Technische Universität München

Fakultät für Chemie

Lehrstuhl für biomolekulare NMR-Spektroskopie

Application of Enhanced Sampling Monte Carlo Methods for Structural Prediction of Protein-Protein Complexes

Zhe Zhang

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Carlo Camilloni

Prüfer der Dissertation:

1. Prof. Dr. Michael Sattler

2. Prof. Dr. Martin Zacharias

Die Dissertation wurde am 12.12.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 01.02.2017 angenommen.

Abstract

Protein-protein interactions are integral to all biological processes in the cell. A full understanding of the underlying mechanism promoting binding requires atomistic high-resolution details of the three dimensional structures of complexes, which in turn can further serve for pharmaceutical purposes. Computational protein-protein docking can help to complement experimental methods to bridge the big gap between the solved atomistic protein-protein complexes and the detected protein-protein interactions. This thesis explores the capacity and efficiency of the Markov chain Monte Carlo based methods in resolving the protein-protein docking problem. Following the three well-known theories for protein interaction, 'Key and Lock', 'induced fit' and 'conformational selection', I have proposed accordingly docking protocols for structural prediction of protein-protein complexes based on Monte Carlo and advanced sampling Monte Carlo methods, which have been implemented in the computational structural prediction software suite Rosetta and the docking engine ATTRACT. The new methods allow efficient flexible protein-protein docking and offer a wild range of applications to systematically generate realistic models of protein-protein complexes. Following Bayesian inference approach, I have derived an integrative docking solution based on the framework of Markov Chain Monte Carlo sampling docking.

Zusammenfassung

Protein-Protein Wechselwirkungen spielen eine Schlüsselrolle in allen biologischen Prozessen der Zelle. Ein volles Verständnis des zugrundeliegenden Mechanismus erfordert atomistische und hochauflösende Details der dreidimensionalen Strukturen von Proteinkomplexen. Diese Strukturen können wiederum für pharmazeutische Zwecke verwendet werden. Rechnergestütztes Protein-Protein Docking kann dazu beitragen, experimentelle Methoden zur Überbrückung der großen Lücke zwischen den registrierten atomistischen Protein-Protein-Komplexen und den erkannten Protein-Protein-Wechselwirkungen zu ergänzen. Diese Arbeit untersucht die Kapazität und Effizienz der Markov-Chain-Monte-Carlo basierten Methoden bei der Lösung des Protein-Protein-Docking-Problems. In Anlehnung an die drei bekannten Theorien zu Protein-Protein-Interaktionen, dem Schlüssel-Schloss-Prinzip, der Induced-Fit-Theorie und der „Conformational Selection“, habe ich Docking-Methoden zur Strukturvorhersage von Protein-Protein-Komplexen entwickelt, die auf dem Monte-Carlo-Verfahren basieren. Die Methoden wurden in die Rosetta Molecular Modelling Software-Suite und das ATTRACT Docking-Programm implementiert. Die neuen Methoden ermöglichen ein effizientes Docking von flexible Protein-Protein Komplexen und bieten eine Vielzahl von Anwendungen, um systematisch realistische Modelle von Protein-Protein-Komplexen zu generieren. Zusätzlich habe ich, basierend auf dem Markov-Chain-Monte-Carlo-Verfahren, eine integrative Docking-Methode mit Bayesschen Ansatz entwickelt. Die Methode ist stabil und fehlertolerant.

Publications

Two of the following research papers have been already published. The third one has been submitted, and at the moment it is under the process of revision after reviewers' comments.

[1] Zhang Z, Lange OF. Replica exchange improves sampling in low-resolution docking stage of RosettaDock. PLoS ONE 2013;8(8):e72096.

[2] Zhang Z, Schindler CEM, Lange OF, Zacharias M. Application of Enhanced Sampling Monte Carlo Methods for High-Resolution Protein-Protein Docking in Rosetta. PLoS ONE 2015;10(6):e0125941.

[3] Zhang Z, Ehmann U, Zacharias M. Monte Carlo Replica Exchange based Ensemble Docking of Protein Conformations. Proteins: Structure, Function, and Bioinformatics, 2017 (*in press*).

Contents

Chapter 1 Introduction

1.1 Protein interactions

Proteins are one of the large biomolecules, which is coded by DNA and translated by mRNA on ribosomes in cell. Most proteins are composed of the 20 amino acids, all of which contain the same backbone atoms, but differ from each other by their side-chains. The specificity of protein structure and function is thus a consequence of the different amino acid sequences.

Proteins are virtually involved in virtually all biological processes that taking place within or between cells, ranging from enzyme catalysis and inhibition to signal transduction and gene regulation. However, they rarely act in isolation, but rather in close association with other biomolecules, including nucleus acids, sugars, lipids, as well as other proteins. In fact, many biological processes are carried out by large molecular machines whose action is coordinated through intricate regulatory networks of transient protein interactions. Hence, it is the inter-relationship between molecules, rather than the individual components, that will eventually determine the behavior of a biological system [1]. Great efforts have thus devoted to unveil molecular interactions, among which protein-protein interactions (PPIs) takes a great proportion.

Genome-wide studies suggest that at least 65% of proteins in the cell work as part of a complex. The number of proteins to constitute a complex typically between two and four, with a few complexes with up to 200 members [2]. However up to date (November 2016), solved complexes in Protein Data Bank [3] count no more than 5% of all deposited entries.

To fully understand the mechanism behind biological process and further serving for pharmaceutical purposes and protein engineering, high resolution structural information is extremely valuable [4,5]. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the classical methods for experimental structural determination at atomic resolution. As the 'gold standard' for structural analysis of biomolecules, X-ray crystallography has so far contributed over 89% entries deposited in the PDB. Despite of some successes in solving complex structures, including even some macromolecular assemblies such as ribosome [6-8], RNA polymerase [9], the RNA exosome [10] and signal-recognition particle complex [11], X-ray crystallography faces nevertheless great challenge when coming to

complex structures. The limitation roots in its need of obtaining diffracting protein crystals of sufficient size and purity. This is especially challenging for (weak, encounter or transient) complexes and membrane proteins. NMR spectroscopy has the advantage over X-ray crystallography of allowing the determination of atomic resolution structures in solution with near-native conditions. It is very useful and powerful to map protein-protein interactions even for large systems, and especially to study dynamics in proteins and dynamics linked to protein complex formation[12]. We will see more contributions from NMR in this regard in the coming years.

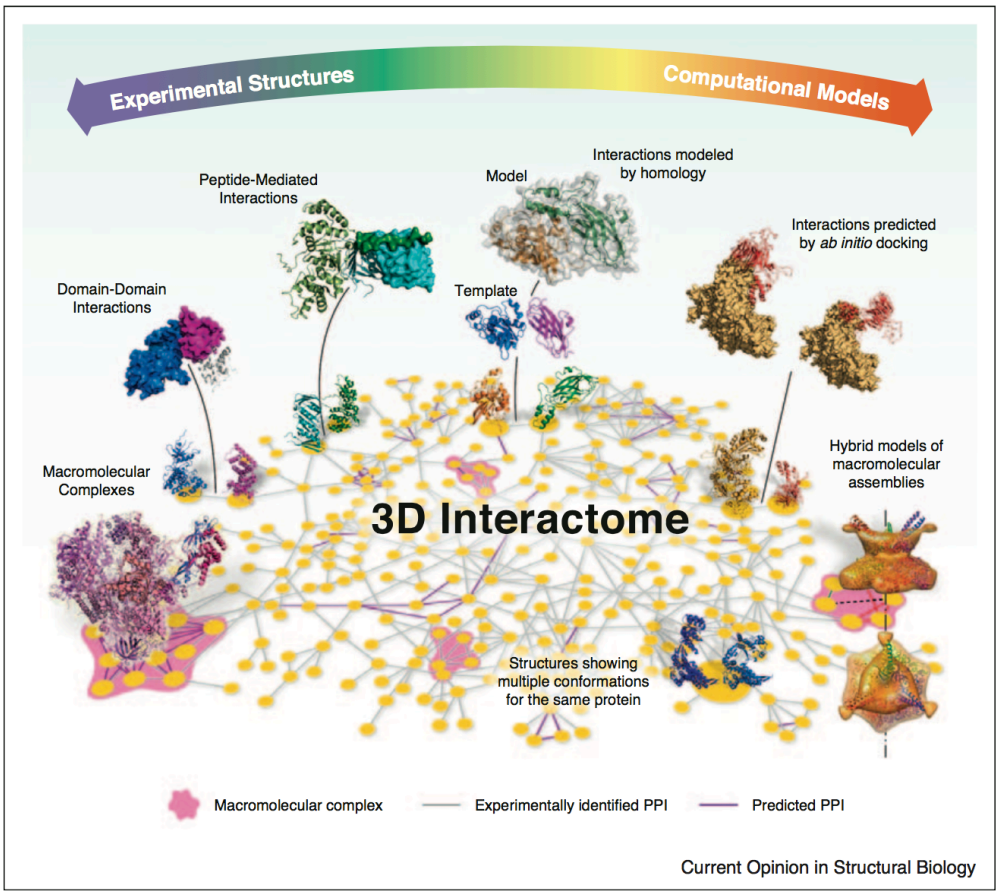

Figure 1.1. Building 3D interactomes from structures experimentally solved as well as modeled, single proteins as well as interactions between them, and interactions experimentally detected or predicted and modeled. This image is taken from [1].

While experimental bottlenecks or time/costing issue surrounding structure determination for complexes at high resolution, computational modeling is urgently needed to bridge the gap between the number of identified interactions and the number of interactions for which the 3D structure is known. Information from high-throughput interactome data sets may not be enough to give high-resolution

structures and accordingly better reveal the underlying mechanism that promotes binding, they can be nevertheless used to assist computational modeling to build the complete picture of 3D interactome (Figure 1.1). Many information about protein complexes can be obtained by different experimental methods. Yeast two hybrid system (Y2H), Tandem Affinity Purification (TAP) as well as mass spectrometry are capable of detecting if two proteins interact. Quantitative immunoblotting and mass spectrometry can be applied to determine the stoichiometry and composition of a complex. Cryo-electron microscopy maps and small-angle X-ray scattering (SAXS) data embody the shape of the complex. In addition, there are also experimental methods such as Förster resonance energy transfer (FRET) and cross-linking which can measure distances. In terms of computational modeling methods studying biomolecular interactions, homology modeling and protein docking work in the level of relatively small scale, with relatively more. In the level of networks scale, modeling suite such as Interactome3D, which integrate interaction data from the main pathway repositories for structural annotation and modeling of protein-protein interactions, is come to life [13].

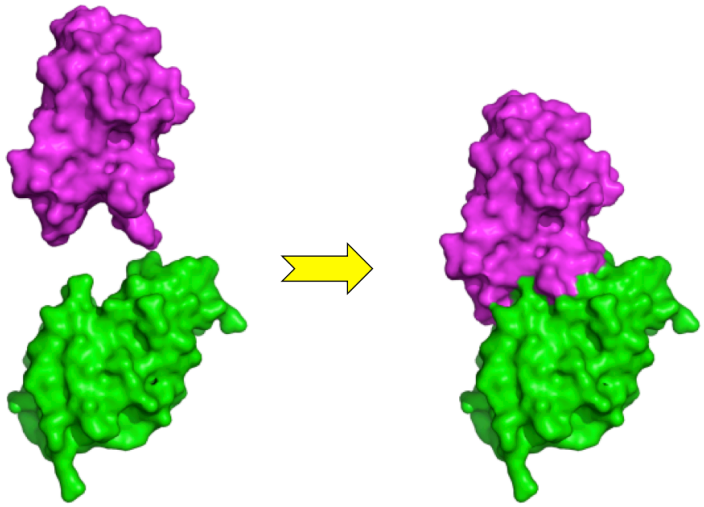

Figure 1.2. The term of DOCKING is to predict the structure of the interaction complex from individually known substitutes.

1.2 Protein-protein docking introduction

Computational protein docking describes any in silico methodology for combining structural knowledge (coordinates) of individual protein components with general knowledge about protein complexes (often in form of potential energy function) to find the best "match" between the two molecules (Figure 1.2). The goal of computational protein-protein docking is to understand and predict molecular

recognition, both structurally, finding likely binding mode, and energetically, predicting binding affinity. The first docking simulation was performed in 1978 by Wodak and Janin [14]. Since then it has emerged with increasing interest from the scientific community.

The most simple form of docking is bound docking, that is to reconstruct the complex using the respective bound component from the native complex structure. For application, unbound docking is however the target. That is to predict the complex structure between the receptor and ligand from the unbound (free) form or bound form resolved with another interaction partner. Docking prediction involves decoy generation and selection of the near-native structures from the generated decoys using energy function or sometimes filters. The success of a predictive docking thus requires an efficient method that samples complex conformations which ideally has sufficient number of decoys close to the native complex, and an accurate energy function which scores the near-native conformations with low/better energies [15].

In docking practice, additional biochemical information, such as binding sites, or distance information from experimental data as well as evolutionary information, are often applied.

1.3 Mechanism behind sampling strategy in docking

Protein-protein docking is commonly discussed according to the complexity it deals with and divided into two main docking stages as shown in the green blocks in Figure 1.3: first rigid-body docking in which internal coordinates are kept rigid, then followed by refinement, which often allows movements of sidechain or backbone and loop motions in selected region on selected decoys. In between and afterwards, there could also be multiple filtering, scoring as well as clustering steps for discriminating and selection. Along with the work flow, the number of decoys that the docking program deals with typically decreases as illustrated by the inverted triangle [16,17].

Advances in docking methods has often gone hand in hand with new insights into the protein-protein binding mechanism. The strategy behind the docking method often mirrors our understanding in the past of the recognition mechanism. In the following the docking strategy will be discussed together with the model of the mechanism behind it.

