.16	0
3	80	0.76	0.11	80
5	98	0.81	0.12	98
10	54	0.84	0.12	54

[1] Results are presented for the proposed weighting factors $\omega_{QCSP} = \{3,5,10\}$ in addition to DrugScore-only docking.
[2] The weighting factor ω_{QCSP}. DrugScore-only docking corresponds to weight 0.
[3] Size of the largest cluster of solutions.
[4] Average RMSD of the poses of the largest cluster with respect to the native crystal structure. In Å.
[5] Average Kendall's correlation coefficient of the poses in the largest cluster.
[6] Percentage of native-like solutions among the whole pool of 100 generated solutions

The limitation in describing CSP due to hydrogen bond formation is well-known, and more work is needed on the theoretical side to improve on this. In the meantime, one possibility to circumvent this limitation is omitting CSP induced by hydrogen bond formation from the hybrid scoring scheme. These CSP can be identified by visually inspecting the experimental CSP pattern. Typically, a hydrogen bond is characterized by a large downfield perturbation, which can be of a similar magnitude as the one due to a nearby aromatic ring. However, a hydrogen bond-induced CSP usually appears isolated

and does not show the typical pattern of CSP on neighboring atoms that an aromatic ring would induce.

5.7.3 Influence of the extent and spatial distribution of CSP assignment

The extent of CSP assignment of HN protons in the binding site area of the three complexes is 33% and 37% for 1ecv and 1ydr, respectively, and 75% for 1a9u. Despite the comparable extent of HN CSP assignments, the highest E_{QCSP} values found for native-like solutions of the first two complexes differ considerably (1ecv: -0.34; 1ydr: -0.11, omitting the HN involved in a hydrogen bond). This difference can be explained by the spatial distribution of the CSP assignment. In the case of 1ecv, the CSP distribute uniformly around the binding site and so capture the traits of the ring-current pattern (Figure 5.8B). On the contrary, for 1ydr, the assignment excludes a large part of the adenosine-binding pocket surface (Langer, et al., 2004), which leads to increased minimal E_{QCSP} scores as the ring-current pattern lacks key reference points in space (Figure 5.8H). Thus, in the case of sparse experimental CSP data, it is the distribution of the CSP rather than the amount of data *per se* that determines the success of QCSP-steered docking.

5.7.4 Influence of the target flexibility

In all three studied cases the proteins undergo re-arrangements upon ligand binding, with RMSD of the binding sites between bound and unbound conformations ranging from 0.73Å in the case of 1ecv and 1.0Å in the case of 1a9u to 2.1Å in the case of 1ydr. Protein re-arrangements are challenging for properly back-calculating CSP in a rigid-protein docking scheme because aromatic ring movements in the protein can produce CSP as large as those induced by a ligand. Rearrangements of solvent molecules contribute to CSP, too. At present, QCSPScore assumes that most of the observed effects can be directly attributed to the ligand as a source and that this proportion suffices to orient the ligand in the native-like position. If the ligand is the major source of CSP, E_{QCSP} scores should be approximately -1 for native-like poses, perfect correlation. In turn, an increase of the minimal E_{QCSP} score is expected with increasing protein movement and this is what is observed for the three experimental cases studied

here). Too large protein movements will thus render the model for back-calculating CSP insufficient. Nonetheless, even movements up to 2.1Å RMSD are still tolerated in the case of 1ydr. In general, I note that target flexibility is not a limitation for the hybrid scoring scheme described here. I believe that using CSP information can be very helpful in a fully-flexible protein-ligand docking scenario, because protein refinement against such data has been shown to improve structural quality (Clore and Gronenborn, 1998), and one ought to expect the same for a complex structure.

It is also interesting to note that not only protein movements upon ligand binding can deteriorate the accuracy of CSP computation but also structural differences of protein structures originating from different experimental sources or conditions. This point is illustrated by the HN proton of Lys-53 of 1a9u. The HN proton experiences a downfield shift in experiment, whereas the model implemented in QCSPScore predicts an upfield shift for the native-like solution. The latter corresponds to a position of the proton in or close to the plane of the aromatic ring. In the native structure, it can be seen that the proton is almost co-planar to the ring, too. Accordingly, the computed CSP for this structure also shows an upfield shift. The experimental downfield shift can be reproduced properly if an alternative crystal structure of the same complex (PDB id: 2ewa) is used, which differs slightly in the mutual orientation of the ring and the HN of Lys-53 (Figure 5.9). Again, as in the case of the hydrogen bonds, local disagreements between experimental and computed CSP can be compensated by a complete and/or well-distributed CSP assignment. It is the global CSP pattern then that still allows finding native-like solutions (Figure 5.8E vs. Figure 5.8F).
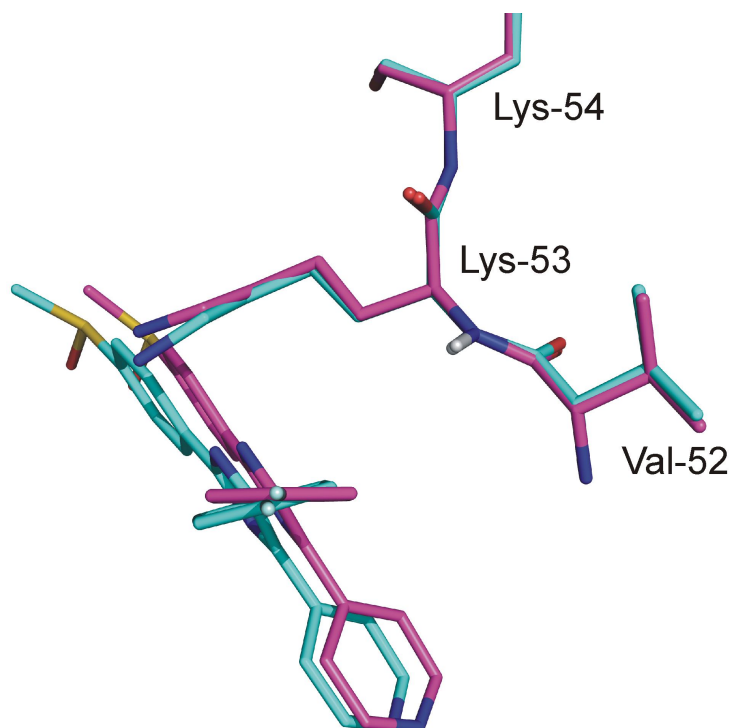
Figure 5.9 Superimposition of p38 complexes 2ewa (model 1) (magenta) and 1a9u (cyan). The angle between the ring plane and a vector from the ring center to the HN of Lys-53 differs by 12° (2ewa: 31°; 1a9u: 23°). This difference is sufficient to predict a downfield shift for 2ewa (CSP: -0.18) and an upfield shift for 1a9u (CSP: 0.07) using the empirical model implemented in QCSPScore.

5.8 Comparison to related methods

In the last years, there has been a growing interest in developing methods capable of harnessing structural information contained in protein CSP upon ligand binding for solving protein-ligand complexes. The methods published so far differ from one another mainly in two aspects: first whether they use CSP quantitatively or quantitatively and, second, whether CSP are employed for pre-/post-filtering (or refinement) or to drive docking. A more thorough review on these methods can be found in the Introduction chapter of this work. Here, a summary of the methods, attending to the aforementioned classification is presented in Figure 5.10.
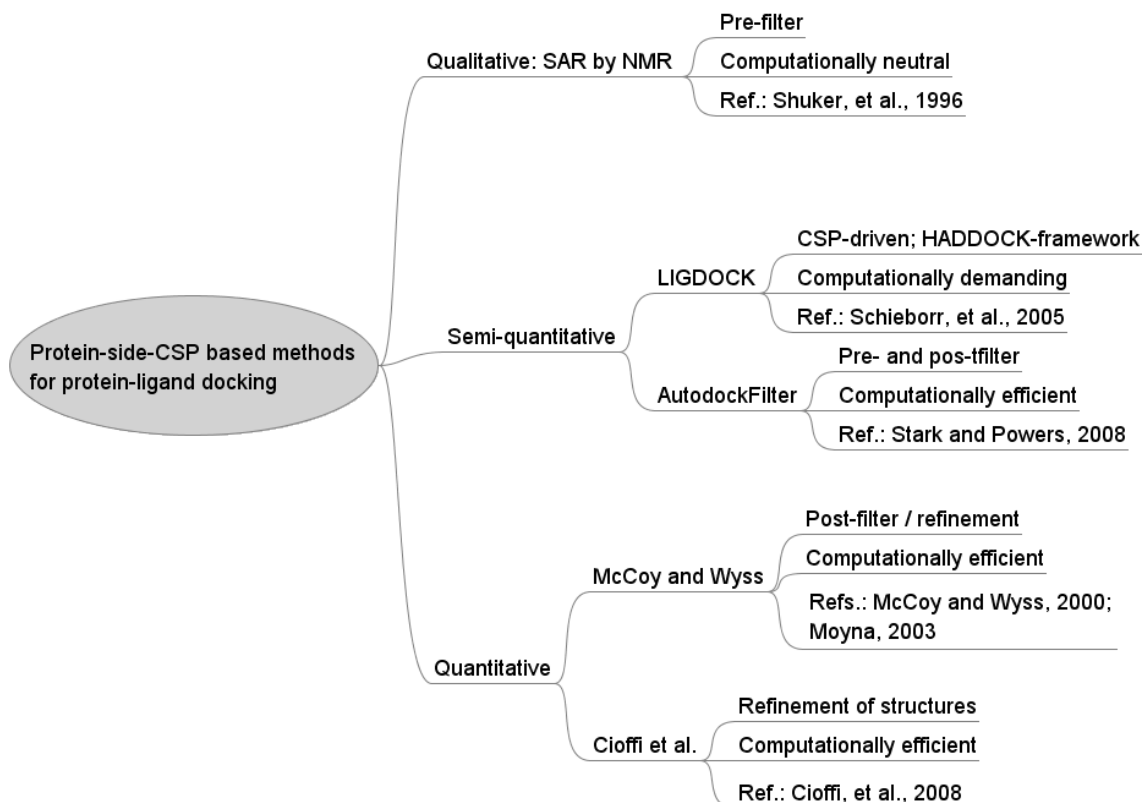
Figure 5.10 Classification of methods that use CSP for protein-ligand docking.