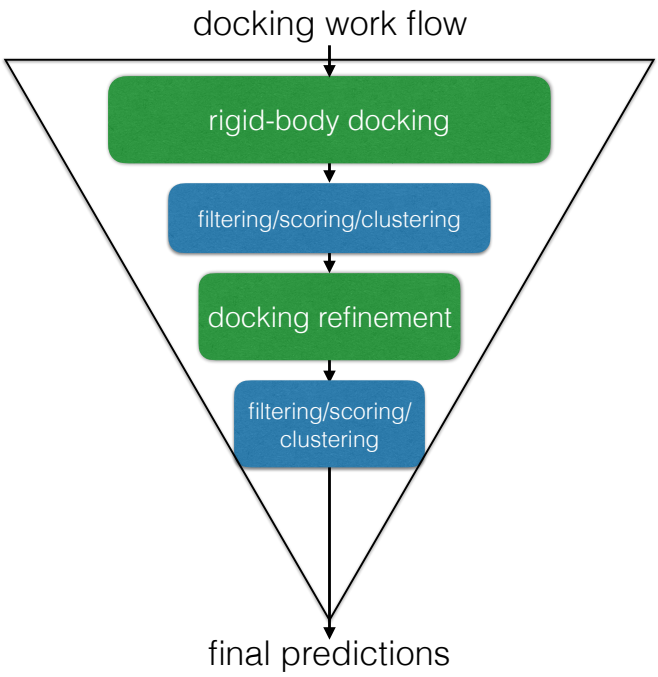

Figure 1.3. Docking work flow in most docking programs. Most docking programs utilize this two-stage docking approach (green blocks), first rigid-body docking then followed by refinement on selected decoys. In between, one or multiple filtering, scoring or clustering procedures can be applied for structure selection (blue blocks).

1.3.1 Lock-and-key and rigid-body docking

The earliest textbook model for protein-protein binding, 'lock-and-key' model, was first proposed in 1894 by Fischer [18]. In the 'lock-and-key' model, the conformation of the free and bound proteins are assumed essentially the same (Figure 1.4). This model has been the most influential model in the development of protein docking algorithms. The rigid-body simplification is stemmed from this model and has been employed in most docking programs. In rigid-body docking, the space of putative docking geometries is sampled broadly, keeping the partner structures rigid, which corresponds to six degrees of freedom, three translational and three rotational ones. It serves for efficient exploration of the orientational space. For a significant number of protein-protein complexes, especially those with neither hinge-motions nor disorder transitions and where the conformational changes are limited to the sidechains, rigid-body docking has worked reasonably well.

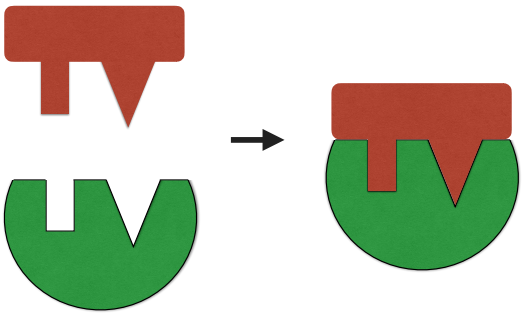

Figure 1.4. Key and lock model of protein-protein interaction, in which the two binding partners are assumed to be rigid without internal conformational changes.

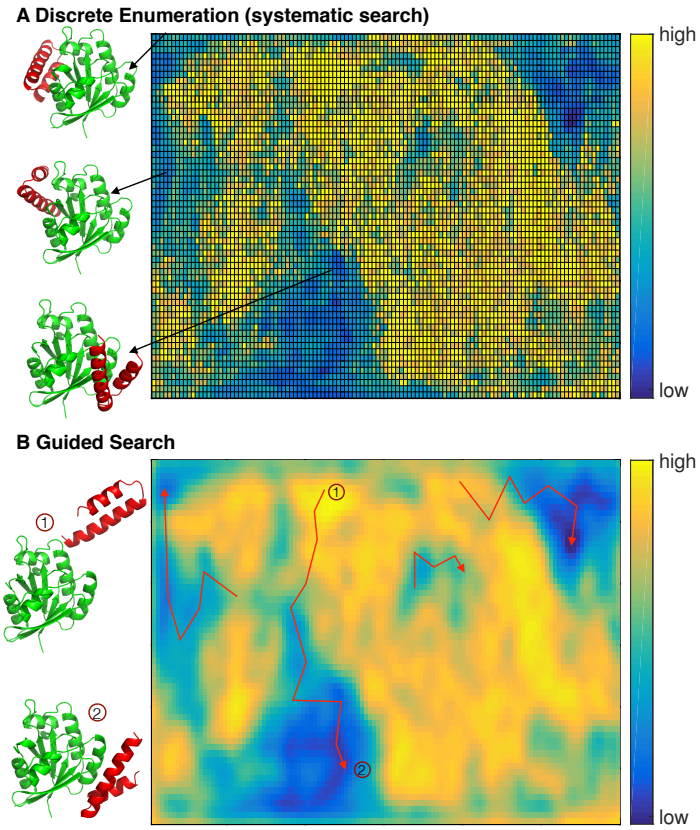

Figure 1.5. 2D illustration of systematic search and guided search. The color represents the energy level in the 2D conformational space. In the systematic search, the conformational space is discretized and enumerated. As for the guided search, the conformational space is explored, starting from random position and driven by the potential function.

The searching algorithm can be divided into two big categories: systematic search and guided search. In the systematic search, it exhaustively enumerates the discretized conformational space, which is illustrated in Figure 1.5 using a 2D space for convenience. For protein-protein docking, the receptor is typically kept fixed. For each discretized rotation angle, the ligand is moved with a certain step size

along the axis between the center of mass of the two proteins. This process is then repeated for all rotation angles. In contrast to the discretized systematic search, the energy landscape revealed from guided search appears to be more continuous and more smoothed (Figure 1.5). In guided search algorithms, it doe not explicitly explore the whole conformational space, but rather it starts from one random position and move to the next, which is guided by the energy function, and eventually converge to local or global energy minima.

1.3.1.1 Systematic search

Correlation methods have been explored to accelerate the systematic search. SY Huang has shown that for docking programs with fast sampling facilitated systematic searching algorithms, at least 50% success rate can be achieved when top-2000 structures are considered [19]. Fast Fourier transform (FFT) correlation based programs such as FTDock [20], DOT [21,22], ZDOCK [23], MolFi t[24,25], SDOCK [26], PIPER [27], F2DOCK [28] and Gramm-X [29], represent the biggest class within this category. While the potential function and the mapping of the potential to grids vary in those docking programs, the basic principle behind are the same. FFT correlation was first in 1992 introduced into protein docking in the group of Vakser by Katchalski- Katzir *et al* [24] and was further explored since then [30]. We illustrate here the principle algorithm using the most basic shape complementary scoring scheme. Both receptor and ligand are digitalized in a 3D grid in Cartesian space with dimensions of *N x N x N* and described by two discrete functions a and b respectively:

$$a_{l,m,n} = \begin{cases} 1 & \text{on the surface of the protein} \\ \rho & \text{in side the molecule} \\ 0 & \text{out side the protein} \end{cases}$$

$$b_{l,m,n} = \begin{cases} 1 & \text{on the surface of the protein} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the protein} \end{cases}$$

and association score for a given displacement (α, β, γ):

$$c_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \cdot b_{l+\alpha,m+\beta,n+\gamma}$$

where l, m, and n are the grid indices. Interior parameters are set $\rho \ll -1$ and $0 < \delta < 1$ to discriminate overlapping regions. The correlation value reflects the

complementarity of the two binding partners. When molecule b penetrates molecule a, multiplication of the negative ρ with positive δ results in a negative value as a penalty to the overall correlation value. In the algorithm in stead of calculating the product of $a_{l,m,n} \cdot b_{l+\alpha,m+\beta,n+\gamma}$ at each grid, the score of c for each displacement (α, β, γ) can be completed in one go three FFT transform:

$$c = FFT^{-1}(\overline{FFT(a)} \times FFT(b))$$

With the development of grid-based FFT correlation method in protein docking, additional pairwise scoring terms are also considered, including electrostatic energy, desolvation effects as well as knowledge-based potentials. They have been implemented in some docking programs with a separate grid and combined together with shape complementary scoring or as a filter afterwards; while there are also groups that have used a single grid, with shape information stored in the real part and pairwise potentials in the imaginary part as follows:

$$a_{l,m,n} = \begin{cases} 1 + i\sqrt{\omega}E_a & \text{on the surface of the protein} \\ \rho & \text{in side the molecule} \\ 0 + i\sqrt{\omega}E_a & \text{out side the protein} \end{cases}$$

and

$$b_{l,m,n} = \begin{cases} 1 - i\sqrt{\omega}E_b & \text{on the surface of the protein} \\ \delta & \text{inside the molecule} \\ 0 - i\sqrt{\omega}E_b & \text{outside the protein} \end{cases}$$

In fact, FFT correlation can not only be applied in Cartesian space, but also can be applied to accelerate the search in the spherical polar space, such as docking program HEX [31,32] and FRODOCK [33] as well as the work by Padhorny D *et al* [34].

In addition to FFT correlation based systematic positioning of the proteins on grids, another systematic search method is geometric-hashing, the focus of which is surface matching. The proteins in this case are described by surface descriptors evolved from Connolly surface to capture the essential features of the surface. The representing docking programs of this category include PatchDock [35], SymmDock [35] and so on.

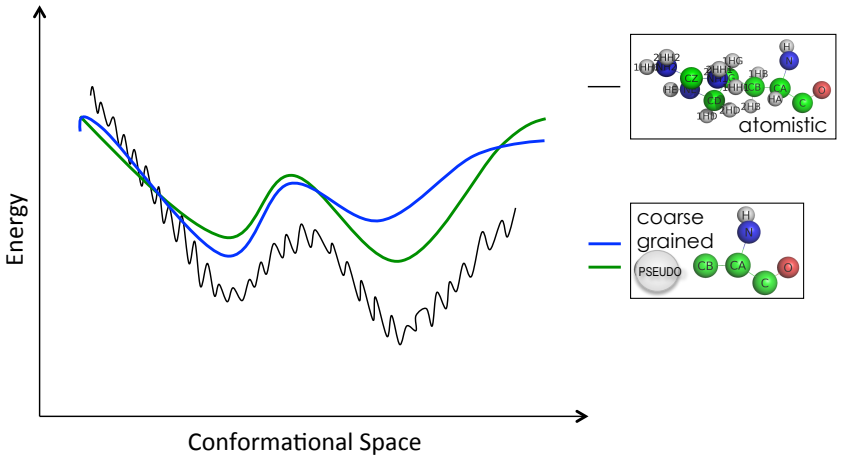

Figure 1.6. Schematic comparison of the energy landscape with atomistic and coarse-grained representation. The figure illustrates the effect of the smoothing of the energy landscape in a coarse-grained model as compared to an atomistic model. The flattening enables efficient exploration of the energy landscape in search for the global minima, while avoiding traps in the local minima. Ideally coarse graining preserve the global minima as shown by the green line. But in practice it is more often the case as shown in blue.

1.3.1.2 Guided search

Guided search, in which the energy is minimized, represents the second important category for rigid-body docking stage. Within the guided search methods, the most popular strategy is the shot gun approach as used in RosettaDock [36], ATTRACT [37], Haddock [38] and ICM-DISCO [39], where many thousands of randomly generated initial conformations are shortly optimized within the energy landscape which could be composed from more sophisticated scoring functions. The algorithms for energy optimization can be gradient based deterministic energy minimization as used in ATTRACT and RosettaDock, molecular dynamics as in Haddock, Monte Carlo simulation as in RosettaDock and ICM-Disco, or genetic algorithm [40].

In the rigid-body docking stage of both systematic sampling and shot-gun guided search, the philosophy behind is to pursue a broad even sampling in the hope of covering near-native structures. A contrasting philosophy is Importance Sampling. It falls also into the guided search category. Here more computer time is spent in the low energy region and cumulates more conformations at the bottom of the energy landscape. Importance Sampling in the form of replica exchange Monte Carlo (REMC) has been studied to sample the rigid-body docking stage of RosettaDock [41], and will be discussed in detail in the respective chapter.

One major disadvantage of guided search is that it is more computationally expensive than the accelerated systematic search due to the rather detailed energy function. In ICM-DISCO and ATTRACT, grid potential is used to speed up the calculation. Each energy term is pre-calculated and stored around the receptor at each point of a grid. In between of the grids, the energy is interpolated. The error caused by the interpolation increases when the two partner proteins get closer. A denser grids is then applied in ATTRACT for the short distance range to counteract this effect.

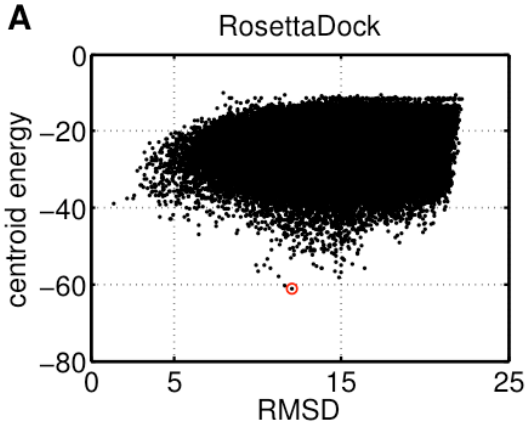

Figure 1.7. Centroid energy versus Ca-RMSD from bound docking on 1EMV from RosettaDock low-resolution docking stage. The centroid energy function in RosettaDock prefers an alternative binding mode with larger buried surface area even for bound docking on target 1EMV.

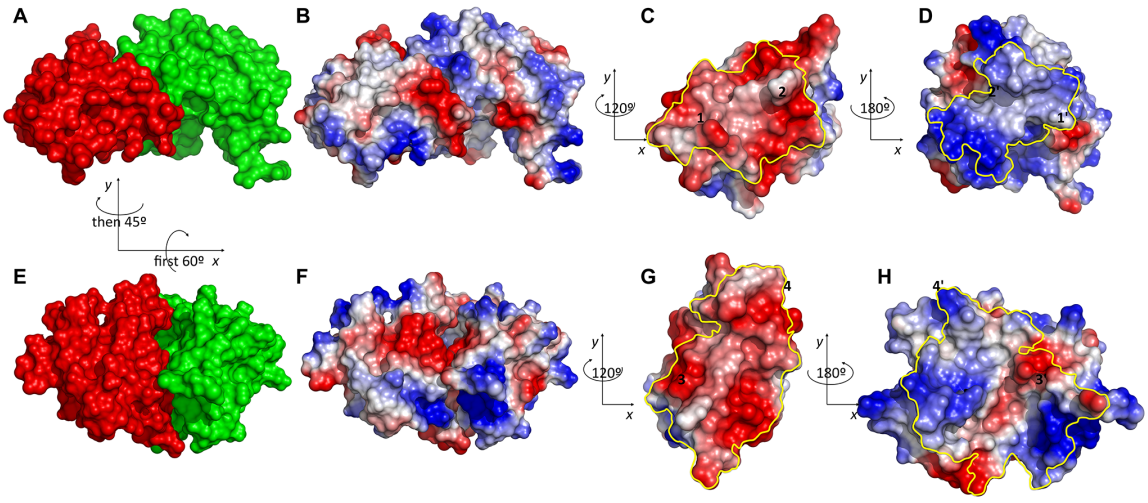

Figure 1.8. Electrostatic surface potential (bound docking 1EMV) of native (A–D) and interchain_cen preferred conformation (E–H) in all-atom representation. A) native complex with receptor in green and ligand in red B) electrostatics map of the native complex C) electrostatics map of native ligand interface D) electrostatics map of native receptor interface E) interchain_cen preferred conformation after refinement F) electrostatics map G) electrostatics map of receptor interface H) electrostatics map of ligand interface. The yellow lines in C, D, G and H indicate the respective interface regions,

and number pairs (e.g. 1 in C and 19 in D) indicate corresponding contact regions. Relations of viewing angle are given between panels where required.

Coarse graining is another option for speeding up the calculation and has been applied in docking programs such as RosettaDock, ATTRACT, pyDockCG [42] as well as scoring potential SIPPER [43]. Detailed reviews have been given by Ravikumar KM *et al* [44] and Sebastian Kmiecik *et al* [45] and so on. Coarse grained models group several atoms into a pseudo atom. Taking RosettaDock for example, all the sidechain atoms are grouped into the so-called *centroid* pseudo atom, to reduce the number of particles to be considered in pairwise potential calculations. While coarse graining considerably reduces the complexity of the potential, the energy landscape is much more smoothed comparing to that of the all-atom representation (Figure 1.6). Ideally, the overall feature and global minima preserves after coarse graining as shown by the green curve in Figure 1.6. But this is seldom the case in practice. Bound docking on 1EMV in Rosetta gives a perfect example. As shown in Figure 1.7 for bound docking on 1EMV with centroid representation in Rosetta, the near-native conformations scores far worse than alternative binding regions with RMSD between 10 and 17 Å. Scrutinizing into the alternative binding region, we found that they are featured with a bigger buried surface area, which is preferred by the energy function with the coarse grained model, meanwhile contain many non-complementary charged atoms as shown in Figure 1.8. While sampled directly with all-atom representation and accordingly the all-atom energy function, bound docking on 1EMV gives lowest energies in the near native region (Figure 1.9). Further more, the near-native solutions rank 2nd by size after clustering. This makes a dramatic contrast with the results in Figure 1.7 with centroid representation, where the near native region got rarely visited.

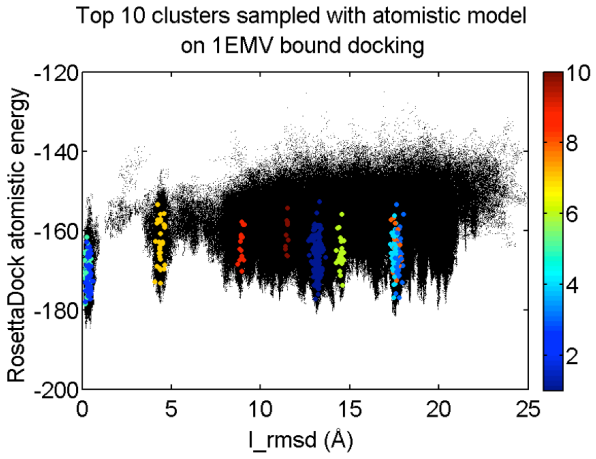

Figure 1.9. Atomistic energy versus I_rmsd of bound docking on 1EMV sampled directly with atomistic representation. Representing decoys of top 10 clusters are labeled with colors. The clusters

are ranked by size and the rank is indicated with the colorbar. Lowest energy appears among the near-native decoys which ranks by size also very well (2nd) after clustering.

1.3.2 Induced-fit and docking refinement

The second model for protein-protein interaction is Koshland's 'induced-fit model' [46]. In 'induced-fit model', ligand binds to receptor to first form an encounter complex, then trigger mutually conformational changes (Figure 1.10). The conformational changes can be local or global, sidechains or backbone. Up to 70% protein structures do not deviated more than 2 Å between the bound and unbound form [47]. The rigid-body docking stage has taken advantage of this fact to narrow down the list of possible docking solutions. A full *ab initio* calculation is in practice computationally prohibitive due to the huge number of degrees of freedom. Flexibility is thus commonly represented in the refinement stage of the two stage docking approaches in selected regions on selected decoys after elimination in the previous rigid-body docking stage. Many docking programs, such as RosttaDock, iATTRACT [48] and Haddock, use atomistic representation in the docking refinement stage to allow the optimization of the sidechains and/or backbones in the selected region (normally interface region), while other studies perform refinement with a coarse-grained model [49]. In either case, refinement protocols require that the starting configurations contain at least one structure that is already fairly close to the native structure for a docking success. This is because, rather than enlarging the structural diversity, refinement stage mainly serves to improve the energetics of the docked candidates for discrimination. However, since initial docking stages commonly use a simplified energy function, it is often misled and does not necessarily produce sufficiently many near-native candidates for the subsequent refinement stage, especially when there is an alternative binding site with bigger buried surface [41,50].

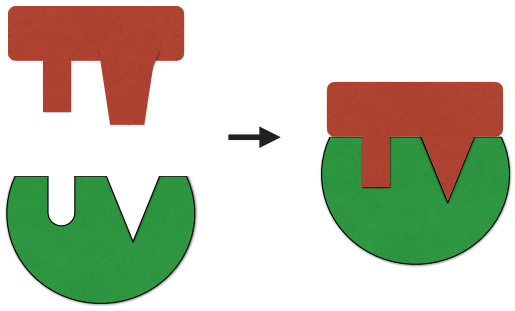

Figure 1.10. Induced fit model of protein-protein interaction, which assumes that binding triggers mutual conformational changes on both binding partners.

In my work, rather than performing refinement on selected decoys from previous docking stage (e.g. rigid-body docking stage), I have applied enhanced sampling Monte Carlo directly for atomistic docking refinement from a random starting position within a certain restricted rigid body space from the native position to mimic the scenario in docking practice, that the binding site is often approximately known from bioinformatics information or experiments [51].

1.3.3 Conformational selection and ensemble docking

An alternative hypothesis of the binding mechanism 'conformational selection', which can be traced back to 1960s, is getting more experimental evidence support recently with the development of experimental techniques [52-55]. It postulates that all protein conformations pre-exist, and the ligand selects the most favored conformation to bind. Following binding the ensemble undergoes a population shift, redistributing the conformational states. Recently, Csermely *et al* has extended the conformational selection model to describe the general scenario, where both selection- and adjustment-type steps follow each other [1,54]. Following this model of molecular recognition, conformational flexibility can be represented implicitly by ensembles of conformers taken from, for example, NMR structures, MD simulations or any other conformational sampling method. The conformations can span various degrees of flexibility, from small sidechain rearrangement to large-scale global backbone motions [2,56]. Docking of the ensemble conformers one by one (cross docking) results in a dramatic increase of the computing time with respect to the size of the conformational ensembles and becomes prohibitively expensive for large ensembles. Thus studies investigating the use of ensembles with cross docking of all ensemble members have only been performed on relatively small benchmarks or single cases, which have shown improvement over single conformation unbound docking in terms of the number of generated near-native decoys [34,57,58]. The representation of the receptor binding region by multiple conformations has been found beneficial for the docking success also in the area of docking of drug molecule candidates to protein binding sites [59-63]. In particular, it has been found that docking in combination with stochastically switching between different receptor structures is both computationally efficient and improves the accuracy of the docking results [60,61]. Monte Carlo type switching between receptor binding site conformations (ordered along the structural similarity) during docking avoids the rapid increase of the computational demand in case of docking ligands separately to each member of the ensemble [58]. The stochastic search rapidly settles towards the regime of receptor binding site conformations compatible with favorable binding of a ligand [60,61].

In my study, I have followed this theory and applied replica exchange Monte Carlo to perform ensemble docking. The docking ensemble space as well as the rigid-body space are both explored by Monte Carlo sampling. For the ensemble space, a model is suggested with a probability calculated based on the similarity to the current model, and accepted according to the metropolis criteria. This will be described in detail in respective chapter.

1.4 Scoring in Docking

Docking prediction involves two aspects: decoy generation (sampling) and selection (scoring). The success of predictive docking requires both, that an efficient method that samples sufficiently many near-native decoys and an accurate energy function that ranks the near-native conformations with better scores. The goal of the scoring is to discriminate between near native and non-near native decoys. Two criteria to assess a scoring function are efficiency and selectivity. On one hand, the scoring function should be fast enough to allow its application on a large number of conformations. This justifies the commonly applied two stage docking strategy, where rigid-body docking quickly identify putative binding region using relatively simple energy function with affordable computer time, while the refinement stage employ a much more sophisticated energy function on less decoys to evaluate more details of the interaction. On the other hand, accurately discrimination between near-native and non near-native conformations is however the most essential and critical need.

Various scoring functions have been developed together with or independently from sampling algorithms. They can be categorized into two groups: "integrated" and "edge" functions based on if they are applied during or at the end of the sampling procedures [3,64]. The difference is that the integrated scoring function will bias the generated ensemble towards energy conformation within the energy landscape. To simplify the method development, sampling and scoring are often decoupled as in the case of "edge" functions. In this way, newly designed and developed scoring functions can be applied on pre-generated and stored decoys. Different "edge" functions can thus be easily evaluated and compared. This has indeed greatly facilitated the development of scoring functions. One potential problem here is that the training decoys may lead to bias of the scoring function. To promote the development of scoring function, the organizers of CAPRI added a separate challenge to test scoring methods, in which scorer groups re-rank all the uploaded models. However, the published results of CAPRI 2009 (rounds 13-19) [4,5,65] and CAPRI 2013 (rounds 20-27) [6-8,66] show that the same group gives substantially

more quality predictions as predictor than as scorer. Based on these results, S. Vajda *et al* suggested that decoupling sampling and scoring may lead to loss of accuracy for docking predictions, and integration of the two-steps will generally lead to better docking results [9,67]. In contrast to "edge" functions, scoring in "integrated" function is integrated into the search stage. It filters emerging solutions and thus forms part of the design of the solutions [10,64]. An example of the integrated approach would be Monte Carlo method, which employs Metropolis Hastings criteria to decide if to accept the newly suggested move based on the energies. The representative docking program utilizing Monte Carlo method include RosettaDock, ICM-DISCO etc.

Over the years, scoring function has gained great improvement. Most of current scoring functions combine a few parameters, such as van der Waals and electrostatic interactions, desolvation energy and hydrogen bonding, as well as empirical statistical potentials from analyzing of the existing complex structures. Which parameters to use in the scoring function strongly depends on the docking scenario. In rigid-body *ab initio* docking, the program needs to sample the entire rotational and translational space, resulting in a six-dimensional search. To keep the method computationally feasible, a relatively simple energy function is typically applied. From the very beginning of protein docking, shape complementarity has been applied to evaluate the generated conformations. It has repeatedly affirmed its important role in docking predictions [11,24,25,68,69]. Indeed, the 3D structures of many protein complexes have revealed good shape complementarity in the interface between receptor and ligand. Then with the development of docking method, electrostatics have become a routine addition to shape complementarity [19,70], such as in docking programs like FTDOCK [13,20], ZDOCK [1,23], RosettaDock [14,36], ATTRACT [15,37], MolFIT [25,42-44], DOT [19,21], and HEX [24,31]. Meanwhile, most docking programs also consider solvation and desolvation effects, which are normally incorporated in the scoring function in the form of an excluded volume model of shape complementarity. Fernandez-Recio *et al* showed that solvation energy is the most important component in the total binding energy. Using accessible surface area based desolvation energy in the ICM scoring function has improved dramatically the predictive results [30,71]. Besides, pyDock and ASPDock [72] have been developed with great emphasize of desolvation energy [35,72,73]. Knowledge-based empirical potentials are also used in programs like ZDOCK, PIPER [27,35], and RosettaDock.

Despite of all the advances in the scoring function [16,36,64,74], there is no scoring function evaluated so far that can discriminate between near native and non-near

native binding mode on all targets. It remains one of the great challenges that hindering docking success. Meanwhile, scoring functions will encounter great new challenges while more and more docking programs strive to include backbone flexibility.

Apart from using the biophysical biochemical or empirical based scoring function to discriminate between near native and non-near native structures, biological information can also be applied to guide the docking process or filter out non-physical conformations. Great success in certain targets in CAPRI test where prior-knowledge about the interaction is available, indicates that with the help of some biological information, it is straight-forward for many current algorithms to make good docking predictions [16,37]. Ambiguous restraints [38] or specific information of the contacting/non-contacting residues can both improve the success rate of docking predictions. Useful biological information could be obtained from experiments such as mutagenesis, chemical cross-linking, or phylogenetic data [39,64]. In this thesis, a integrative docking solution derived from Bayesian inference in the framework of Monte Carlo based docking will be given in **Chapter 5**.