A "qualitative" use of CSP considers exclusively information regarding presence or absence of perturbation in a residue's CS. In "semi-quantitative" methods CSP are only considered in absolute magnitude, thus neglecting the sign of the perturbation and not making a difference upfield or downfield effects. QCSPScore, the method described in this work, belongs to the "quantitative" category. In what follows I compare directly the semi-quantitative use of CSP vs. the quantitative use done in this work and CSP-driven approaches vs. post-filtering strategies.

5.8.1 Semi-quantitative vs. quantitative scoring

Both AutoDockFilter and LIGDOCK share common principles. They rely on the idea of ambiguous restraints which can be defined from the observed CSP. In AutoDockFilter it is assumed that the intensity of observed protein CSP upon ligand binding are linearly distant-dependent to any ligand atom. Thus, the authors have defined a violation energy E_{NMR} as $E_{NMR} = k \sum_{i=1}^{n} (\Delta_{Dist})^2$, where Δ_{Dist} penalizes ligand poses as long as any of the "observed" CSP are not in close contact to at least one ligand atom. I have evaluated the

performance of such a scoring method on the same data set used in the theoretical development of QCSPScore. For that, I considered the subset of the Astex data set that generated both native-like and non-native-like solutions using standard DrugScore, as done in the initial scoring function assessment (see section 5.2, above). As the developers of AutoDockFilter did, I also simulated CSP for these complexes considering a simple linear relationship between residues in the binding site of the protein and the nearest ligand atom. Then, I calculated the corresponding E_{NMR} for each pose in a post-filtering fashion. I considered a very relaxed criterion and accepted as "successful" those cases in which the native-like solutions obtained the lowest E_{NMR}, independently if this lowest score was also shared by a non-native solution.

Simplifying the relationship of observed CSP to a linear distance dependency to the source of the perturbation is a very crude approximation. On the one hand it circumvents the necessity of expensive calculations for predicting CSP, but on the other hand, that simplification can only provide information about the location of the binding site and anchor the ligand to it. Such approaches have a deep impact in protein-protein docking where possible relative orientations of the interacting partners are considerably larger than in the protein-ligand case (van Dijk, et al., 2005a). However, in the protein-ligand docking case, the real problem comes when the ligand is properly placed in the binding site but two or more (symmetrically or not) alternative conformations are plausible, since CSP information used in that fashion does not provide additional clues. Schieborr et al. faced that problem when trying to re-dock 1a9u (Schieborr, et al., 2005). I have seen the very same effect in the cases analyzed here using AutoDockFilter as illustrated in Figure 5.11.
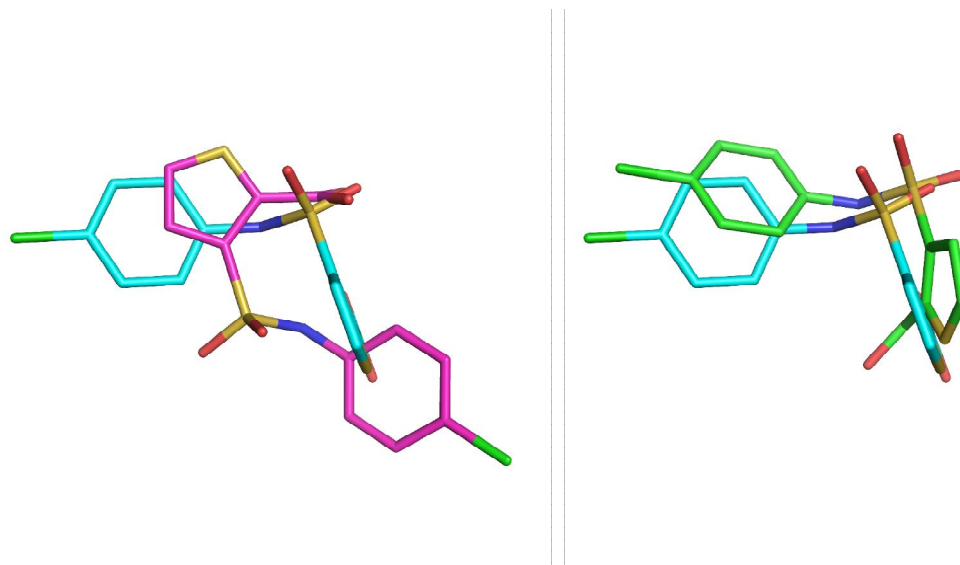
Figure 5.11 A decoy (left) and a native-like solution (right) for 1l2s complex. Both conformations are indistinguishable to AutoDockFilter, whereas Kendall's correlation coefficient for decoy and native-like solutions are 0.13 and 0.70 respectively.

What happened to 1l2s complex is a clear example of the limitations of AutoDockFilter. Whereas for this scoring function both solutions shown in Figure 5.11 (decoy and native-like) are indistinguishable. As a matter of fact, the decoy gets a slightly better solution than the native-like solution. They both generate clearly different CSP patterns, properly captured by the Kendall's correlation based scheme implemented in QCSPScore ($\tau = 0.13$ for the decoy and $\tau = 0.70$ for the native-like pose).

In summary, despite the limitations discussed in the previous section, I believe these results show that quantitative CSP is a remarkable advantage towards successful docking results, compared to qualitative or semi-quantitative use. In fact, if semi-quantitative data is used in a pre-filtering step in order to narrow the binding site definition, all structural data contained in it is already exploited.

5.8.2 CSP-driven vs. post-filtering approaches

The first approaches in which quantitative CSP data was used consisted of post-filtering methodologies (McCoy and Wyss, 2000). Arguably, such approaches are convenient as long as one is confident that the generator of conformations is going to sample and produce native-like solutions for the starting pool. In the case of rigid-ligand docking it is feasible to exhaustively enumerate the space of ligand orientations; however, rigid-

ligand docking is not a common practice anymore, unless one is sure of the bioactive conformation of the ligand. Running a docking program in a standard-fashion, the most likely scenario does not ensure that pre-condition: for the data set analyzed here 9 out of 70 cases (13%), no native-like solution was generated at all. Similarly, Stark and Powers reported that blind AutoDock calculations on 19 complexes resulted in 5 cases not generating any native-like poses (120 dockings runs for each complex) (Stark and Powers, 2008).

But incorporating CSP at docking time imposes the challenge of defining how to weight the experimental information. Schieborr et al. (Schieborr, et al., 2005) resorted to a standard NMR-like methodology, switching on and off the contribution along the simulations, followed by a "selection plot" for scoring poses. The selection plot was necessary for example to evaluate 1ydr poses, where no single criterion, van der Waals interactions or matching experimental data, sufficed. Despite their success, the method incurs in very low efficiency. Cioffi et al. (Cioffi, et al., 2008) seem to acknowledge the aforementioned problems of post-filtering and resort to non-standard use of the docking program: first, only vdW contributions of the scoring function are left on; second, so generated poses are not re-scored but optimized in CSP terms. This also reflects that CSP alone do not suffice to orient protein and ligand starting from a given random orientation. QCSPScore is more efficient in since the whole simulation is done in a single step.

5.8.3 QCSPScore vs. other quantitative approaches

Probably, the most sophisticated method among the so far cited, and also more similar to QCSPScore, is the one by Cioffi et al., being a fundamental difference the functional form of the scoring function. As long as difficulties remain to model more correctly CSP (e.g. with respect to hydrogen formation) a Kendall's correlation based scheme, is preferable, as I have shown here. Unfortunately, the authors did not report cases in which they re-docked protein-ligand complexes and hydrogen-bond effects were present, to properly evaluate the impact of such effect. Additionally, Cioffi et al. used a model for back-calculating CSP which included ring-current effects, electrostatic and other magnetic effects. I have shown that the ring-currents-only model works better in the current setting of crude electrostatic modelling and sparse experimental data. It is

difficult to evaluate differences also in this respect, since they have reported results only for one case. Additionally, it would have been also interesting to see whether their method really outperformed standard docking, without any CSP information, as is the case for QCSPScore.

6   Summary, conclusions and outlook

In this work, I have presented the development of QCSPScore, a new computational method that exploits quantitatively observed chemical shifts perturbations of amide protons on the protein side to steer a protein-ligand docking engine. QCSPScore is implemented within AutoDock, which permits a convenient straightforward use.

QCSPScore has been conceived as a hybrid scoring function that combines linearly DrugScore potentials with a measurement of the agreement between experimentally observed and back-calculated QCSP for a given docking pose. The back-calculation of CSP is done using an empirical model, in order to keep the whole approach efficient. The core of the work has therefore been devoted to study: a) how that agreement between experimental and back-calculated QCSP can be measured and b) determine how to relatively weight QCSP contribution with respect to DrugScore (Gohlke, et al., 2000).

Since large collections of experimental data were not available for developing and testing purposes, I have followed a two-step strategy. First, I considered a comprehensive data set of 70 protein-ligand complexes from the Astex diverse data set (Hartshorn, et al., 2007) for which I computed QCSP data for reference. Second, results obtained in this step were translated and re-evaluated using three protein-ligand complexes for which experimental QCSP data was available.

In the theoretical development part, complexes were re-docked using DrugScore-only as scoring function, which produces correct docking results in 71% of the cases (section 5.2). I used the structures of those correct cases and the incorrect ones (decoys) for establishing an appropriate functional form of the QCSP contribution. Kendall's rank correlation coefficient proved to be appropriate (section 5.3). Using the rank correlation, it is not necessary to account for the degree of protein saturation by a bound ligand, which may vary from complex to complex and influence the magnitude of CSP. In addition, the rank correlation is a robust statistic measure, which compares favorably to scoring schemes based on minimizing squared deviations between experimental and computed CSP. I have ascertained that these latter scoring schemes show a high sensitivity to the presence of outliers in the back-calculated CSP data.