1.5 Monte Carlo: a natural coupling of sampling and scoring

The task of quickly and accurately exploring large regions of the conformational space in protein-protein docking remains to be a big challenge, especially for large systems. Grid based FFT method and geometric hashing can quickly and efficiently explore the entire rotational and translational space in *ab initio* docking, but limited in combining certain prior information and including conformational flexibility. Monte Carlo methods, which use random numbers to examine various problems from a stochastic perspective, is however not constrained with the two problems. Further more, it is also suitable for all kinds of potential functions, not restricted to derivative ones as required for deterministic minimization algorithms.

In Monte Carlo simulation, comply with detailed balance, an unbiased random move is suggested. The acceptance of the suggested move is decided by the Metropolis criterion with the probability:

$$P = min(1, exp(-\Delta E(x, x')/kT))$$

where $\Delta E(x, x')$ is the energy change from current state to the suggested one, and kT denoting the inverse temperature, is the product of Boltzmann's constant k and thermodynamic temperature T.

1.5.1 Short randomized Monte Carlo sampling docking problem

Monte Carlo method is especially useful in simulating systems with many coupled degrees of freedom, and has been applied in many scientific disciplines. For protein-protein docking, Monte Carlo method has been applied in docking programs such as RosettaDock, ICM-DISCO [39], FireDock [41,75] and ROTAFIT [47,76], for rigid-body docking stage and/or docking refinement.

In RosettaDock [36,48], short Monte Carlo simulation (500 MC steps) is applied to optimize the rigid-body orientations on many randomly generated initial placements with centroid representation. After the low-resolution rigid-body docking stage, the residues are restored on the retained structures with the sidechains from the input unbound structure. The sidechain optimization is then achieved by Monte Carlo and minimization of the rotamers, which are the sidechain conformations that are found with a high propensity in the protein data bank [49,77]. Meanwhile, the rigid-body orientation is simultaneously optimized with Monte Carlo Minimization (MCM).

In ICM-DISCO [39,41,50], ligand is initially placed around the fixed receptor with an average distance of 15 Å between neighbors. Then the sampling of the rotational and translational degrees of freedom of the ligand starting from the initial placements is performed by a pseudo-Brownian Monte Carlo minimization. The interaction energy considers van der Waals potential, electrostatic potential corrected for the solvation effect, hydrogen-bonding potential and a hydrophobicity potential, plus a atomic solvent-accessible surfaces based solvation energy added to revaluate the docking solutions obtained from unbound subunits [39,51]. For computational efficiency, the interaction energy was represented by grid potentials and pre-calculated at each point of a grid. After the initial rigid-body docking stage, retained structures after clustering are further refined allowing full flexibility of the interface ligand sidechains using biased probability Monte Carlo for the minimization in the side-chain torsion angles space [52-55,78]. The intramolecular energy for the refinement is based on the same grid potentials as in the initial rigid-body docking stage with increased truncated threshold for van der Waals potential. Besides, internal energy accounting the flexibility of the ligand is added. For both rigid-body stage and refinement, ICM-DISCO has used full-atomistic representation.

FireDock [75] refines the rigid-body docking results from PatchDock allowing rearrangement of the interface sidechains and adjustment of the relative orientation of the molecules. The flexibility of the sidechains is also modeled as in RosettaDock by rotamers and optimized by integer linear programming based on Dunbrack

rotamer library. Following the rearrangement of the sidechains, the rigid-body orientation is refined by Monte Carlo minimization.

In the *ab initio* docking applications summarized above, relatively short Monte Carlo simulation is applied in RosettaDock with coarse-grained representation featured with more smoothed energy landscape, in ICM with carefully designed initial placements. Each Monte Carlo routine starts from one conformation and after short simulation, output at the end one structure as the prediction. In docking refinement of RosettaDock, ICM-DISCO and FireDock, Monte Carlo adjusts the rigid-body orientation after sidechain movement by Monte Carlo or other methods on random selected sidechains based on the probability in Dunbrack rotamer library. The task of docking refinement is minimal yet trivial adjustment on the primary predictions from rigid-body docking stage with the goal of discrimination between near native and non-near natives.

Most of the times, the docking refinement is done with full-atom representation as in the three refinement procedures summarized above. The energy landscape in the full-atom representation is generally characterized by many local minima that are separated by free-energy barriers. For this kind of systems, Monte Carlo often suffers from long impractical equilibrium time due to the suppression of tunneling through these barriers. To overcome this problem, enhanced sampling method can be combined with Monte Carlo. One of such method is parallel tempering Monte Carlo, or replica exchange Monte Carlo. The idea of the algorithm is to simulate the system with multi-copies (replicas) at different temperatures and by exchange between different replicas to overcome the free-energy barriers in the energy landscape. The high temperature replicas serves to sample broadly in the phase space, whereas low temperature replicas having precise sampling in the local region of energy minima. By exchange with the higher temperature replicas, low temperature ones are prevented from being trapped in local minima.

S Lorenzen and Y Zhang have thus combined Monte Carlo with replica exchange in ROTAFIT [76] to refine the rigid-body docking solutions from FFT-based docking method ZDOCK. In this work, Monte Carlo is applied in combination with replica exchange to be more efficient. For each replica 2500 composite Monte Carlo steps are run: first unbiased translation and rotation with scaled magnitude such that the displacement of the interface atoms is under control. Following the rigid-body move conformations of two randomly selected sidechains respectively from the receptor and ligand in the interface are randomly changed based on the probability in Dunbrack rotamer library [79]. The acceptance of the composite move is evaluated

after rigid-body move and sidechain move by the Metropolis criterion. Replica exchange is attempted between neighboring replicas after each composite move. To ensure a smooth refinement process, a smoothed van der Waals energy function is gradually roughened to the standard Lennard-Jones potential.

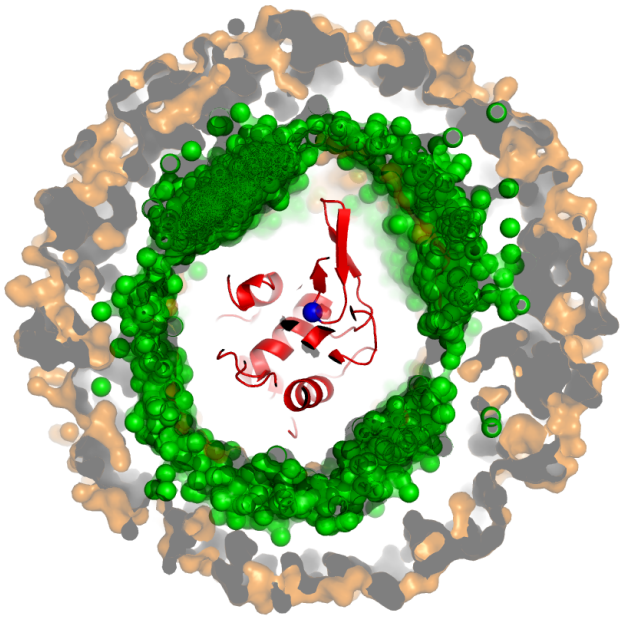

Figure 1.11. The population of mass center of the ligand from lowest (green) and highest (yellow) temperature replicas around the receptor in replica exchange Monte Carlo sampling low-resolution docking stage of RosettaDock. The mass center is represented by sphere and the picture is the cross section view. In the middle is the receptor shown in red with blue sphere indicates its mass center.

1.5.2 Long trajectory Monte Carlo for protein docking

In all the applications in the above section, including even ROTAFIT which has though combined replica exchange, Monte Carlo mainly serves to search in the space of local minima, with each of many trajectories starting from one single structure and evolves with relatively short Monte Carlo steps to the end output structure as one single prediction. Using the evolution of long trajectories of Monte Carlo simulation, is a contrasting strategy for its application in protein-protein docking. To this end, the snapshots (evolution) along the long trajectory after every certain Monte Carlo steps are taken as predictions for final analysis. In this way, the population of the generated decoys is more physically meaningful. When unbiased sampling guaranteed, the generated ensemble is canonical ensemble with Boltzmann distribution. Following this idea, long trajectory Monte Carlo was applied in combination with replica exchange to sample the low-resolution docking stage of RosettaDock (**Chapter 2**) [41]. Attribute to the relatively smooth energy landscape

of the coarse-grained representation, replica exchange Monte Carlo has achieved thorough sampling for the rigid-body global docking. In the highest temperature, the interaction between the ligand and the receptor is weak. With the receptor fixed, the ligand is restricted with a distance threshold from diffusing away from the receptor, such that the ligand can freely move within the restricted space and populates relatively even. In the lowest temperature, the interaction between the ligand and receptor is strong. The ligands are thus much closer to the receptor with denser population in several low energy regions (Figure 1.11).

The general idea of parallel tempering is not limited to exchange or swap between different temperatures. Apart from using temperature as the variable parameter for replicas, other parameters which can help to minimize the barriers in the energy landscape have also been considered. For example, Fukunishi *et al*. developed a Hamiltonian replica exchange method for biomolecular systems, in which they scale the interaction energy for different replicas [80]. Besides, multi-dimensional replica exchange has also been proposed. Thus in our work for high-resolution docking refinement in Rosetta, two dimensional replica exchange Monte Carlo has been employed, with temperature as the parameter for one dimension and scaling of the van der Waals potential as parameter for the second dimension (**Chapter 3**). With full-atom representation in high-resolution docking refinement, the energy landscape is too rugged. Even a small perturbation can produce huge change on the system's energy. This makes the exchange extremely difficult, requiring not only big number of replicas but also very well designed spacing between the replicas. Well-tempered ensemble introduced by M. Bonomi and M. Parrinello is the biased ensemble sampled by well-tempered metadynamics when the energy is used as the collective variable [81]. One great feature of well-tempered ensemble is approximately the same average energy as the canonical ensemble but much larger fluctuations. This can extremely facilitate the exchange in parallel tempering. Hence, to avoid the complex designing for the settings of the replicas, well-tempered-ensemble technique is combined upon parallel tempering in our work. This has greatly reduced the required number of replicas and improved the round-trip time and thus improved the diffusion efficiency of the system. Monte Carlo is here used to sample the rigid-body degrees of freedom as well as the rotamers of randomly selected sidechains each time based the probability in Dunbrack rotamer library. We have one rigid body mover, and three sidechain movers for perturbing Chi angle, perturbing rotamer and jumping to a random rotamer. The four movers have for each a pre-defined sampling weight. During the simulation, one of the four movers

are randomly selected based on the pre-defined weight and applied for each Monte Carlo step.

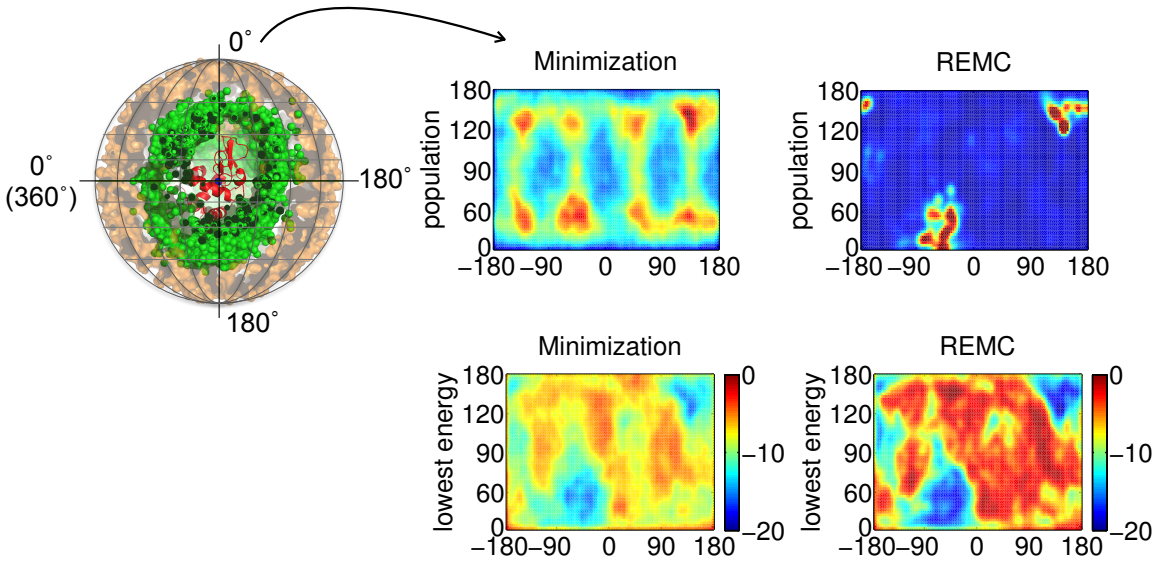

Figure 1.12. ATTRACT ensemble docking on unbound target 1J2J with minimization and REMC sampling method. We projected the mass centers of sampled ligands on the unit sphere around the mass center of the receptor, and plotted the population of the mass center of the ligand on this unit sphere in the first row. In the second row, the lowest energy of the decoys from the respective grid is plotted.

In both replica exchange Monte Carlo sampling low-resolution docking stage of RosettaDock and enhanced Monte Carlo sampling docking refinement in Rosetta, Monte Carlo simulation always starts from one single conformation, and each replica runs on a separate processor. The position of the starting conformation has shown to be irrelevant with the population of the sampled ensemble for the simulation length we have used. One disadvantage is that the waiting time is bit of too long. In ATTRACT, de Vries has implemented the energy function based on potential grids, which has yielded tremendous speedup. To make full use of this grid potential, we have thus made a compromise between long trajectory Monte Carlo and the many randomized starting points Monte Carlo (short and only the end output as prediction), and implemented the new protocol in docking program ATTRACT (**Chapter 4**). We have used several hundred starting positions. All replicas start from the same several hundred initial placements. Replica exchange is attempted between the decoys started originally from the same position and done for all simultaneously. Monte Carlo simulation is used to sample in the conformer ensemble docking space simultaneously or independently with rigid-body orientation space based on a pre-defined probability. In Figure 1.12, we plotted the ensemble docking results using ATTRACT with minimization and REMC sampling.

From the figure of population and respective lowest energy from REMC sampling ensemble docking, we can see that the compromise and the on-the-fly replica exchange scheme although do not give strict canonical ensemble, the lowest energies still reflects the most populated region. In the results of minimization, the low energy regions are similar as in REMC, but the population of the sampled ligands is irrelevant with the low energy regions.

Chapter 2 Replica exchange improves sampling in low-resolution docking stage of RosettaDock[§]

2.1 Introduction

Protein-protein interactions are one of the fundamental molecular mechanisms of life, and to investigate them it is important to know the atomic structures of the formed complexes. Since many proteins have multiple interaction partners, the number of protein complexes is far larger than the number of individually folded proteins. At the same time, the number of known complex structures is far lower than that of monomeric proteins, illustrating the experimental challenges involved in solving the structure of protein complexes [1,82,83].

Computational protein docking describes any *in silico* methodology for combining structural knowledge of individual protein components with general knowledge about protein complexes (often in form of a potential energy function) and, if available, sparse data of the complex [64,84-86]. Popular sources of sparse data include, cryo-EM, SAXS, NMR chemical shift perturbations or chemical crosslinking [1,82,83,87,88].

Currently, many popular docking programs employ a two-stage approach: First, conformational space is sampled broadly, keeping partner structures rigid. Second, structures are refined in one or multiple steps [16,74,89-91]. To account for possible side-chain or loop motion, many docking methods employ a low-resolution model during their initial rigid-body search to create the required level of softness without adding extra degrees of freedom [76,90,92]. Whereas some methods use solely a low-resolution representation, others refine structures in an all-atom representation, often allowing also side-chain and loop motion [23,36-38]. While the all-atom representation allows a more exact modeling of the energetics of protein-protein interfaces, it also leads to a rugged energy landscape that is hard to sample [93]. As a result, the high-resolution stage generally serves only to discriminate conformations, not to generate them.

The initial (low-resolution) structural exploration stage of common docking programs such as Haddock [38], Attract [37], ICM-DISCO [39] or RosettaDock [36] is driven by a shotgun approach of short energy minimizations started from many thousands of randomly generated initial conformations. Another large class of

[§] The work presented in this chapter has been published [41]

programs such as DOT [21], ZDOCK [23], 3D-DOCK [94], and Gramm-X [29] employ grid-based fast Fourier transform (FFT) search of rigid-body degrees of freedom to find the low-energy conformations [74,91]. Recently, geometric hashing has been applied to quickly identify possible binding modes [95].

The philosophy behind the low-resolution sampling in most docking programs, and in particular those that employ shotgun sampling, is to guarantee an even sampling in the low-resolution stage. A contrasting philosophy is Importance Sampling, which is constructed to spend more computer time in regions of low energy than in those with high energy. It is often argued that using Importance Sampling, too much computer time is spent in a small number of low-energy regions of a potentially misleading low-resolution energy function, while a thorough exploration of conformational space is neglected. Temperature Replica Exchange, however, might overcome the lack of exploration. Thus we address the following questions in this study: First, does Importance Sampling in the form of Replica Exchange have a benefit over shotgun sampling for the low-resolution stage of protein-protein docking despite the potentially misleading low-energy function. Second, whether the highly skewed populations of conformations generated by Importance Sampling are advantageous or disadvantageous for the subsequent refinement stage.

We thus introduced *ReplicaDock*, a replica exchange Metropolis-Monte Carlo method [96,97] for the low-resolution stage of protein-protein docking, which has been implemented within the RosettaDock program. We chose temperature levels such that the lowest temperature reflects a bound state and the highest temperature an unbound state within the RosettaDock *centroid* energy function. Within the unbound state, the binding partners are free to sample the whole surface. In the bound state, the binding partners stick together and explore local conformational space to find the lowest energy conformation accessible within the current binding mode [98].

Of course protein-binding partners would freely diffuse in the unbound state, rendering collision events rather rare. To counter-act this physical but undesired behavior we introduced an artificial restraint energy. This *encounter constraint* is a flat-bottom restraint energy that acts on the distance of the center of mass of both binding partners. As it penalizes only those conformations that are too far away to touch, the encounter constraint has no effect on the bound conformational ensemble whatsoever. However, by constricting the available conformational space volume, it increases the local concentration of the binding partners and enhances their collision rate dramatically.

As prediction of protein complex structures is still an unsolved problem [65], protein-protein docking programs are most useful in combination with sparse experimental data [38,85]. Nevertheless, here, we tested ReplicaDock without any additional experimental data to fully focus on the sampling strategy and exclude any other possible interference factors. Performance is compared to shotgun sampling, and enumerative sampling (represented by ZDOCK) foremost via the ability to sample the lowest-energy structures after refinement and secondly by the ability to sample near-native conformations. Furthermore, we analyzed how well accurate predictions of native conformations are possible after all-atom refinement with Rosetta.

The manuscript is organized as follows. At first, we analyze the shotgun sampling employed currently in RosettaDock [36] and demonstrate that it strongly depends on the initial random placement and little on the energy function (Section 2.3.1). Subsequently, ReplicaDock sampling is introduced and it is showcased at hand of target 1ppf (Section 2.3.2). In Section 2.3.3, we show that near-native sampling below 4Å I_rms to the native conformation is a necessary condition to achieve a positive effect on structural accuracy in subsequent refinement. Thus, we analyze the frequency of "hits" over a benchmark of 30 proteins for shotgun and ReplicaDock sampling (Section 2.3.4), and compare to ZDOCK (Section 2.3.5). Next, we consider all-atom refinement and analyze the sampling of distinct energy landscape features and the recovery of the native energy basins (Section 2.3.6). Finally, we analyze the capability to predict accurate complex structures from the refined shotgun, ReplicaDock and ZDOCK ensembles, respectively, by employing commonly employed metrics of prediction quality (Section 2.3.7). As discussed above, Importance Sampling might suffer from a misleading low-resolution energy function, and indeed we find some targets in the benchmark where this is the case. We discuss some of the shortcomings of the low-resolution energy function of RosettaDock that lead to alternative non-native binding modes in Section 2.3.8.

2.2 Methods

2.2.1 Energy Function

2.2.1.1 Low-resolution energy

The low-resolution stage uses the *interchain_cen* energy function which has been previously introduced for RosettaDock [36,77]. This energy function consists of a term to reward contacting residues (*interchain_contact),* a penalty term for overlapping residues *(interchain_vdw),* a docking-specific statistical residue

environment (*interchain_env*) and residue-residue pair-wise potentials (*interchain_pair*) with weights 2.0, 1.0, 1.0 and 1.0, respectively.

The *interchain_contact* component of the low-resolution energy function in Rosetta has originally been capped at -10 Rosetta Energy Units (REU) [36], which corresponds to 40 contacting residues within a distance cutoff of 6Å between centroid-interaction centers of the two binding partners. The centroid pseudo atom is the interaction center representing all sidechain atoms in Rosetta's low-resolution representation. The cap avoids over-stabilization of spurious binding interfaces with large contact area, but also has the disadvantage that perturbations away from an optimal conformation with a large contact area are no longer penalized. Here, we obtained optimal performance by combining conformations sampled with and without energy capping. Thus, if not otherwise noted, for Shotgun (Section 2.2.3) and ReplicaDock (Section 2.2.4) always half of the generated conformations in low-resolution stage are sampled with capping at -10 REU. For ReplicaDock this is realized by running 2 of 4 trajectories with the capped energy function.

2.2.1.2 All-atom energy

The high-resolution stage uses the standard all-atom energy for RosettaDock as given by the weight-set *docking* [77]. For final analysis an interface energy is computed by subtracting the all-atom energy of non-interacting partners from the all-atom energy of the interacting binding partners. To compute the energy of non-interacting partners the two binding partners are moved far away from each other while keeping all internal degrees of freedom fixed.

2.2.2 Generating initial conformations

To generate initial conformations we randomly perturb the orientation of both binding partners. To this end, we uniformly draw rotation matrices from the rotation group $SO(3)$ by generating Euler angles α,β,γ (in z,y,z notation) with α,γ drawn uniformly from the interval $[-\pi,\pi]$ and setting $\beta=\cos^{-1}z$ with z drawn uniformly from interval $[-1,1]$

Subsequently, the binding partners are slid into contact using steps of 1Å by first increasing the distance between the binding partners until the energy term $interchain\_vdw<0.1$, and then again decreasing until $interchain\_vdw>0.1$.

2.2.3 Shotgun Protocol (low-resolution stage)

Shotgun sampling in RosettaDock's low-resolution stage proceeds in two steps. First, a random initial conformation is generated as described in the preceding section. Second, a Monte-Carlo (MC) sampling procedure with 500 steps is applied to optimize the low-resolution energy. At step i a proposed conformation $\boldsymbol{x}_*$ is accepted according to the Metropolis Criterion $p_{accept}(\boldsymbol{x}_* \mid \boldsymbol{x}_i) = \min(1, \exp(\beta[V(\boldsymbol{x}_i) - V(\boldsymbol{x}_*)]))$ where $V(\boldsymbol{x})$ denotes the potential energy of conformation $\boldsymbol{x}$, and β the inverse temperature. The inverse temperature is kept constant at $\beta = 0.8^{-1}\text{kcal}^{-1} \cdot \text{mol}$ and the step-sizes are adjusted every 50 steps to maintain a 50% acceptance rate. The initial step-sizes are drawn from normal distributions with mean value of 0.7Å (translation along all the three axes) and 5° (rotation around the axis of protein centers and tilt off this axis in a randomly-chosen direction) [36]. At the end, the lowest energy conformation observed during the 500 MC-steps is recorded as final output.

By presetting the number of generated decoys *(-nstruct)* we adjusted the computer time expense to match ReplicaDock's expense. We have generated about 120,000 decoys with shotgun sampling for each target. Decoys with *interchain_contact* >10 are discarded and the top 40,000 in energy are selected for analysis or refinement. (*Appendix* Method S1: protocol_capture/rosetta_dock/)

2.2.4 ReplicaDock (alternative low-resolution stage)

As an alternative to the shotgun sampling described above we applied here a replica exchange procedure [96,97]. Inverse temperatures are set to, β, of $2^{-1}\text{kcal}^{-1} \cdot \text{mol}$, $3^{-1}\text{kcal}^{-1} \cdot \text{mol}$ and $5^{-1}\text{kcal}^{-1} \cdot \text{mol}$, and swaps are attempted every 1,000 Monte-Carlo steps [99,100]. 4 trajectories with 3 temperature levels are run for 5×10^6 Monte-Carlo steps, and snapshots are stored every 1,000 steps. In total, 60,000 decoys are generated for each target with this protocol. For further analysis or refinement, the highest temperature level (inverse temperature $5^{-1}\text{kcal}^{-1} \cdot \text{mol}$) and decoys with *interchain_cen* >10 were excluded. Initial configurations of trajectories were generated as described in Section 2.2.2. All targets of the benchmark are sampled with the same three temperature levels. The choice of temperature levels is discussed in Results Section 2.3.2. For all targets good exchange rates (~25%) are achieved and no further target dependent optimization is required.

To avoid unbounded diffusion of the two binding partners away from each other, we generated an *encounter constraint*. This constraint is realized as flat-bottom distance

restraint between the C_α-atoms closest to the center of mass of the respective binding partners. These center atoms are denoted in the following as $C_\alpha^{\text{cen}}(i)$ with $i = 1,2$. The constraint does not penalize the conformations unless they are further than $d_{\text{lim}} = s_1 + s_2 + g$ apart, where g denotes the chosen gap parameter (here 8Å), and s_i denotes the furthest distance of a surface C_α-atoms of binding partner i to its center $C_\alpha^{\text{cen}}(i)$. For distances $d > d_{\text{lim}}$ the harmonic penalty energy $V_{\text{enc}} = k(d - d_{\text{lim}})^2$ is applied. (*Appendix* Method S2: protocol_capture/replica_dock/centroid)

2.2.5 ZDOCK

To prepare for ZDOCK, protons are removed and the individual binding partners are marked using *mark_sur* from ZDOCK's toolbox. 54,000 decoys are generated with ZDOCK3.0.2 for each target and top 36,000 decoys by ZDOCK score are evaluated and refined (Section 2.2.6).

2.2.6 Refinement (high-resolution stage)

The high-resolution stage of RosettaDock described by *Gray et al.* [36] was applied without alteration and used to refine conformations generated with shotgun, ZDOCK or ReplicaDock. (*Appendix* Method S3: protocol_capture/rosetta_dock/refine and Method S4: protocol_capture/replica_dock/refine)

2.2.7 Construction of Benchmark

30 Targets were selected from the Dockground Benchmark [47], for which the x-ray resolution of the bound complex is no worse than 2.5Å and the x-ray resolution of the individual unbound partner is no worse than 2.2Å. Since the current method does not allow backbone motion, we restricted the benchmark to targets where the monomer C_α-RMSD between bound and unbound structure is ≤ 1.5Å for both binding partners (*Appendix* Table S1).