To establishing the weight of the QCSP contribution for linearly combining it with DrugScore I analyzed again DrugScore results to assess *energy gaps* between decoys and native-like docking solutions (section 5.4). For the data set under study, the maximum energy gap in DrugScore terms was rather low, 1.31 DrugScore units. Kendall's rank correlation values for decoys spread around 0.29 and around 0.82 for native-like solutions, which suggested a weight scan between 3 and 10 (section 5.5). I note that in these ideal cases no inaccuracies occur upon back-calculating the CSP and all amide protons in the binding site have a CSP value assigned, which permits isolating the theoretical influence of the CSP contribution on the docking outcome. At the smallest weighting factor tested ($\omega_{QCSP} = 3$) a docking success rate of 87% was achieved, already 16% higher than DrugScore-only scoring. This rate improved to 100% at the largest weighting factor tested ($\omega_{QCSP} = 10$). Additionally, results generated with QCSPScore have a higher native-likeness than the cases of DrugScore-only docking. This is also corroborated using as a more stringent criterion of success an RMSD to the native structure of ≤ 1.0 Å: 47% of the DrugScore-only cases would have been successful vs. 86% of the cases driven by QCSP at $\omega_{QCSP} = 10$ (section 5.6). In total, this part of the study revealed that, despite the QCSP contribution being a "hard" scoring term, global optimization performs satisfactorily on the combined E_{DS}/E_{QCSP} docking energy landscape. In particular, it increases the sampling of native-like conformations.

Applying the hybrid scoring scheme with a weighting factor of 10 to the three cases for which experimental CSP reference data was available, resulted in successful docking if HN involved in hydrogen bonds with the ligand were discarded from the data pool, (section 5.7). This is a significant improvement with respect to DrugScore-only docking, which was unsuccessful in all three cases. In addition, only in the hybrid docking scheme, native-like solutions were generated for complexes 1a9u and 1ydr. This justifies the importance of including the experimental information at docking time, and not just as a pre-filter or a post-filter as has been typically done until now.

Assessing QCSPScore with real experimental data gave the opportunity to gain some insight in possible limitations of the approach and how to overcome them. To start with, HN CSP involved in hydrogen bonds, when neglected, resulted in better performance of the calculation. This was expected since the empirical model implemented in

QCSPScore for back-calculating CSP does not explicitly model this effect. However, an *a priori* surprising result was that neglecting also the electrostatic contribution included in the model was necessary to obtain successful results. Since HN CSP in an experimental setting might distribute sparsely and asymmetrically, the appearance of clear *patterns* marking the appearance of the ligand in the binding site is hindered. For this reason, it can be rationalized that using a ring-current-only model for back-calculating produced better results. Ring-current effects, when present, are the most significant components of the total observed QCSP. In addition, they produce a characteristic pattern of magnetic perturbation in their vicinity for which they can be used for effectively orienting ligands in their binding sites. When analyzing additional factors that influence the docking success, it was encouraging to see that binding site movements of up to 2 Å did not deteriorate the docking success when considering CSP information. Furthermore, an extent of CSP assignments of HN protons in the binding site region below 40% can still be tolerated, if the CSP are rather uniformly distributed.

The limited experimental data set used in the development of QCSPScore, despite the successful results, does not permit making a definitive assessment of QCSPScore's merits. However, it permitted discovering some limitations, from a theoretical and experimental point of view, that need being addressed in the future in order to refine and expand its applicability.

The low accuracy in the prediction of protein HN only has a limited impact on the success of the approach, compared to incomplete and irregular patterns of HN CSP assignments. However, dealing with the specific case of hydrogen-bond formation, better understanding and modelling electrostatic effects on CSP would be desirable. Likewise, an extension to account for protein flexibility is needed. With respect to the problem of HN involved in hydrogen-bonds with the ligand, Moon and Case (Moon and Case, 2007) recently proposed an extended empirical model. It remains to be investigated how much the efficiency in the calculation is compromised by such an alternative. The same is true for more sophisticated treatments of the electrostatics. So-called Generalized-Born models have become a popular alternative to more demanding explicit solvent models in recent years. The question is whether any of such models could be efficiently implemented into QCSPScore, and if the gain in accuracy will really compensate the computational effort in this respect. More accurate methods for

back-calculating CSP, in this case on the ligand site, have been recently reported. NMRScore(Wang, et al., 2007) uses quantum-mechanics for this task, which is useful in a structural refinement effort, but not in a medium-throughput docking campaign.

Relying on ring-currents effects only effectively limits the number of small-molecules that can be studied. Yet, more than 95% of the compounds in the MDL Drug Data Report database (http://symyx.com) contain an aromatic ring and would thus be amenable. Finally, flexibility of the protein has not been considered. However, this is not a limitation of QCSPScore itself, but from the underlying docking engine.

From an experimental point of view, relying on CSP permits studying complexes for which NOE cannot be seen, however observing and assigning CSP is still a demanding task. For this reason, approaches like QCSPScore will benefit from future developments in high-throughput assignment of proton chemical shifts and peak tracking methodologies (Fukui and Chen, 2007).

Almost ten years ago McCoy and Wyss pioneered the use of QCSP for protein-ligand complex elucidation. They wrote: "We anticipate that ligand-induced chemical shifts perturbations can be used as restraints in structure calculations and can be energy minimized with NOEs, van der Waals and electrostatics to give more accurate protein/ligand structures" (McCoy and Wyss, 2000). Since then, increasing attention has been paid to ligand induced CSP. A very recent work by Cioffi et al. (Cioffi, et al., 2008) describes how QCSP can be used for optimizing protein-ligand geometries, but, to the best of my knowledge, the present study is the first work in which QCSP are incorporated into a docking scoring function for effectively driving the simulation from the beginning. The relevance of the direction taken in this work has been already stated. First, simulations where CSP were used only qualitatively or semi-quantitatively do not effectively exploit the experimental data to orient the ligand in the binding site; rather, they only anchor it to that area. Second, there is a risk for quantitative approaches following a post-filtering strategy to work on a pool of solutions where a native-like solution is not present.

I believe that the approaches like QCSPScore can have a relevant impact when dealing with challenging protein-ligand complexes, which nowadays coincide to represent appealing avenues to finding new therapeutic applications. I am referring to fragment-

based drug-design (Congreve, et al., 2008), where small fragments are known to be a challenge for classical scoring functions (Hubbard, et al., 2007; Chen and Shoichet, 2009), and protein-protein interfaces (Arkin and Wells, 2004), which typically present rather flat binding sites, difficult to address for current docking programs (Gonzalez-Ruiz and Gohlke, 2006).

7   Zusammenfassung

Das Docken von Liganden an Proteine ist ein in der Pharmaforschung routinemäßig angewendetes rechnerisches Verfahren. Eine beträchtliche Anzahl publizierter Studien zeigt den erfolgreichen Einsatz von Docking beim virtuellen Screening nach Wirkstoffkandidaten (Kitchen, et al., 2004; Cavasotto and Singh, 2008). Auch bei der Optimierung potentieller Wirkstoffe hat sich der Einsatz von Dockingverfahren bewährt. Als ein bekanntes erfolgreiches Beispiel ist der inzwischen kommerzialisierte Wirkstoff Imatinib (Gleevec) zu nennen (Capdeville, et al., 2002). Trotzdem besteht auch weiterhin Bedarf und die Notwendigkeit, Dockingverfahren weiter zu verbessern.

Das Docking erfordert einerseits die Vorhersage der Konfiguration oder Geometrie des Protein-Ligand-Komplexes, andererseits die akkurate Bewertung oder Abschätzung der Wechselwirkungsenergie zwischen den Bindungspartnern, das so genannte Scoring. Was die erste Anforderung betrifft, so sind Erfolgsraten von bis zu 80% berichtet worden (Warren, et al., 2006). Für die zweite, schwierigere Anforderung liegen die Erfolgsraten bei der Unterscheidung einer nativen von einer nicht-nativen Komplexstruktur nur zwischen 10 und 70%. Bei der Reihung einer Serie von Liganden gemäß ihrer Wechselwirkungsenergie ist die Erfolgsrate sogar noch geringer (Warren, et al., 2006). Aufgrund dieser Tatsachen wird allgemein davon ausgegangen, dass das Problem der Geometrievorhersage weitestgehend gelöst sei, das Problem der Bewertung dagegen nicht. Velec und Kollegen haben kürzlich gezeigt, dass gute Geometrievorhersagen Voraussetzung für die Verbesserung von Bewertungsfunktionen sind (Velec, et al., 2005).

Eine Tendenz bei der Verbesserung von Bewertungsfunktionen ist die Anpassung der Funktion an ein bestimmtes System, d.h. die Entwicklung so genannter maßgeschneiderter Bewertungsfunktionen. Dabei werden verfügbare Daten für eine Reihe bereits bekannter Liganden genutzt (Fradera and Mestres, 2004; Jansen and Martin, 2004; Radestock, et al., 2005). Alternativ können auch experimentelle Daten zur Verbesserung des Dockings benutzt werden, die im Rahmen der Untersuchung eines bestimmten Systems neu erzeugt wurden (van Dijk, et al., 2005a). Das Kombinieren theoretischer Dockingmethoden mit neu erzeugten experimentellen Daten hat die Untersuchung biologischer Systeme ermöglicht, die bislang als zu schwierig für eine

Untersuchung mit Dockingverfahren galten. Bemerkenswerterweise haben diese Ansätze der Erforschung und Entwicklung von Protein-Protein-Dockingmethoden neuen Auftrieb gegeben (van Dijk, et al., 2005a). Aber auch im Bereich des Protein-Ligand-Dockings werden solche Methoden angewendet, insbesondere in Verbindung mit Kernmagnetresonanz (NMR)-spektroskopischen Daten (Carlomagno, 2005; Powers, 2007).

NMR-spektroskopischen Untersuchungen liefern Informationen über die Struktur eines biomolekularen Komplexes. Chemischen Verschiebungen (CS) kommt dabei eine Schlüsselrolle bei der Zuweisung von NMR-Spektren zu. Für die eigentliche Strukturaufklärung wurden sie bislang aber als weniger brauchbar angesehen (Szilagyi, 1995), da ihre Interpretation hinsichtlich der Struktur schwierig ist. Seit einigen Jahren finden CS aber immer mehr Anwendung auch bei der Strukturaufklärung (Wishart and Case, 2001; Hunter, et al., 2005). Zudem hat sich die Beobachtung von Veränderungen der CS (CSP) bezüglich eines Referenzzustands zur Untersuchung biomolekularer Wechselwirkungen als sehr interessant herausgestellt. Die Messung von Veränderungen der CS sind auch Grundlage eines unter dem Namen „SAR by NMR" bekannt gewordenen fragmentbasierten Ansatzes zur Entwicklung von Liganden (Shuker, et al., 1996). Schwierigkeiten bei der Interpretation und Modellierung von CSP können umgangen werden, indem nur qualitative Aussagen getroffen werden. Dieser Ansatz liefert aber keine Informationen über die Konfiguration des Protein-Ligand-Komplexes, die von einem Dockingverfahren zur Geometrievorhersage genutzt werden könnten.

Ich habe im Rahmen dieser Arbeit einen neuen rechnerischen Ansatz entwickelt, der quantitative CSP von Amidprotonen auf der Proteinseite zur Verbesserung eines Protein-Ligand-Dockingverfahrens benutzt. Die Methode nennt sich QCSPScore. Sie wurde in das AutoDock Softwarepaket implementiert, was eine komfortable Nutzung erlaubt.

QCSPScore wurde als eine Hybridbewertungsfunktion entwickelt, die DrugScore-Potentiale (Gohlke, et al., 2000) mit einem Maß für die Übereinstimmung zwischen experimentell gemessenen und vorhergesagten quantitativen CSP für eine Docking-Pose linear kombiniert. Die Vorhersage der CSP erfolgt dabei mit Hilfe eines empirischen Modells, um die Effizienz des Ansatzes zu gewährleisten. Die vorliegende Arbeit konzentriert sich auf die Behandlung zweier Fragen. Erstens, wie die

Übereinstimung zwischen experimentell gemessenen und vorhergesagten QCSP erfolgen kann, und, zweitens, wie der QCSP-Beitrag gegenüber DrugScore-Potentialen gewichtet werden muss.

Da für die Entwicklung und Testung der Methode keine großen Datenmengen zur Verfügung standen, habe ich eine zweischrittige Strategie verfolgt. In einem ersten Schritt habe ich mit 70 Protein-Ligand-Komplexen aus dem „Astex diverse" Datensatz gearbeitet, für die ich CSP vorhergesagt habe. In einem zweiten Schritt habe ich die Ergebnisse der Untersuchung dreier Protein-Ligand-Komplexe ausgewertet, für die experimentelle Daten vorlagen.

Während der Entwicklung der Methode wurden die untersuchten Protein-Ligand-Komplexe auch mit herkömmlichen DrugScore-Potentialen bewertet. Dies führte in 71% der Fälle zu korrekt vorhergesagten Geometrien (Abschnitt 5.2). Diese nativ-ähnlich vorhergesagten Komplexe habe ich anschließend zusammen mit den falsch vorhergesagten Komplexen zur Ableitung der funktionellen Form des QCSP-Beitrags zur Bewertungsfunktion benutzt. Dabei stellte sich der Rangkorrelationskoeffizient nach Kendall als nützlich heraus (Abschnitt 5.3). Durch die Verwendung dieses Korrelationskoeffizienten muss die Sättigung der Signale auf Proteinseite nicht mehr berücksichtigt werden, die bei verschiedenen Protein-Ligand-Komplexen unterschiedlich sein und die Werte der CSP beeinflussen kann. Außerdem ist der Rangkorrelationskoeffizient ein robustes Maß, verglichen mit Bewertungschemata, die auf der Minimierung quadratischer Differenzen zwischen experimentellen und berechneten CSP beruhen. Meine Untersuchungen haben ergeben, dass die Sensitivität dieser Bewertungsfunktionen gegenüber Ausreißern in vorhergesagten Daten ein Problem darstellt.

Um die Gewichtung des QCSP-Beitrags in der linearen Kombination mit DrugScore-Potentialen zu bestimmen, habe ich die „Energielücke" (*energy gap*) zwischen den mit herkömmlichen DrugScore-Potentialen nativ-ähnlich bzw. nicht korrekt vorhergesagten Komplexen untersucht (Abschnitt 5.4). Für den untersuchten Datensatz war die maximale Energielücke für DrugScore mit 1,31 DrugScore-Einheiten recht niedrig. Werte für den Rangkorrelationskoeffizienten nach Kendall lagen dagegen zwischen 0,29 für die nicht-korrekt vorhergesagten Komplexe und 0,82 für die native-änlich vorhergesagten Komplexe, was auf einen Gewichtungsfaktor zwischen drei und zehn

hindeutete (Abschnitt 5.5). Die Tatsache, dass für diese idealen Fälle keine Ungenauigkeiten durch die Vorhersage von CSP auftreten und alle Amidprotonen zugewiesen werden können, erlaubt die alleinige Beobachtung des Einflusses der CSP auf das Dockingergebnis. Mit dem niedrigsten getesteten Gewichtungsfaktor ($\omega_{QCSP} = 3$) wurde eine Erfolgsrate beim Docking von 87% erzielt. Das bedeutet eine Verbesserung um 16% im Vergleich zum DrugScore-Ergebnis. Die Erfolgsrate konnte sogar auf 100% gesteigert werden, wenn der größte getestete Gewichtungsfaktor verwendet wurde ($\omega_{QCSP} = 10$). Hinzu kommt, dass die mit QCSPScore vorhergesagten Geometrien der nativen Komplexstruktur ähnlicher waren als die mit herkömmlichen DrugScore-Potentialen vorhergesagten. Dies wird deutlich, wenn ein RMSD $\leq 1,0$ Å als ein strengeres Kriterium zum Beurteilen der Posen herangezogen wird. In diesem Fall wurden 47% der Komplexstrukturen mit herkömmlichen DrugScore korrekt vorhergesagt, aber 86% mit QCSPScore und dem Gewichtungsfaktor zehn (Abschnitt 5.6). In der Schlussfolgerung hat dieser Teil der Arbeit gezeigt, dass die globale Optimierung auf der kombinierten E_{DS}/E_{QCSP}-Hyperfläche erfolgreich ist, was bei der Suche nach nativ-ähnlichen Lösungen zu besseren Ergebnissen führt.

Die Anwendung der hybriden Bewertungsfunktion mit dem Gewichtungsfaktor zehn auf die drei Systeme, für die experimentell bestimmte Daten vorliegen, ist erfolgreich, wenn solche Amidprotonen unberücksichtigt bleiben, die bei der Ausbildung von Wasserstoffbrücken beteiligt sind (Abschnitt 5.7). Die Verbesserung gegenüber der Verwendung herkömmlicher DrugScore-Potentiale ist signifikant, da letzteres in allen drei Fällen nicht erfolgreich ist. Außerdem werden nur bei Anwendung der hybriden Bewertungsfunktion nativ-ähnliche Komplexstrukturen für 1a9u und 1ydr vorhergesagt. Das betont wiederum die Wichtigkeit, experimentelle Daten in das Docking mit einzubeziehen, und zwar nicht nur als ein Vor- oder Nachfilter, wie es in der Vergangenheit bereits praktiziert wurde.

Der Vergleich der QCSPScore-Vorhersagen mit experimentell bestimmten Daten ermöglichte die Erkennung von Einschränkungen des Ansatzes und Überlegungen, wie diese Einschränkungen überwunden werden können. So führte beispielsweise die Nichtberücksichtigung von CSP von Amidprotonen, die an der Ausbildung von Wassersoffbrücken beteiligt sind, zu einer Verbesserung der Dockingergebnisse. Das war zu erwarten, da die Methode zur Vorhersage von CSP diesen Effekt nicht

modellieren kann. Interessanter war die Beobachtung, dass auch die Nichtberücksichtigung elektrostatischer Effekte notwendig für gute Ergebnisse war. Da CSP von Amidprotonen im Experiment breit und nicht symmetrisch verteilt sind, können nur schwer eindeutige Muster beobachtet werden, welche die Position des Liganden markieren. Die alleinige Berücksichtigung von Ringstromeffekten führt dabei zu besseren Ergebnissen: Ringstromeffekte (sofern vorhanden) liefern den größten signifikanten Beitrag zu den gesamten beobachteten CSP. Außerdem erzeugen sie ein charakteristisches Muster von Veränderungen der CS in ihrer Umgebung und können somit effektiv zur genauen Positionierung eines Liganden in der Bindetasche des Proteins genutzt werden. Bei der Untersuchung weiterer Faktoren, die das Ergebnis des Dockings beeinflussen könnten, stellten sich Bindetaschenbewegungen von bis zu 2 Å RMSD als den Erfolg des Dockings nicht negativ beeinflussend heraus. Wenn die CSP gleichmäßig verteilt sind, kann zudem eine Zuweisung von nur 40% oder weniger toleriert werden.

Die geringe Menge verfügbarer experimentell bestimmter Daten erlaubt es trotz dieser Erfolge nicht, alle Vorteile der Methode endgültig herauszustellen. Einige Einschränkungen des Verfahrens konnten aber festgestellt werden, sowohl was die Theorie hinter der Methode, aber auch die experimentell bestimmten Daten betrifft. Um die Anwendungsmöglichkeiten der Methode zu verfeinern und zu erweitern, müssen diese Einschränkungen in zukünftigen Entwicklungen reduziert werden.

Eine geringe Genauigkeit bei der Vorhersage von CSP von Amidprotonen zeigt dabei nur einen beschränkten Einfluss auf den Erfolg des Verfahrens verglichen mit unvollständigen und unregelmäßigen Zuweisungsmustern. Hinsichtlich der Wasserstoffbrücken wären ein besseres Verständnis und eine bessere Behandlung elektrostatischer Effekte wünschenswert. Gleiches gilt für eine Erweiterung des Ansatzes unter Berücksichtigung von Proteinflexibilität. Hinsichtlich des Problems der Wasserstoffbrücken haben Moon und Case (Moon and Case, 2007) erst kürzlich ein erweitertes empirisches Modell vorgestellt. Es bleibt aber nachzuprüfen, ob die Verwendung einer solchen Methode die Effizienz des Verfahrens nicht beeinträchtigt. Generalisierte Born-Modelle sind eine Alternative zu expliziten Lösungsmittelmodellen. Es ist offen, ob solche Modelle effizient in QCSPScore implementiert werden können und ob der Gewinn an Genauigkeit tatsächlich die Kosten

ausgleicht, die ein solches Verfahren verursacht. Effizientere Methoden zur Vorhersagen von CSP auf der Ligandenseite sind vor kurzem ebenfalls vorgestellt worden. NMRScore (Wang, et al., 2007) benutzt dafür quantenmechanische Ansätze. Dies ist nützlich zur Verfeinerung der Struktur, aber nicht für den Einsatz in Docking-Verfahren mit mittlerem Durchsatz.