2.2.8 Implementation of ReplicaDock in Rosetta3

We implemented replica exchange within the general Metropolis-Hastings framework of the Rosetta3 software package. The replica-exchange module is accessible through the *RosettaScripts* [101] interface and can be combined with any conformational moves that are implemented as children of the *ThermodynamicMover* class. To ensure detailed balance, either Movers have to yield unbiased conformational perturbations or they have to provide the proposal density

of the perturbation through implementation of an abstract virtual function in the *ThermodynamicMover* interface. For docking we have provided the *ThermodynamicRigidBodyPerturbNoCenterMover* which performs unbiased rotational and translational moves. Random translations drawn from Gaussian distribution are performed along all three axes. The axis-angle notation is used to represent rotations. A rotation axis is generated using

$$\alpha = 2\pi \cdot x$$

$$\beta = \cos^{-1}(1 - 2 \cdot y)$$

$$\vec{V} = (\sin\beta \cdot \sin\alpha, \sin\beta \cdot \cos\alpha, \cos\beta)$$

where x, y are randomly drawn from uniform distribution. This is sufficient to guarantee unbiased rotational sampling, and a distribution for the positive rotation angle can be chosen freely. In order to be consistent in rotational step sizes with the parameterization of the original RigidBodyPerturbNoCenterMover, we first draw Euler angles from a Gaussian distribution with specified magnitude, and then transform the resulting rotation into axis-angle representation. Second, we combine the rotation axis obtained using our unbiased sampling method, with the rotation angle that corresponds to the Gaussian Euler angles. The resulting distribution of rotation angles is shown for different parameters in *Appendix* Figure S1.

Additionally, we have implemented the *DockSetupMover* to select the rigid-body degrees of freedom to be sampled (via *FoldTree*) and implemented the *parse_my_tag* method of the existing *DockingInitialPerturbationMover* [77] to render it accessible through the RosettaScripts interface.

2.2.9 Metrics for structural accuracy and docking performance

The metrics interface RMSD (I_rms), ligand RMSD (L_rms), fraction of native contacts (f_{nat}) and fraction of non-native contacts ($f_{non\text{-}nat}$) are defined as in CAPRI [102] and are calculated against the bound complex.

For the low-resolution ensemble we consider a decoy with I_rms ≤ 2.5Å as 'hit'.

2.2.10 Sampling the native energy basin

Additionally, we were interested in the native energy basin accessible by the applied fixed-backbone, flexible-sidechain docking protocol from the unbound monomeric starting structures. Accordingly, we started 1000 trajectories of refinement (Section

2.2.6) from the unbound monomers superimposed onto the bound complex to generate *RelaxedNative* ensembles. (*Appendix* Method S5: protocol_capture/relax_native)

To assess how well a protocol samples the native energy basin, we count conformations that overlap with the RelaxedNative ensembles as follows. A lower left region in the interface energy vs. I_rms plots relative to the RelaxedNative ensemble was defined by the 50%-tile interface energy of RelaxedNatives as upper confinement and by the 75%-tile I_rms of the RelaxedNatives as the right confinement (*Appendix* Figure S2). All conformations in this region are counted.

We define 4 categories based on the number of conformations n that overlap with the RelaxedNative ensemble: none (0), magic points ($0 < n \leq 5$), sporadic ($5 < n \leq 20$), dense ($20 < n$).

2.2.11 Clustering after all-atom refinement

The top-2000 conformations by interface energy after all-atom refinement were clustered using a cutoff of 5.0Å of C_α-atoms RMSD. Clusters are ranked by size and are represented by the decoy with lowest interface energy within the cluster.

Ranking clusters by interface energy resulted in slightly worse performance when evaluated using CAPRI criteria (*Appendix* Table S2).

2.2.12 Automated Setup

The automated setup tools available with the CS-Rosetta toolbox (www.csrosetta.org) have been used to generate all production runs of the benchmark. We advise users to install this toolbox from *protocol_capture/ 2012/ replica_docking/ csrosetta3* or from the website. If installed from the website, the docking plugins *_docking_base, rosetta_dock* and *replica_dock* have to be copied into the *flag_library/ methods* folder from */protocol_capture/ 2012/ replica_docking/ csrosetta3/ flag_library/ methods*. All methods presented in this work are implemented as plugins in *csrosetta3/flag_library* and can be accessed through the –*method* option of the *setup_xxx* commands. Example usage of these tools for docking can be found in the protocol capture section (*Appendix* Figure S3, *Appendix* Method S7: Automated Setup) and general documentation is provided at www.csrosetta.org/manual.

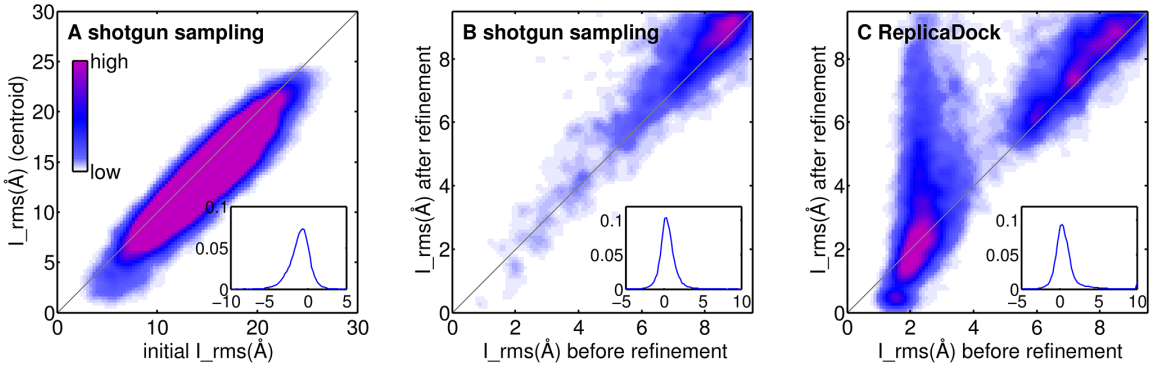

Figure 2.1. Detailed analysis of individual docking stages on bound target 1sq2. A) Interface RMSD (I_rms) before and after the Monte-Carlo optimization in the low-resolution stage of RosettaDock's shotgun sampling, B–C) I_rms before and after all-atom refinement for shotgun and ReplicaDock sampled ensembles, respectively. The colorbar indicates the density of data points at given position of the scatter plot, A–C) use a same colorbar range. The insets show the distribution of differences between I_rms after and before the respective sampling stage has been applied (negative values reflect an improvement in I_rms).

2.2.13 Computational Cost

For shotgun sampling and ReplicaDock the same amount of computer time was used. The amount of computer time required depends on the size of the binding partners and ranges from 66 core-hours to 900 core-hours on 2.6 GHz AMD Opteron Processors. ZDOCK requires significantly less computer time, with 1.25-7.25 core-hours on the same machine. We refined 40,000, 36,000 and ~36,000 conformations for shotgun, ZDOCK and ReplicaDock, respectively. Refinement requires between 14-120 core-seconds per conformation, again depending strongly on the protein sizes.

2.3 Results

2.3.1 Shotgun sampling is dominated by initial random placement

In this section we address the question how conformational space is explored in the shotgun approach of RosettaDock. Shotgun sampling consists of a random initial placement followed by a short energy optimization procedure using 500 steps of Monte Carlo sampling with adjusted step-sizes and constant inverse temperature of $\beta = 0.8^{-1}\text{kcal}^{-1} \cdot \text{mol}$ (Methods).

The short Monte-Carlo sampling explores only a small region of conformational space around the respective starting position, which is reflected by a strong correlation between I_rms to the native complex structure before and after the Monte Carlo optimization (Figure 2.1A). Thus, the generated ensemble of structures

is highly biased by the initial starting structures, rather than by the low-resolution energy (Figure 2.2C+G; *Appendix* Figure S4C+G).

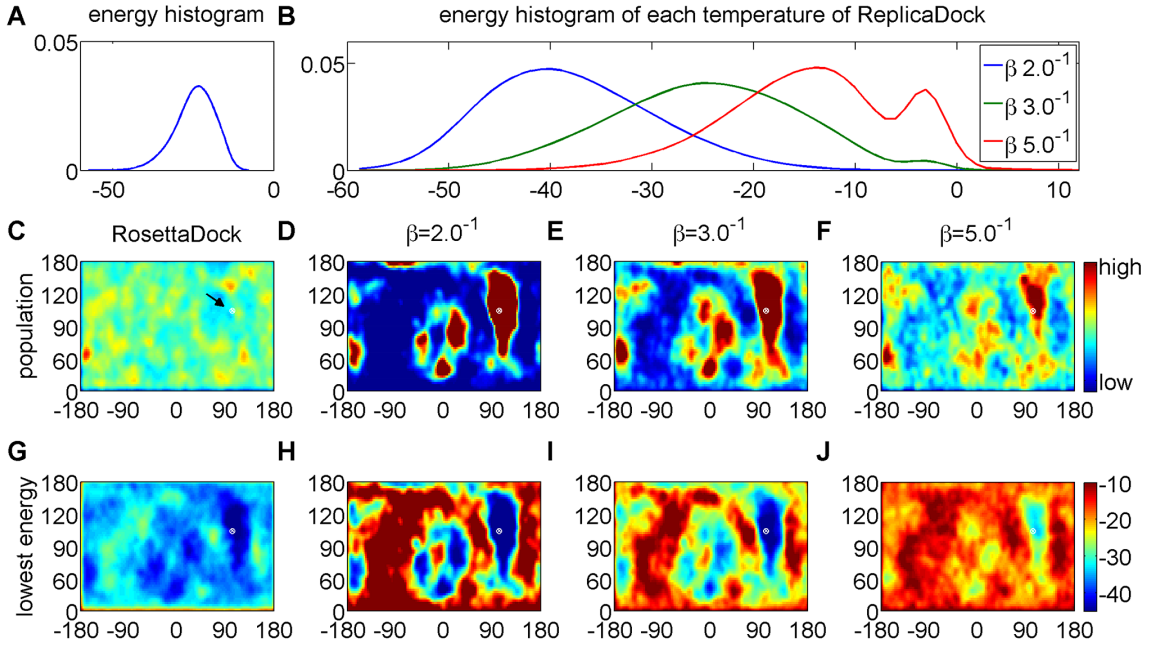

Figure 2.2. Detailed analysis of shotgun and ReplicaDock sampling on target 1ppf. A) energy distribution of shotgun sampling generated low-resolution decoys. B) energy distribution of conformations sampled by ReplicaDock at respective inverse temperatures. C-F) Population of sampled conformations in spherical coordinates. Partner A is fixed at the center and the position of Partner B with respect to an idealized spherical surface around Partner A is recorded. The native structure is labeled as white dot (arrow in C). G-J) Conformations are assigned to grid-cells as in C-F, but shown is the lowest energy of all conformations assigned to the respective grid cell. The same color-scale is used for each plot of a row, and the colorbars are attached to the rightmost panel.

Subsequently, the decoys of the low-resolution phase are refined using an all-atom model and allowing side-chain flexibility on top of the rigid body motion. The energy landscape in this phase is very rugged and the applied refinement protocol is not very explorative. Indeed, the I_rms of low-resolution input structures and all-atom refined structures are also strongly correlated (Figure 2.1B). Thus, a necessary requirement for the prediction of the native complex structure is that the initial random placement already generates sufficiently many near-native conformations.

2.3.2 Sampling with ReplicaDock generates energy-biased populations

The previous section showed that the relative populations of conformations in the shotgun ensemble are dominated by the initial perturbation and do not reflect the low-resolution energy landscape. Nevertheless, differences in energy are apparent and the native structure is found in one of the low-energy basins (Figure 2.2G).

We argued that Importance Sampling might generate ensembles of better quality with more near-native conformations that are better suited for continuation in the high-resolution refinement stage. Accordingly, we employ unbiased rigid body moves and replica exchange Monte-Carlo (REXMC). The temperature levels are chosen such that the lowest and highest temperatures reflect the bound and unbound state, respectively.

To achieve efficient exchange rates between replica's, the energy distribution at any given temperature level has to overlap with the energy distributions of neighboring temperature levels. Generally, in applications of replica exchange to biomolecular systems this can only be achieved by employing a large number of temperature levels combined with careful optimization for each individual system [103-105]. Here, however, only the six rigid body degrees of freedom are sampled rendering energy distributions rather broad, such that only a small number of temperature levels is required [80]. In an initial survey, simulations on bound target 1sq2 were carried out with inverse temperatures, β , $0.6^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $0.8^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $1^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $1.2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $1.5^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $2.5^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $5^{-1}\text{kcal}^{-1}\cdot\text{mol}$ and $10^{-1}\text{kcal}^{-1}\cdot\text{mol}$ (Section 2.2.3, *Appendix* Figure S5). We found that the lowest temperature levels lead to sharp energy distributions centered at relatively low energies. Inspection of the corresponding conformations reveals predominantly large number of contacts consistent with bound conformations. Another peak in the energy distribution is found at the higher temperature levels, and inspection of the corresponding conformations shows that binding partners have little or no contact. Inverse temperatures between $2^{-1}\text{kcal}^{-1}\cdot\text{mol}$ and $5^{-1}\text{kcal}^{-1}\cdot\text{mol}$ yield broad distributions that connect the low-energies of bound conformations with the sharp energy-peak of unbound conformations. In our initial tests, this temperature range turned out to be most suitable to broadly sample different binding modes, and we chose the inverse temperatures ($2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $3^{-1}\text{kcal}^{-1}\cdot\text{mol}$ and $5^{-1}\text{kcal}^{-1}\cdot\text{mol}$) for the final protocol. This choice yields consistent results across the full range of benchmark cases in terms of energy distributions (*Appendix* Figure S6) and exchange rates around 25% between replica's with little variation between targets (*Appendix* Table S3).

Replica Exchange with inverse temperatures $2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $3^{-1}\text{kcal}^{-1}\cdot\text{mol}$ and $5^{-1}\text{kcal}^{-1}\cdot\text{mol}$ is show-cased at the example of target 1ppf in Figure 2.2. Indeed, ReplicaDock achieves good overlap in the energy distributions (Figure 2.2B) and frequent exchanges (*Appendix* Figure S7). Obviously, at high temperature the

unbound (non-contacting) conformations are more populated, whereas at low temperatures bound (contacting) conformations are more populated. Thus, thermodynamically speaking, the high-temperature state reflects the unbound state, whereas the low-temperature state reflects the bound state, and the replica exchange scheme achieves frequent exchange between both.