Die Konzentration auf Ringstromeffekte schränkt die Fälle an Liganden ein, die untersucht werden können. Dies gilt aber nur diejenigen 5% der Strukturen aus der „MDL Drug Data Report" Datenbank, die keine aromatischen Systeme tragen. Die Nichtberücksichtigung der Proteinflexibilität schränkt das Verfahren ebenfalls ein, ist aber weniger eine Beschränkung des QCSPScore als des zugrunde liegenden Docking-Verfahrens.

Ich bin davon überzeugt, dass Verfahren wie QCSPScore für mit Dockingverfahren schwierig vorherzusagende Protein-Ligand-Komplexe deutliche Verbesserungen bringen können. Dies gilt beispielsweise im Fall des fragmentbasierten Ligandenentwurfs (Congreve, et al., 2008), weil kleine Fragmente eine Herausforderung für klassische Bewertungsfunktionen darstellen (Hubbard, et al., 2007; Chen and Shoichet, 2009), und für Protein-Protein-Wechselwirkungsflächen (Arkin and Wells, 2004), die typischerweise relativ flache Bindestellen haben, die für gängige Docking-Programme schwer anzugreifen sind (Gonzalez-Ruiz and Gohlke, 2006).

8   Bibliography

Abagyan, R.; Totrov, M.; Kuznetsov, D., Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J Comput Chem* **1994,** 15, (5), 488-506.

Abraham, R.; Reid, M., 1H chemical shifts in NMR. Part 18. Ring currents and pi-electron effects in hetero-aromatics. *J Chem Soc [Perkin 1]* **2002**, 1081-1091.

Abraham, R. J.; Canton, M.; Reid, M.; Griffiths, L., Proton chemical shifts in NMR. Part 14. Proton chemical shifts, ring currents and pi electron effects in condensed aromatic hydrocarbons and substituted benzenes. *J Chem Soc [Perkin 1]* **2000,** 4, 803-812.

Acharya, K. R.; Lloyd, M. D., The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci* **2005,** 26, (1), 10-14.

Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C., Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci U S A* **2003,** 100, (4), 1603-1608.

Arkin, M. R.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **2004,** 3, (4), 301-317.

Bax, A.; Grzesiek, S., Methodological Advances in Protein Nmr. *Accounts of Chemical Research* **1993,** 26, (4), 131-138.

Becattini, B.; Pellecchia, M., SAR by ILOEs: an NMR-based approach to reverse chemical genetics. *Chemistry* **2006,** 12, (10), 2658-62.

Beddell, C. R.; Goodford, P. J.; Norrington, F. E.; Wilkinson, S.; Wootton, R., Compounds designed to fit a site of known structure in human haemoglobin. *Br J Pharmacol* **1976,** 57, (2), 201-9.

Beddell, C. R., *The Design of drugs to macromolecular targets*. Wiley: Chichester ; New York, 1992.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** 28, (1), 235-242.

Bertini, I.; Fragai, M.; Giachetti, A.; Luchinat, C.; Maletta, M.; Parigi, G.; Yeo, K. J., Combining in Silico Tools and NMR Data To Validate Protein-Ligand Structural Models: Application to Matrix Metalloproteinases. *J Med Chem* **2005,** 48, (24), 7544-7559.

Bertini, I.; Calderone, V.; Fragai, M.; Giachetti, A.; Loconte, M.; Luchinat, C.; Maletta, M.; Nativi, C.; Yeo, K. J., Exploring the subtleties of drug-receptor interactions: the case of matrix metalloproteinases. *J Am Chem Soc* **2007,** 129, (9), 2466-75.

Blundell, T. L.; Jhoti, H.; Abell, C., High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Disc* **2002,** 1, (1), 45-54.

Bohm, H. J., The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J Comput-Aided Mol Des* **1994,** 8, (3), 243-256.

Brunger, A. T., Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992,** 355, (6359), 472-5.

Brunger, A. T.; Nilges, M., Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q Rev Biophys* **1993,** 26, (1), 49-125.

Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **1998,** 54, (Pt 5), 905-21.

Campbell, R. L. http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/ (Accessed June, 2009),

Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A., Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov* **2002,** 1, (7), 493-502.

Card, S. K.; Mackinlay, J. D.; Shneiderman, B., *Readings in information visualization : using vision to think*. Morgan Kaufmann Publishers: San Francisco, Calif., 1999; p xvii, 686 p.

Carlomagno, T., Ligand-target interactions: what can we learn from NMR? *Annu Rev Biophys Biomol Struct* **2005,** 34, 245-266.

Case, D. A., Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* **1995,** 6, (4), 341-6.

Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M., Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci U S A* **2007,** 104, (23), 9615-9620.

Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skelton, N. J., *Protein NMR Spectroscopy: Principles and Practice*. 2nd ed.; Elsevier Academic Press: 2007.

Cavasotto, C. N.; Orry, A. J., Ligand docking and structure-based virtual screening in drug discovery. *Curr Top Med Chem* **2007,** 7, (10), 1006-14.

Cavasotto, C. N.; Singh, N., Docking and high throughput docking: successes and the challenge of protein flexibility. *Current Computer-Aided Drug Design* **2008,** 4, 221-234.

Cioffi, M.; Hunter, C. A.; Packer, M. J.; Spitaleri, A., Determination of protein-ligand binding modes using complexation-induced changes in (1)H NMR chemical shift. *J Med Chem* **2008,** 51, (8), 2512-7.

Ciulli, A.; Abell, C., Fragment-based approaches to enzyme inhibition. *Curr Opin Biotechnol* **2007,** 18, (6), 489-96.

Clore, G. M.; Gronenborn, A. M., New methods of structure refinement for macromolecular structure determination by NMR. *Proc Natl Acad Sci U S A* **1998,** 95, (11), 5891-5898.

Cole, J. C.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R., Comparing protein-ligand docking programs is difficult. *Proteins* **2005**.

Coles, M.; Heller, M.; Kessler, H., NMR-based screening technologies. *Drug Discov Today* **2003,** 8, (17), 803-810.

Congreve, M.; Murray, C. W.; Blundell, T. L., Structural biology and drug discovery. *Drug Discov Today* **2005,** 10, (13), 895-907.

Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J., Recent developments in fragment-based drug discovery. *J Med Chem* **2008,** 51, (13), 3661-80.

Constantine, K. L.; Davis, M. E.; Metzler, W. J.; Mueller, L.; Claus, B. L., Protein-ligand NOE matching: a high-throughput method for binding pose evaluation that does not require protein NMR resonance assignments. *J Am Chem Soc* **2006,** 128, (22), 7252-7263.

Cornilescu, G.; Delaglio, F.; Bax, A., Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **1999,** 13, (3), 289-302.

Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A., Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* **2008,** 51, (20), 6237-55.

Chen, J. H.; Im, W.; Brooks, C. L., Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J Am Chem Soc* **2004,** 126, (49), 16038-16047.

Chen, Y.; Shoichet, B. K., Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat Chem Biol* **2009,** 5, (5), 358-64.

Choong, I. C.; Lew, W.; Lee, D.; Pham, P.; Burdett, M. T.; Lam, J. W.; Wiesmann, C.; Luong, T. N.; Fahr, B.; DeLano, W. L.; McDowell, R. S.; Allen, D. A.; Erlanson, D. A.; Gordon, E. M.; O'Brien, T., Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J Med Chem* **2002,** 45, (23), 5005-22.

Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie-International Edition* **2003,** 42, (24), 2718-2736.

DeLano, W. L. *The PyMOL Molecular Graphics System*, DeLano Scientific LLC: San Carlos, CA, USA.

Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K., Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* **2002,** 45, (11), 2213-2221.

Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J., HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003,** 125, (7), 1731-1737.

Erickson, J.; Neidhart, D. J.; Vandrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M., Design, Activity, and 2.8 a Crystal-Structure of a C2 Symmetrical Inhibitor Complexed to Hiv-1 Protease. *Science* **1990,** 249, (4968), 527-533.

Fernandez, C.; Wider, G., TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **2003,** 13, (5), 570-80.

Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., Assessing scoring functions for protein-ligand interactions. *J Med Chem* **2004,** 47, (12), 3032-3047.

Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K., NMR analysis of a 900K GroEL GroES complex. *Nature* **2002,** 418, (6894), 207-11.

Fradera, X.; Mestres, J., Guided Docking Approaches to Structure-Based Design and Screening. *Current Topics in Medicinal Chemistry* **2004,** 4, (7), 687-700.

Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** 47, (7), 1739-1749.

Fukui, L.; Chen, Y., NvMap: automated analysis of NMR chemical shift perturbation data. *Bioinformatics* **2007,** 23, (3), 378-380.

Gargaro, A. R.; Soteriou, A.; Frenkiel, T. A.; Bauer, C. J.; Birdsall, B.; Polshakov, V. I.; Barsukov, I. L.; Roberts, G. C.; Feeney, J., The solution structure of the complex of Lactobacillus casei dihydrofolate reductase with methotrexate. *J Mol Biol* **1998,** 277, (1), 119-34.

Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000,** 295, (2), 337-356.

Gohlke, H.; Klebe, G., Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie-International Edition* **2002,** 41, (15), 2645-2676.

Gonzalez-Ruiz, D.; Gohlke, H., Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem* **2006,** 13, (22), 2607-2625.

Gorczynski, M. J.; Grembecka, J.; Zhou, Y. P.; Kong, Y.; Roudaia, L.; Douvas, M. G.; Newman, M.; Bielnicka, I.; Baber, G.; Corpora, T.; Shi, J.; Sridharan, M.; Lilien, R.; Donald, B. R.; Speck, N. A.; Brown, M. L.; Bushweller, J. H., Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins Runx1 and CBF beta. *Chem Biol* **2007,** 14, (10), 1186-1197.

Goto, N. K.; Kay, L. E., New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* **2000,** 10, (5), 585-592.

Graves, A. P.; Brenk, R.; Shoichet, B. K., Decoys for docking. *J Med Chem* **2005,** 48, (11), 3714-3728.

Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L., The Blue Obelisk-interoperability in chemical informatics. *J Chem Inf Model* **2006,** 46, (3), 991-8.

Guntert, P., Automated NMR protein structure calculation. *Prog Nucl Magn Reson Spectrosc* **2003,** 43, (3-4), 105-125.

Habeck, M.; Rieping, W.; Nilges, M., Weighting of experimental evidence in macromolecular structure determination. *Proc Natl Acad Sci U S A* **2006,** 103, (6), 1756-61.

Haigh, C. W.; Mallion, R. B., Ring current theories in nuclear magnetic resonance. *Prog Nucl Magn Reson Spectrosc* **1980,** 13, 303.

Hajduk, P.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Disc* **2007,** 6, (3), 211-219.

Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W., One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *J Am Chem Soc* **1997,** 119, (50), 12257-12261.

Hajduk, P. J.; Mack, J. C.; Olejniczak, E. T.; Park, C.; Dandliker, P. J.; Beutel, B. A., SOS-NMR: A saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc* **2004,** 126, (8), 2390-2398.

Hamelryck, T.; Manderick, B., PDB file parser and structure class implemented in Python. *Bioinformatics* **2003,** 19, (17), 2308-10.

Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H., Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **2005,** 48, (2), 403-413.

Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W., Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem* **2007,** 50, (4), 726-41.

Haselhorst, T.; Weimar, T.; Peters, T., Molecular recognition of sialyl Lewis(x) and related saccharides by two lectins. *J Am Chem Soc* **2001,** 123, (43), 10705-14.

Hiller, N.; Fritz-Wolf, K.; Deponte, M.; Wende, W.; Zimmermann, H.; Becker, K., Plasmodium falciparum glutathione S-transferase - Structural and mechanistic studies on ligand binding and enzyme inhibition. *Protein Sci* **2006,** 15, (2), 281-289.

Hu, L. G.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A., Binding MOAD (Mother of All Databases). *Proteins-Structure Function and Bioinformatics* **2005,** 60, (3), 333-340.

Huang, N.; Jacobson, M. P., Physics-based methods for studying protein-ligand interactions. *Curr Opin Drug Discov Devel* **2007,** 10, (3), 325-331.

Hubbard, R. E.; Chen, I.; Davis, B., Informatics and modeling challenges in fragment-based drug discovery. *Curr Opin Drug Discov Devel* **2007,** 10, (3), 289-97.

Hunter, C. A.; Packer, M. J.; Zonta, C., From structure to chemical shift and vice-versa. *Prog Nucl Magn Reson Spectrosc* **2005,** 47, (1-2), 27-39.

Ihaka, R.; Gentleman, R., R: A Language for Data Analysis and Graphics. *J Comput Graph Stat* **1996,** 5, (3), 299-314.

Ilari, A.; Savino, C., Protein structure determination by x-ray crystallography. *Methods Mol Biol* **2008,** 452, 63-87.

Imming, P.; Sinning, C.; Meyer, A., Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **2006,** 5, (10), 821-34.

Jack, A.; Levitt, M., Refinement of Large Structures by Simultaneous Minimization of Energy and R-Factor. *Acta Crystallogr A* **1978,** 34, (Nov), 931-935.

Jahnke, W., Perspectives of biomolecular NMR in drug discovery: the blessing and curse of versatility. *J Biomol NMR* **2007,** 39, (2), 87-90.

Jansen, J. M.; Martin, E. J., Target-biased scoring approaches and expert systems in structure-based virtual screening. *Curr Opin Chem Biol* **2004,** 8, (4), 359-64.

Jayalakshmi, V.; Biet, T.; Peters, T.; Krishna, N. R., Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization. *J Am Chem Soc* **2004,** 126, (28), 8610-8611.

Jayalakshmi, V.; Krishna, N. R., CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J Magn Reson* **2004,** 168, (1), 36-45.

Jensen, C.; Herold, P.; Brunner, H. R., Aliskiren: the first renin inhibitor for clinical treatment. *Nat Rev Drug Discov* **2008,** 7, (5), 399-410.

Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G., Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr Opin Chem Biol* **2007,** 11, (5), 485-493.

Jiang, W.; Baker, M. L.; Jakana, J.; Weigele, P. R.; King, J.; Chiu, W., Backbone structure of the infectious epsilon 15 virus capsid revealed by electron cryomicroscopy. *Nature* **2008,** 451, (7182), 1130-U12.

Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** 267, (3), 727-48.

Kendall, M. G., A new measure of rank correlation. *Biometrika* **1938,** 30, 81-93.

Kendrew, J. C.; Perutz, M. F., A comparative X-ray study of foetal and adult sheep haemoglobins. *Proc R Soc Lond A Math Phys Sci* **1948,** 194, (1038), 375-98.

Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004,** 3, (11), 935-949.

Klebe, G., Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* **2006,** 11, (13-14), 580-94.

Klein, J.; Meinecke, R.; Mayer, M.; Meyer, B., Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR. *J Am Chem Soc* **1999,** 121, (22), 5336-5337.

Kuhn, P.; Wilson, K.; Patch, M. G.; Stevens, R. C., The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr Opin Chem Biol* **2002,** 6, (5), 704-710.

Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982,** 161, (2), 269-88.

Langer, T.; Vogtherr, M.; Elshorst, B.; Betz, M.; Schieborr, U.; Saxena, K.; Schwalbe, H., NMR backbone assignment of a protein kinase catalytic domain by a combination of several approaches: Application to the catalytic subunit of cAMP-dependent protein kinase. *Chembiochem* **2004,** 5, (11), 1508-1516.

Leach, A.; Shoichet, B.; Peishoff, C., Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *J Med Chem* **2006a,** 49, (20), 5851-5855.

Leach, A. R.; Hann, M. M.; Burrows, J. N.; Griffen, E. J., Fragment screening: an introduction. *Molecular Biosystems* **2006b,** 2, (9), 430-446.

Levitt, M. H., *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons: Chichester [England] ; New York, 2001.

Lin, M.; Shapiro, M. J.; Wareing, J. R., Diffusion-Edited NMR-Affinity NMR for Direct Observation of Molecular Interactions. *J Am Chem Soc* **1997,** 119, (22), 5249-5250.

London, R. E., Theoretical analysis of the inter-ligand overhauser effect: a new approach for mapping structural relationships of macromolecular ligands. *J Magn Reson* **1999,** 141, (2), 301-11.

Maaheimo, H.; Kosma, P.; Brade, L.; Brade, H.; Peters, T., Mapping the binding of synthetic disaccharides representing epitopes of chlamydial lipopolysaccharide to antibodies with NMR. *Biochemistry* **2000,** 39, (42), 12778-88.

Marcou, G.; Rognan, D., Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* **2007,** 47, (1), 195-207.

Mayer, M.; Meyer, B., Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **1999,** 38, (12), 1784-1788.

McCoy, M. A.; Wyss, D. F., Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *J Biomol NMR* **2000,** 18, (3), 189-198.

McCoy, M. A.; Wyss, D. F., Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J Am Chem Soc* **2002,** 124, (39), 11758-11763.

Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W., The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *J Am Chem Soc* **2000,** 122, 1241-1242.

Meng, E. C.; Shoichet, B. K.; Kuntz, I. D., Automated Docking with Grid-Based Energy Evaluation. *J Comput Chem* **1992,** 13, (4), 505-524.

Meyer, B.; Peters, T., NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* **2003,** 42, (8), 864-890.

Miyazawa, S.; Jernigan, R. L., Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation. *Macromolecules* **1985,** 18, 534-552.

Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L., Docking: Successes and challenges. *Curr Pharm Des* **2005,** 11, (3), 323-333.

Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R., Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol* **2008,** 153 Suppl 1, S7-26.

Moon, S.; Case, D. A., A new model for chemical shifts of amide hydrogens in proteins. *J Biomol NMR* **2007,** 38, (2), 139-150.

Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **1998,** 19, (14), 1639-1662.

Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009**.

Moy, F. J.; Chanda, P. K.; Chen, J. M.; Cosmi, S.; Edris, W.; Skotnicki, J. S.; Wilhelm, J.; Powers, R., NMR solution structure of the catalytic fragment of human fibroblast collagenase complexed with a sulfonamide derivative of a hydroxamic acid compound. *Biochemistry* **1999,** 38, (22), 7085-96.

Moy, F. J.; Chanda, P. K.; Cosmi, S.; Edris, W.; Levin, J. I.; Powers, R., Letter to the editor: 1H, 15N, 13C, and 13CO assignments and secondary structure determination of collagenase-3 (MMP-13) complexed with a hydroxamic acid inhibitor. *J Biomol NMR* **2000,** 17, (3), 269-70.

Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I., Comparison of ring current methods for use in molecular modeling refinement of NMR derived three-dimensional structures. *J Chem Inf Comput Sci* **1998**, 38, (4), 702-9.

Moyna, G. sdilicon.c. http://tonga.usip.edu/gmoyna/sdilicon/ (Accessed Accessed June, 2009),

Muegge, I., A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. *Perspectives in Drug Discovery and Design* **2000,** 20, (1), 99-114.

Neal, S.; Berjanskii, M.; Zhang, H.; Wishart, D., Accurate prediction of protein torsion angles using chemical shifts and sequence homology. *Magn Reson Chem* **2006,** 44, (S1), S158-S167.

Neuhaus, D.; Williamson, M., *The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd Edition*. Wiley-VCH: 2000.

Nilges, M.; O'Donoghue, S. I., Ambiguous NOEs and automated NOE assignment. *Prog Nucl Magn Reson Spectrosc* **1998,** 32, 107-139.

Orry, A. J.; Abagyan, R. A.; Cavasotto, C. N., Structure-based development of target-specific compound libraries. *Drug Discov Today* **2006,** 11, (5-6), 261-6.