Independent trajectories started from different random conformations converge to the same populations (*Appendix* Figure S8) demonstrating that convergence is achieved within the simulation time. Accordingly, all generated trajectories can be combined to improve statistics (Figure 2.2D-F). As expected from Importance Sampling, areas of high population in the ReplicaDock ensemble coincide with the low-energy regions in the conformational energy landscape (Figure 2.2H-J), which is in stark contrast to the shotgun ensemble.

Surprisingly, with the lowest temperature set to $\beta = 2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, ReplicaDock achieved significantly lower energies than shotgun sampling with temperatures set to $\beta = 0.8^{-1}\text{kcal}^{-1}\cdot\text{mol}$. This demonstrates that significantly more than 500 Monte-Carlo steps, as employed in the shotgun protocol, are required to consistently equilibrate at the chosen temperature.

2.3.3 Only near-native conformations are pulled into the native energy funnel in the refinement stage

Refinement of the ReplicaDock ensemble results in a dramatically different behavior than refinement of the corresponding shotgun ensemble (Figure 2.1B+C). When conformations below I_rms 3Å are refined, they are likely to be pulled into the native energy funnel, resulting in I_rms between 0 and 1Å (Figure 2.1B+C). Decoys further than 4Å do not systematically improve their I_rms in refinement.

Since the shotgun centroid ensemble has a very low number of conformations with I_rms ≤ 3Å refinement is less beneficial for shotgun than for ReplicaDock ensembles. However, for ReplicaDock decoys, some refinement trajectories also seem to move away from the near-native regions. This seems to be caused by clashes after switching to the high-resolution representation (data not shown). As discussed later (Section 2.3.8), the centroid energy is dominated by *interchain_cen* which may cause some overly compact conformations.

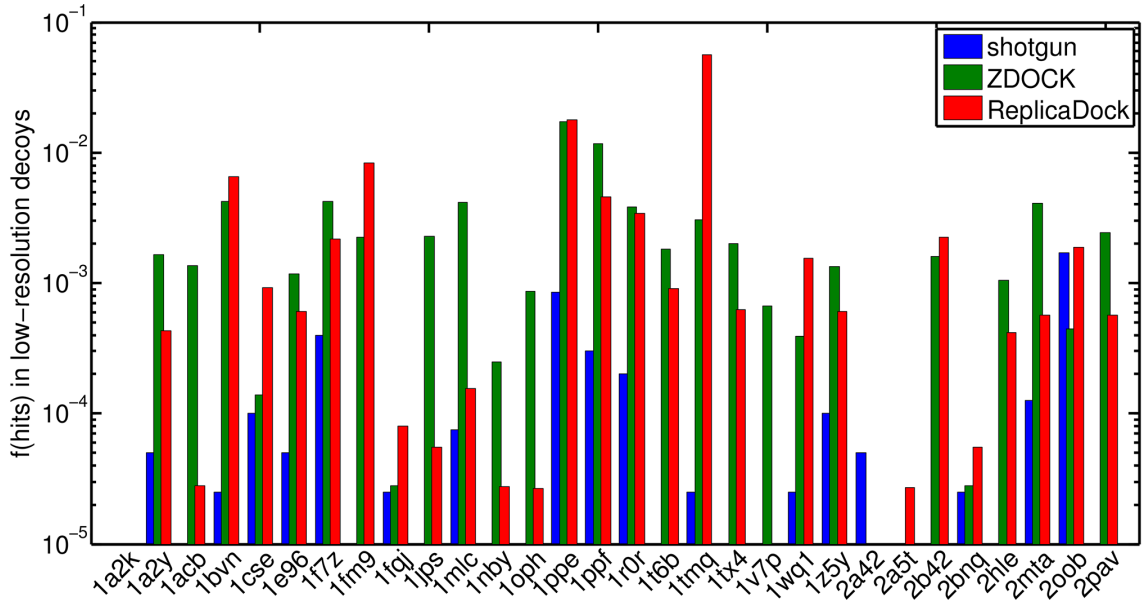

Figure 2.3. Fraction of hits in low-resolution docking. Conformation with I_rms ≤ 2.5Å to the bound complex are considered as a hit (Section 2.2.9). Blue, green and red represent the results of shotgun sampling, ZDOCK and ReplicaDock, respectively.

2.3.4 ReplicaDock efficiently samples near native conformations

In the previous section, we showed dense sampling of near native conformations during the low-resolution phase is a necessary condition for identifying the native energy basin in the refinement phase. In the following, we analyze how successful the shotgun approach and ReplicaDock sample near-native conformations in the low-resolution phase on a benchmark of 30 protein-protein complexes (Methods), where the docking partners are given as PDB-structures of the unbound proteins.

We applied ReplicaDock with inverse temperatures, β, $2^{-1}\text{kcal}^{-1}\cdot\text{mol}$, $3^{-1}\text{kcal}^{-1}\cdot\text{mol}$ and $5^{-1}\text{kcal}^{-1}\cdot\text{mol}$ for low-resolution sampling. For ReplicaDock, 4 trajectories with 5×10^{6} MC-steps were run using different amounts of computer time depending on the target size. The same amount of computer time per target was given to shotgun sampling by adjusting the number of generated decoys *(-nstruct)*, accordingly.

As shown in Figure 2.3, ReplicaDock generates significantly higher fractions of near-native decoys for nearly all targets than shotgun sampling. In three cases (1a2k, 1v7p, and 2a42) ReplicaDock did not generate any near-native decoys. Scrutinizing the targets in detail, we find that for 1a2k a few residues on the terminus of the receptor stick into the binding pocket and cause clashes between backbone atoms in the superimposed structures. Similarly, in complex structure 1v7p and 2a42 loops

in the interface area deviate substantially from the position they occupy in the unbound monomer conformation. Without these backbone changes crucial side-chains cannot be packed without clashes in the interface, causing the binding partners being pushed apart. In these cases shotgun sampling also produced only a small fraction of near-native decoys. Obviously, these problems cannot be overcome with rigid-body docking, and thus explain the bad performance of the methods.

Additionally, if we impose further energy cut-offs and only keep the low-20k, low-5k or low-2k of conformations (*Appendix* Figure S9), we find that a) for an increasing number of targets no near-native conformations are retained, and b) that the fraction of near-native decoys become more similar between shotgun and ReplicaDock sampling. This shows that ranking by low-resolution energy is problematic and only generous filters should be applied before refinement.

In summary, ReplicaDock yields a dramatic improvement of near-native sampling for most targets. Some targets, however, are not improved or get slightly worse, and aggressive energy-cuts have a negative impact. Possible reason for these failures of the low-resolution energy function will be discussed further in Section 2.3.8.

2.3.5 Comparison to ZDOCK

As shown above, ReplicaDock yields ensembles with a higher fraction of near-native decoys than RosettaDock's shotgun sampling method. Another large class of docking programs uses enumerative sampling on a translational and rotational grid. Here, we run ZDOCK 3.0 as a representative of these programs. ZDOCK has been shown to be one of the most successful docking programs [106-108].

As shown in Figure 2.3, the performance of ZDOCK and ReplicaDock is comparable. For most targets both methods yield fractions of near-native decoys between 10^{-4} and 10^{-2}. Moreover, both methods fail in some targets: 2a42 and 2a5t for ZDOCK, and 1a2k, 1v7p and 2a42 for ReplicaDock. However, one should mention that ZDOCK is more efficient due to its FFT based sampling of translational degrees of freedom and requires only a small fraction (~1%) of the computational expense (Methods).

2.3.6 All-atom refinement of ReplicaDock ensembles

As shown above, ReplicaDock and ZDOCK yield ensembles with a higher fraction of near-native decoys than shotgun sampling. To analyze whether these improvements also have a positive impact on the all-atom refinement stage of the docking protocol,

we refined the shotgun, ZDOCK and ReplicaDock ensembles for all targets in the benchmark. For ReplicaDock and shotgun sampling we have made sure that ReplicaDock use always equal or slightly less overall computer time compared to the shotgun based approach. In any case the computational expense for refinement of any conformation, whether generated by ZDOCK, ReplicaDock or shotgun sampling is approximately the same.

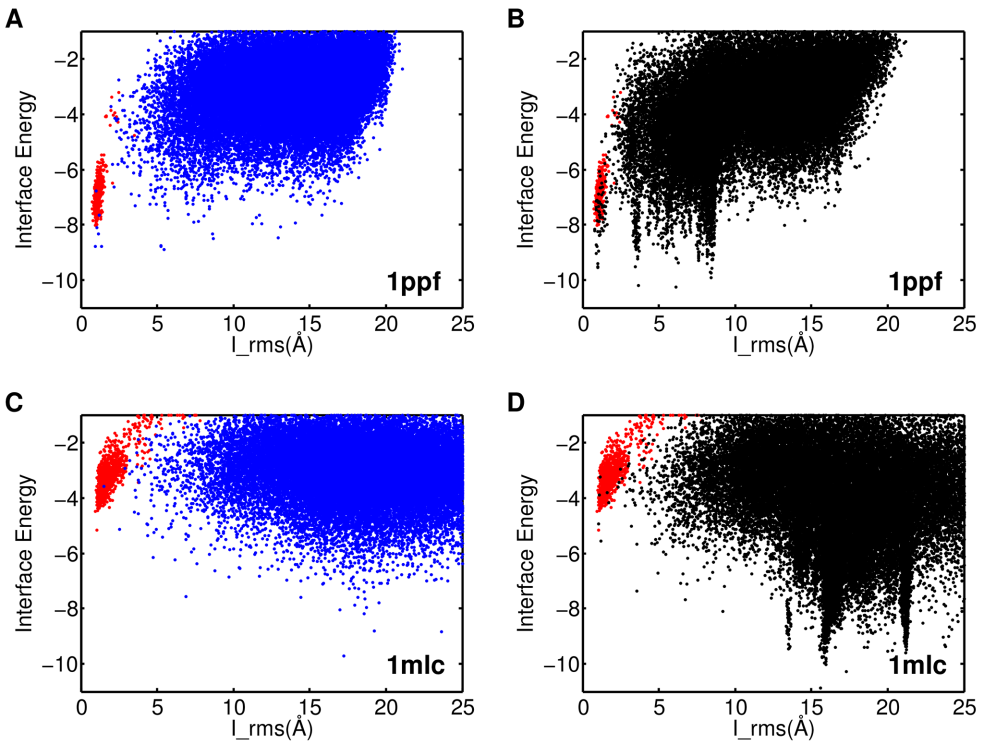

Figure 2.4. Interface RMSD vs. Interface Energy after refinement on target 1ppf and 1mlc. A) and C) refinement of shotgun sampling generated ensembles, B) and D) refinement of ReplicaDock generated ensembles. The red dots represent the RelaxedNative ensembles (Results).

Next we want to test the respective sampling strategies for their ability to detect the native energy basin of the all-atom energy function after refinement. To show the position and form of the native energy funnel, we also generated RelaxedNative ensembles (Methods) by starting multiple trajectories of refinement from unbound monomers superimposed onto the bound complex.

We first introduce the typical differences between refined shotgun and ReplicaDock ensembles at the example of target 1ppf and 1mlc. As shown in Figure 2.4A, the refined shotgun ensemble of target 1ppf (blue) is higher in energy than the RelaxedNative ensemble (red). Only 3-4 isolated conformations reach significantly lower energies. The ReplicaDock ensemble, on the contrary, shows three distinct

energy funnels that are well sampled (Figure 2.4B). One of the funnels coincides in form and position with the funnel formed by the RelaxedNative structures, demonstrating that the native energy basin has been found and is well sampled. But unfortunately it is neither the lowest nor the most pronounced energy funnel, rendering discrimination of native from non-native decoys challenging. Apparently, the Rosetta all-atom energy function features at least three well-resolved energy funnels for this target complex, which is confirmed by finding corresponding clusters that are well populated (data not shown). All three of these funnels remain poorly sampled with shotgun sampling. For target 1mlc the RelaxedNative conformations have higher energies than the shotgun or ReplicaDock ensembles (Figure 2.4C+D). This points to deficiencies of the energy function. However, also for 1mlc shotgun sampling produces only sporadic sampling of low energies, whereas ReplicaDock detects distinct funnels in the energy landscape.

The main observations for ReplicaDock for targets 1ppf and 1mlc are a) that much lower energies are sampled, b) that distinct energy funnels are sampled densely, and c) that for 1ppf the native energy funnel is sampled densely. Next, we ask whether similar differences in behavior between shotgun and ReplicaDock are observable for all 30 targets. Indeed, equivalent scatter plots of all targets (Figure 2.5) show similar differences between shotgun and ReplicaDock as already observed for targets 1ppf and 1mlc. To quantify, we computed histograms of the lowest energies sampled per target by the respective approaches (shotgun, ZDOCK, ReplicaDock and RelaxedNative). Whether we focus on the lowest 0.1%, 1% or 5% of decoys, energies of shotgun ensembles are higher for all targets, and even the RelaxedNative ensembles often do not reach energies as low as ReplicaDock (Figure 2.6). Energies of refined ZDOCK conformations are in-between those of ReplicaDock and shotgun. These results demonstrate that the conformations in the centroid ReplicaDock ensemble are well poised to reach low interface energies in the subsequent all-atom refinement.

For target 1ppf and 1mlc, ReplicaDock detected funnels in the energy landscape that were not revealed by shotgun ensembles. The same impression is reached from a comparison of the sampling methods on all targets (Figure 2.5). Refined ZDOCK ensembles are more comparable with refined ReplicaDock ensembles than shotgun ensembles, but also show slightly less distinct energy funnels overall (*Appendix* Figure S10).