Osapay, K.; Case, D. A., A New Analysis of Proton Chemical-Shifts in Proteins. *J Am Chem Soc* **1991,** 113, (25), 9436-9444.

Pagliaro, L.; Felding, J.; Audouze, K.; Nielsen, S. J.; Terry, R. B.; Krog-Jensen, C.; Butcher, S., Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol* **2004,** 8, (4), 442-9.

Parker, L. L.; Houk, A. R.; Jensen, J. H., Cooperative hydrogen bonding effects are key determinants of backbone amide proton chemical shifts in proteins. *J Am Chem Soc* **2006,** 128, (30), 9863-9872.

Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D. S., NMR-based structural characterization of large protein-ligand interactions. *J Biomol NMR* **2002,** 22, (2), 165-173.

Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G., Perspectives on NMR in drug discovery: a technique comes of age. *Nat Rev Drug Discov* **2008,** 7, (9), 738-45.

Perutz, M. F., An X-ray study of horse methemoglobin. *Proc R Soc Lond A Math Phys Sci* **1949,** 195, (1043), 474-99.

Perutz, M. F., *Protein structure : new approaches to disease and therapy*. W.H. Freeman and Co.: New York, 1992.

Pettit, F. K.; Bowie, J. U., Protein surface roughness and small molecular binding sites. *J Mol Biol* **1999,** 285, (4), 1377-82.

Polshakov, V. I.; Birdsall, B.; Frenkiel, T. A.; Gargaro, A. R.; Feeney, J., Structure and dynamics in solution of the complex of Lactobacillus casei dihydrofolate reductase with the new lipophilic antifolate drug trimetrexate. *Protein Sci* **1999a,** 8, (3), 467-81.

Polshakov, V. I.; Morgan, W. D.; Birdsall, B.; Feeney, J., Validation of a new restraint docking method for solution structure determinations of protein-ligand complexes. *J Biomol NMR* **1999b,** 14, (2), 115-122.

Powers, R., Functional genomics and NMR spectroscopy. *Comb Chem High Throughput Screen* **2007,** 10, (8), 676-697.

Radestock, S.; Bohm, M.; Gohlke, H., Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J Med Chem* **2005,** 48, (17), 5466-5479.

Reibarkh, M.; Malia, T. J.; Hopkins, B. T.; Wagner, G., Identification of individual protein-ligand NOEs in the limit of intermediate exchange. *J Biomol NMR* **2006,** 36, (1), 1-11.

Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; Lambert, R. W.; Merrett, J. H.; Mills, J. S.; Parkes, K. E. B.; Redshaw, S.; Ritchie, A. J.; Taylor, D. L.; Thomas, G. J.; Machin, P. J., Rational Design of Peptide-Based Hiv Proteinase-Inhibitors. *Science* **1990,** 248, (4953), 358-361.

Russ, A. P.; Lampel, S., The druggable genome: an update. *Drug Discov Today* **2005,** 10, (23-24), 1607-1610.

Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T., The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. *Angewandte Chemie-International Edition* **2005,** 44, (27), 4172-4175.

Scapin, G., Structural biology and drug discovery. *Curr Pharm Des* **2006,** 12, (17), 2087-97.

Schieborr, U.; Vogtherr, M.; Elshorst, B.; Betz, M.; Grimme, S.; Pescatore, B.; Langer, T.; Saxena, K.; Schwalbe, H., How much NMR data is required to determine a protein-ligand complex structure? *Chembiochem* **2005,** 6, (10), 1891-1898.

Schneider, G.; Fechner, U., Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* **2005,** 4, (8), 649-63.

Schulz-Gasch, T.; Stahl, M., Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Mod* **2003,** 9, (1), 47-57.

Schüttelkopf; Aalten, V., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, (8), 1355-1363.

Seifert, M. H., Assessing the discriminatory power of scoring functions for virtual screening. *J Chem Inf Model* **2006,** 46, (3), 1456-65.

Sharff, A.; Jhoti, H., High-throughput crystallography to enhance drug discovery. *Curr Opin Chem Biol* **2003,** 7, (3), 340-345.

Shoichet, B. K.; Bussiere, D. E., The role of macromolecular crystallography and structure for drug discovery: Advances and caveats. *Current Opinion in Drug Discovery and Development* **2000,** 3, 408-422.

Shoichet, B. K., Virtual screening of chemical libraries. *Nature* **2004,** 432, (7019), 862-5.

Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** 274, (5292), 1531-1534.

Simonson, T.; Archontis, G.; Karplus, M., Free energy simulations come of age: protein-ligand recognition. *Acc Chem Res* **2002,** 35, (6), 430-7.

Sotriffer, C. A.; Gohlke, H.; Klebe, G., Docking into knowledge-based potential fields: A comparative evaluation of DrugScore. *J Med Chem* **2002,** 45, (10), 1967-1970.

Stahl, M.; Rarey, M., Detailed analysis of scoring functions for virtual screening. *J Med Chem* **2001,** 44, (7), 1035-42.

Stark, J.; Powers, R., Rapid protein-ligand costructures using chemical shift perturbations. *J Am Chem Soc* **2008,** 130, (2), 535-545.

Stevens, R. C.; Yokoyama, S.; Wilson, I. A., Global efforts in structural genomics. *Science* **2001,** 294, (5540), 89-92.

Szilagyi, L., Chemical-Shifts in Proteins Come of Age. *Prog Nucl Magn Reson Spectrosc* **1995,** 27, 325-443.

Takeuchi, K.; Wagner, G., NMR studies of protein interactions. *Curr Opin Struct Biol* **2006,** 16, (1), 109-17.

Tirado-Rives, J.; Jorgensen, W. L., Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J Med Chem* **2006,** 49, (20), 5880-5884.

Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H. Y.; Markley, J. L., BioMagResBank. *Nucleic Acids Res* **2008,** 36, D402-D408.

van Dijk, A. D.; Boelens, R.; Bonvin, A. M., Data-driven docking for the study of biomolecular complexes. *FEBS J* **2005a,** 272, (2), 293-312.

van Dijk, A. D.; de Vries, S. J.; Dominguez, C.; Chen, H.; Zhou, H. X.; Bonvin, A. M., Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins* **2005b,** 60, (2), 232-8.

van Dijk, A. D.; Fushman, D.; Bonvin, A. M., Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked di-ubiquitin and validation against 15N-relaxation data. *Proteins* **2005c,** 60, (3), 367-81.

van Dijk, A. D.; Kaptein, R.; Boelens, R.; Bonvin, A. M., Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking. *J Biomol NMR* **2006,** 34, (4), 237-44.

Velec, H. F. G.; Gohlke, H.; Klebe, G., DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* **2005,** 48, (20), 6296-6303.

Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins-Structure Function and Genetics* **2003,** 52, (4), 609-623.

von Itzstein, M., The war against influenza: discovery and development of sialidase inhibitors. *Nat Rev Drug Discov* **2007,** 6, (12), 967-74.

Wang, B.; Brothers, E. N.; van der Vaart, A.; Merz, K. M., Jr., Fast semiempirical calculations for nuclear magnetic resonance chemical shifts: a divide-and-conquer approach. *J Chem Phys* **2004a,** 120, (24), 11392-400.

Wang, B.; Raha, K.; Merz, K. M., Pose scoring by NMR. *J Am Chem Soc* **2004b,** 126, (37), 11430-11431.

Wang, B.; Westerhoff, L. M.; Merz, K. M., A critical assessment of the performance of protein-ligand scoring functions based on NMR chemical shift perturbations. *J Med Chem* **2007,** 50, (21), 5128-5134.

Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A., Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J Am Chem Soc* **2001,** 123, (22), 5221-5230.

Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **2006,** 25, (2), 247-260.

Wang, R.; Lu, Y.; Wang, S., Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **2003,** 46, (12), 2287-2303.

Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A critical assessment of docking programs and scoring functions. *J Med Chem* **2006,** 49, (20), 5912-5931.

Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K., Testing a flexible-receptor docking algorithm in a model binding site. *J Mol Biol* **2004,** 337, (5), 1161-1182.

Weigelt, J.; McBroom-Cerajewski, L. D.; Schapira, M.; Zhao, Y.; Arrowmsmith, C. H., Structural genomics and drug discovery: all in the family. *Curr Opin Chem Biol* **2008,** 12, (1), 32-39.

Wen, X.; Yuan, Y.; Kuntz, D. A.; Rose, D. R.; Pinto, B. M., A combined STD-NMR/molecular modeling protocol for predicting the binding modes of the glycosidase inhibitors kifunensine and salacinol to Golgi alpha-mannosidase II. *Biochemistry* **2005,** 44, (18), 6729-6737.

Wider, G., Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Prog Nucl Magn Reson Spectrosc* **1998,** 32, 193-275.

Wider, G., Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *BioTechniques* **2000,** 29, (6), 1278-+.

Wider, G., NMR techniques used with very large biological macromolecules in solution. *Nuclear Magnetic Resonance of Biological Macromolecules, Part C* **2005,** 394, 382-398.

Wishart, D. S.; Case, D. A., Use of chemical shifts in macromolecular structure determination. *Methods Enzymol* **2001,** 338, 3-34.

Wong, S.; Witte, O. N., The BCR-ABL story: bench to bedside and back. *Annu Rev Immunol* **2004,** 22, 247-306.

Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* **1999,** 285, (4), 1735-47.

Wuthrich, K., The way to NMR structures of proteins. *Nat Struct Biol* **2001,** 8, (11), 923-925.

Wyss, D. F.; Arasappan, A.; Senior, M. M.; Wang, Y. S.; Beyer, B. M.; Njoroge, F. G.; McCoy, M. A., Non-peptidic small-molecule inhibitors of the single-chain hepatitis C virus NS3 protease/NS4A cofactor complex discovered by structure-based NMR screening. *J Med Chem* **2004,** 47, (10), 2486-2498.

Zhu, X. T.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H. L.; Morgenstern, K. A., Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure with Folding & Design* **1999,** 7, (6), 651-661.

Appendix: Implementation of QCSPScore

All dockings were done using AutoDock 3.0.5, with DrugScore fields as base scoring function. In order to incorporate the E_{QCSP} term and construct the hybrid scoring function we have extended AutoDock's original code. The most salient features of the extended implementation are specified in the UML diagram in Figure A.1. The core of the modification consists of two new classes: AtomP and Protein_csp. The existing class Eval (used by AutoDock for evaluations of the scoring function) has been extended to contain Protein_csp. A Protein_csp object is made of a number of AtomP objects. It has methods for reading both PDB structural data and the experimental CSP. Finally and for convenience, it also holds specific information about the ligand: which atoms are involved in aromatic rings and in groups generating anisotropy. AtomP objects hold the individual protons for which CSP have to be predicted (those for which experimental information is provided).
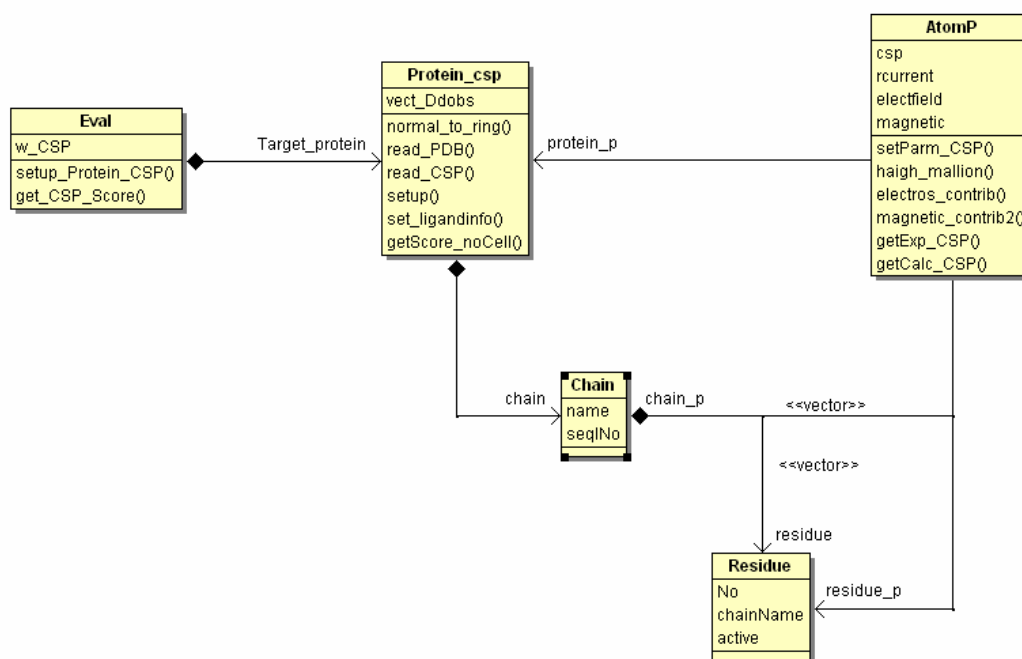

Figure A.1 UML diagram of the most relevant modifications done to AutoDock 3, in order to incorporate the information about CSP to the scoring function (diagram generated with BOUML (http://bouml.free.fr)).

Experimental CSPs, PDB with protein coordinates and ligand chemical groups relevant for the calculation of the CSP are passed to the program through a new line in the AutoDock's input file. The weight for the E_{QCSP} contribution is also specified at this point. A new DPF keyword ("shifts") has been defined to request AutoDock to setup a calculation involving CSP. The example of the new input line:
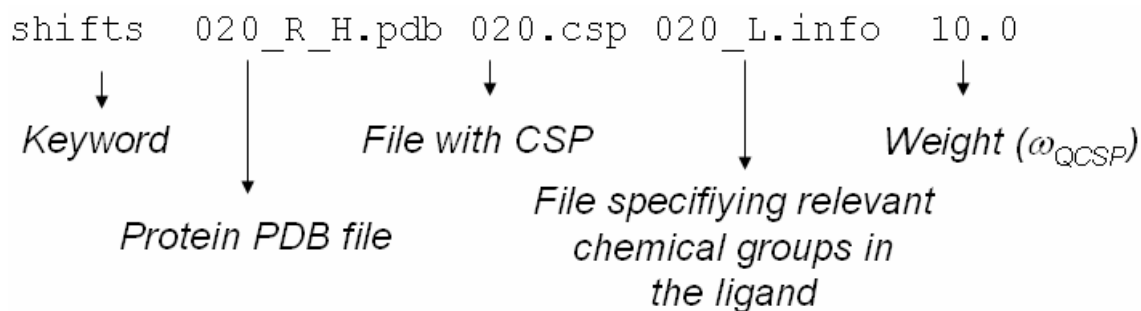

Figure A.2 Example of the new input line required to use AutoDock with our extension incorporating CSP.

CSP are specified as in the example below (example line in the "file with CSP"):

```
9 TYR  H   -0.010
```

where the first column corresponds to residue number, the second to residue name, the third to the atom name, and the fourth to the experimental CSP. Chemical information of the ligand is specified as:

```
ring 0.85 6 28 29 30 32 34 35
```

where "ring" is a keyword that identifies an aromatic ring. Groups that are sources of anisotropic effects are labelled "cgroup". What follows is the intensity of the ring factor (or the anisotropy), the number of atoms, and the atom numbers as in the pdbq file specified at the "move" keyword in a standard AutoDock run.

At docking time for every conformation with favourable interacting energy ($E_{DS} < 0$) the program calls the get_CSP_Score() procedure from Eval class, passing to it the coordinates of the new trial ligand pose to compute E_{QCSP}. E_{QCSP} is only contributing to the objective function during docking and is not used in the final re-ranking of solutions, if more than one docking is performed. For that final scoring and re-ranking, only DrugScore and the internal energy of the ligand as computed by AutoDock is

considered. All implementations have been done in C++ and a flag `CSPDOCK` has been defined for conditional compilation of the program (with or without CSP).

Curriculum vitae

Personal information

Surnames / First name	**González Ruiz, Domingo**
Address	c/Cayetano Duréndez, 1, 03390, Benejúzar, Spain
Telephone	+34.630.91.77.00
E-mail	dgruiz@gmail.com
Nationality	Spanish
Date and place of birth	13th of November, 1978 in Benejúzar (Alicante)

Work experience

November 2008 –	Community Pharmacist in Benejúzar, Spain
November 2007 – November 2008	Research Assistant at Max-Planck-Institut für molekulare Physiologie, Dortmund, Germany Chemical Biology Department (Prof. Herbert Waldmann)

Education and training

May 2004 –	Doctoral candidate (wissenschaftlicher Mitarbeiter) at J. W. Goethe University, Frankfurt am Main, Germany Adviser: Dr. Holger Gohlke Topic: *Protein-ligand docking with quantitative chemical shifts perturbations*
October 2002 – May 2004	Post-graduate diploma (Spanish D.E. A.) at "Severo Ochoa" Center for Molecular Biology, UAM - CSIC, Madrid, Spain Adviser: Dr. Angel Ramírez Ortiz Topic: *Implementation of an implicit solvent model for protein-ligand docking within the context of the AMBER forcefield*
October 2001 – July 2002	Pre-doctoral taught courses on Molecular Biology at the Miguel Hernández University, Elche, Alicante, Spain
October 1996 – July 2001	Degree in Pharmacy ("Licenciado") at the University San Pablo-CEU, Madrid, Spain

Publications

Papers	· González Ruiz, D., Gohlke, H. *Steering protein-ligand docking with quantitative chemical shift perturbations.* Submitted **2009**. · González Ruiz, D., Gohlke, H. *Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding.* Curr. Med. Chem. **2006**, 13, 2607-2625

Oral Presentations	· González Ruiz, D. Gohlke, H. *Driving protein-ligand docking by quantitative chemical shifts perturbations.* Young Modellers Forum 2006, 1st December, **2006**, London (United Kingdom)
	· González Ruiz, D. Gohlke, H. *Protein-ligand docking with quantitative chemical shifts* 20th Darmstaedter Molecular Modeling Workshop, May 23 – May 24, **2006**, Erlangen (Germany) (Best oral presentation award)
	· González Ruiz, D., Ortiz, A. R. *A New Implicit Solvent Model For Ligand-Receptor Electrostatic Binding Free Energy Calculations* 2nd meeting of the Catalan Bioinformatics Network, December 1 – December 2, **2003**, Les Avellanes (Spain)
	· González Ruiz, D., de Pascual-Teresa Fernández, B., *Methotrexate and folic acid interactions with DHFR: a pedagogical model for molecular visualization in medicinal chemistry.* II Congreso de aplicación de las nuevas tecnologías en la docencia presencial y eLearning, September 10 – September 11, **2002**, Alfara del Patriarca – Valencia (Spain)
Posters	· González Ruiz, D., Gohlke, H. *Incorporating quantitative experimental chemical shifts information into protein-ligand docking algorithms* 19th Darmstaedter Molecular Modeling Workshop, May 3 – May 4, **2005**
	· González Ruiz, D., Ortiz, A. R. *A New Implicit Solvent Model For Ligand-Receptor Electrostatic Binding Free Energy Calculations"* 1st International Conference, Zaragoza (Spain) *Biology after the genome: a physical view,* Februrary 11 – Februrary 13, **2004**
Patents	
	World Patent Application WO/2008/009758: "*Compound that can inhibit UBC13-UEV interactions, pharmaceutical compositions and therapeutic uses*", Scheper, J., Guerra-Rebollo, M., Sanvicens, G., Moure, A., González Ruiz, D., Rubio, N., Crosas, B., Loukili, N., Plans, V., Morreale, A., Blanco J., Ortiz, A. R., Messeguer, A. and Thomson, T. M. Publication date: 24th January, 2008
Scholarships and Awards	
	· FPU-fellowship from the Spanish Ministry of Education for post-graduate studies at the Centre for Molecular Biology "Severo Ochoa", 2003 – 2004, Madrid.
	· Second prize at the "Leche Pascual Professorship Awards for Young Researchers" (Universidad San Pablo – CEU) for*: Interactions of Folic Acid and Methotrexate with Dihydrofolate Reductase: A pedagogical method of Visualization in Medicinal Chemistry.* November 2001 Adviser: Dr. Beatriz de Pascual-Teresa.