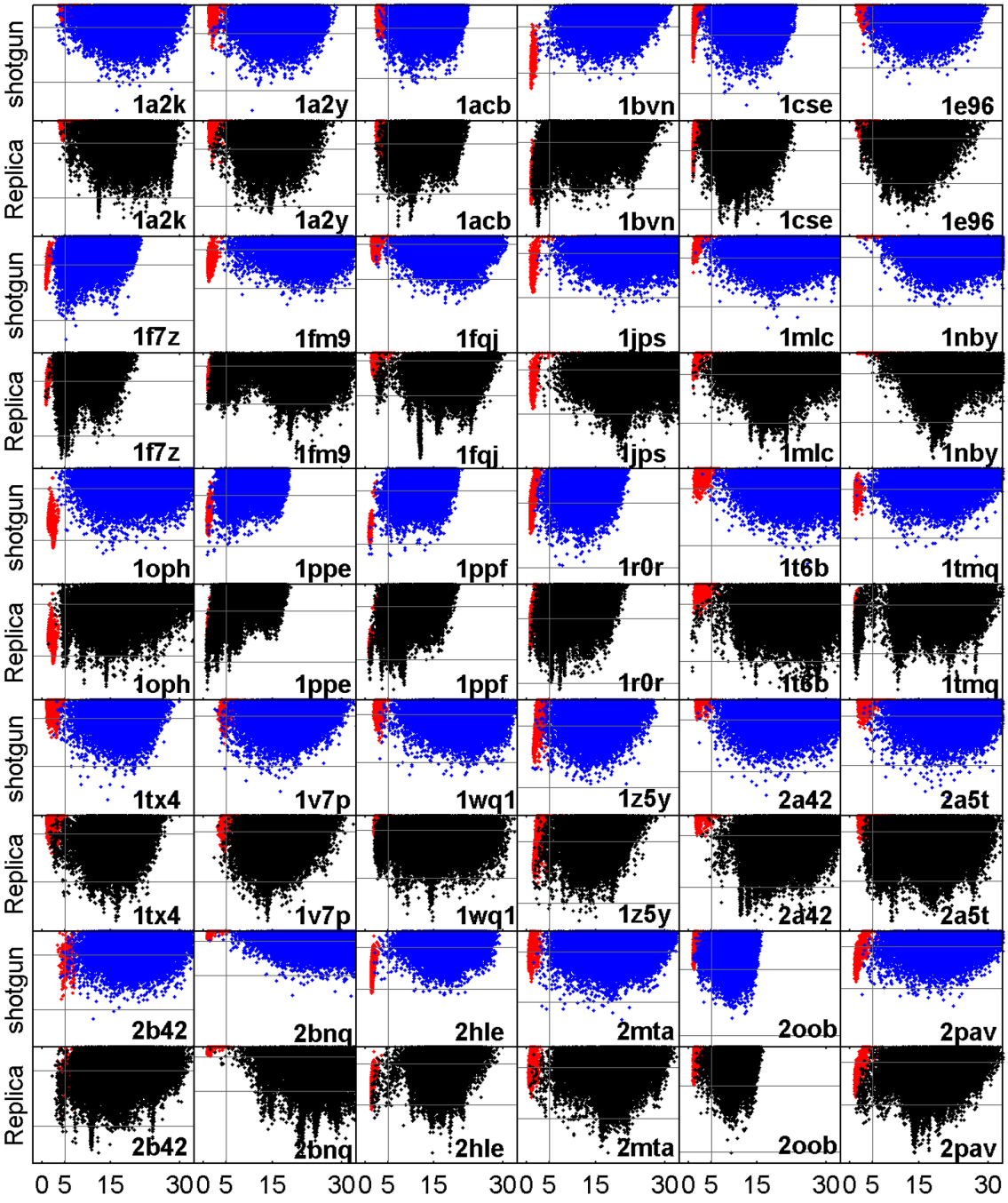

Figure 2.5. Interface RMSD vs. Interface Energy after refinement for all the 30 unbound docking targets. The red dots represent the RelaxedNative ensembles(Results). The interface RMSD is shown on the x-axis, the interface energy on the y-axis. The same energy range is used for displaying both, shotgun sampling (blue) and ReplicaDock (black), results of each target, respectively. The vertical gray lines correspond to I_rms of 5.0 Å, and the two horizontal gray lines correspond to interface energy -4 and -8 Rosetta Energy Units.

Table 2.1. Qualitative classification in ability to sample native energy basin

target	shotgun	ZDOCK	ReplicaDock
1a2k	dense	magic	dense
1a2y	magic	dense	sporadic
1acb	magic	dense	dense
1bvn	none	dense	dense
1cse	magic	magic	dense
1e96	sporadic	dense	dense
1f7z	none	dense	sporadic
1fm9	none	dense	dense
1fqj	magic	magic	sporadic
1jps	none	sporadic	none
1mlc	magic	dense	sporadic
1nby	sporadic	dense	dense
1oph	none	dense	magic
1ppe	sporadic	dense	dense
1ppf	magic	dense	dense
1r0r	magic	dense	dense
1t6b	magic	dense	dense
1tmq	magic	dense	dense
1tx4	magic	dense	dense
1v7p	dense	dense	dense
1wq1	magic	dense	dense
1z5y	sporadic	dense	dense
2a42	magic	none	none
2a5t	sporadic	dense	dense
2b42	sporadic	dense	dense
2bnq	magic	none	magic
2hle	none	sporadic	sporadic
2mta	none	dense	magic
2oob	sporadic	magic	magic
2pav	none	dense	dense
dense	2	22	19
sporadic	7	2	5
magic	13	4	4
none	8	2	2

At the end is the summary of each category achieved by each method.

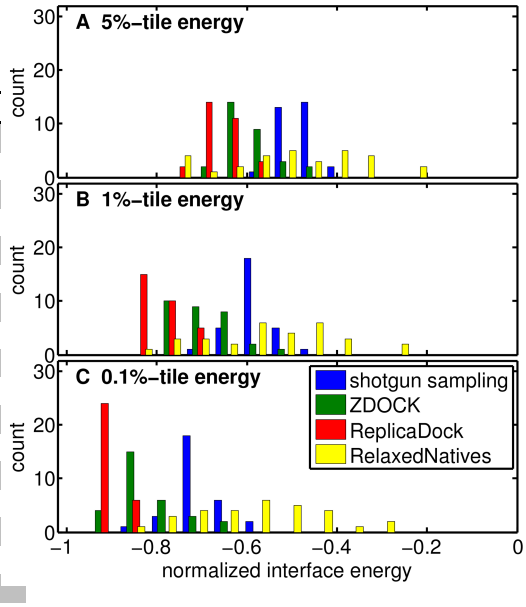

Figure 2.6. Interface energy distribution obtained with different simulation protocols. The energies for each target and method have been normallized to the dynamic range of interface energies observed for the respective target across all methods. The interface energies are normalized by the absolute value of mean energy of the 10 lowest observed energies for this target (highest energy is always 0). The x-percentile energy is the scaled energy value that separates off x% of the lowest energy decoys for a given simulation result. Shown are the distributions of x-percentile energies across all 30 targets for a) the 5%-tile, b) the 1%-tile and c) the 0.1%-tile, respectively.

We would like to quantify the performance of the two sampling strategies in detecting funnels in the energy landscape. However, *a priori* it is impossible to tell how many and which funnels should be detected by an optimal sampling strategy. The only funnel whose existence is known *a priori* is the funnel formed by the RelaxedNative ensemble (red). We thus restrict ourselves to ask whether overlap with the RelaxedNative ensemble is achieved. Model selection is often based on clustering, and it has been shown that cluster size is an important criterion for ranking that it is often found to be more reliable than energies [109]. Thus a relatively high population is crucial for the confident selection of a binding mode. Accordingly, we discriminate four cases (Methods): a) *dense*, the native funnel is densely sampled (1ppf; ReplicaDock) b) *sporadic,* the native funnel is sampled sporadically (1ppe; shotgun or 1a2y; ReplicaDock) c) *magic points,* individual decoy

structures coincide with the native funnel (1fqj; shotgun) and d) *none,* no decoy structures fall within the native funnel (2hle; shotgun).

Table 2.2. Summary of structure prediction accuracy in unbound docking.

	shotgun refined					ZDOCK refined					ReplicaDock refined				
Target	CQ[1]	L_ rms	I_ rms	f_{nat}	$f_{non-nat}$	CQ[1]	L_ rms	I_ rms	f_{nat}	$f_{non-nat}$	CQ[1]	L_ rms	I_ rms	f_{nat}	$f_{non-nat}$
1a2k	0	43.8	13.5	0.000	1.00	0	38.9	15.4	0.000	1.06	0	30.0	12.3	0.146	1.72
1a2y	0	21.7	12.4	0.000	1.29	0	22.9	7.0	0.068	1.45	0	24.5	7.6	0.023	1.27
1acb	0	17.5	7.6	0.014	0.66	*	9.8	3.9	0.145	0.89	**	4.9	2.7	0.333	0.58
1bvn	0	17.8	7.7	0.068	0.75	0	14.7	6.1	0.110	1.01	*	4.7	2.5	0.274	0.78
1cse	0	17.6	8.0	0.013	1.22	0	18.9	7.8	0.000	0.82	*	12.4	3.9	0.241	0.54
1e96	0	39.2	8.2	0.000	1.36	0	17.8	6.3	0.122	1.14	0	27.7	7.0	0.000	1.95
1f7z	0	17.7	3.8	0.098	0.70	0	14.4	3.7	0.066	1.06	0	17.7	4.2	0.197	0.83
1fm9	0	49.6	17.7	0.000	0.60	**	4.0	2.6	0.364	0.45	0	58.3	19.3	0.000	1.24
1fqj	0	40.8	12.3	0.000	1.28	0	43.1	13.3	0.000	1.15	0	27.7	11.9	0.000	2.25
1jps	0	28.5	7.4	0.014	0.80	*	13.2	3.2	0.254	0.62	0	25.4	10.7	0.000	1.07
1mlc	0	56.0	15.9	0.000	1.03	0	26.8	6.9	0.054	0.96	0	56.5	12.4	0.000	1.25
1nby	0	46.0	13.1	0.000	1.00	0	19.9	10.6	0.041	0.74	0	52.9	14.7	0.000	1.13
1oph	0	50.1	17.5	0.000	0.80	**	5.1	2.6	0.581	0.30	0	19.4	4.6	0.145	0.93
1ppe	***	2.7	0.95	0.746	0.08	***	1.8	0.79	0.761	0.19	***	2.11	0.88	0.775	0.11
1ppf	***	2.7	0.94	0.824	0.19	**	3.6	1.12	0.824	0.27	***	2.9	0.99	0.745	0.25
1r0r	0	15.1	7.1	0.106	0.62	*	10.4	2.76	0.394	0.42	*	9.2	2.47	0.455	0.39
1t6b	0	69.1	24.1	0.000	1.33	0	17.9	10.2	0.000	0.81	0	69.1	19.7	0.000	1.41
1tmq	0	19.8	11.9	0.026	0.77	**	4.9	1.4	0.579	0.44	**	4.5	1.36	0.618	0.51
1tx4	0	27.2	10.0	0.000	1.01	0	28.4	12.4	0.031	0.95	0	22.1	11.1	0.062	1.20
1v7p	0	26.6	14.6	0.016	1.08	0	24.0	8.6	0.145	1.08	0	23.5	11.9	0.048	1.29
1wq1	0	14.0	7.1	0.121	0.52	*	7.3	3.8	0.220	0.39	0	11.5	6.9	0.011	1.01
1z5y	*	4.1	1.9	0.273	0.42	*	6.2	3.2	0.333	0.50	0	10.2	5.7	0.076	0.69
2a42	0	42.3	12.1	0.000	1.25	0	82.2	23.7	0.000	1.50	0	51.9	12.8	0.021	1.66
2a5t	0	17.5	4.8	0.186	0.79	0	23.5	14.1	0.000	0.93	0	25.5	9.0	0.068	1.10
2b42	0	22.0	11.4	0.079	0.49	*	9.2	4.2	0.157	0.74	0	17.6	7.3	0.079	0.73
2bnq	0	74.9	29.6	0.000	1.18	0	57.2	16.5	0.000	1.97	0	70.1	14.8	0.000	2.34
2hle	0	24.9	12.8	0.000	0.77	*	6.6	2.9	0.464	0.40	0	36.6	10.3	0.083	0.61
2mta	*	7.3	3.3	0.457	0.63	**	10.1	2.5	0.543	0.56	0	47.2	17.2	0.000	1.54
2oob	*	8.4	3.4	0.000	0.70	0	11.0	5.5	0.000	0.85	0	11.5	7.0	0.000	0.96
2pav	0	34.0	13.7	0.000	1.15	**	3.4	1.45	0.068	0.33	*	7.65	3.00	0.421	0.87
summary	3*/2***					7*/6**/1***					4*/2**/2***				

Clusters are ranked by size and represented by the lowest interface energy decoy. In column 'CQ' (CAPRI Quality),
'0' indicates that none of the top 10 models was of accetable quality,
'*', '**' and '***' indicates that at least one of the top 10 models is of acceptable, medium or high quality, respectively (Section 2.3.7).
Columns 'L_rms', 'I_rms', 'f_{nat}' and '$f_{non-nat}$' record the respective information of the best model within these top 10 models.
[1] CQ refers to CAPRI quality

The individual classifications for shotgun, ZDOCK and ReplicaDock are shown in Table 2.1. In summary, we found 2 *densely* sampled native funnels for shotgun but 22 and 19 for ZDOCK and ReplicaDock, respectively. *No funnel* is found for 8 targets in shotgun ensembles, and for 2 targets in ZDOCK or ReplicaDock ensembles. For 13 targets, ReplicaDock improves the category by at least two steps, and for 11 targets

by one step. ReplicaDock and ZDOCK dramatically improves the ability to detect existing funnels in the energy landscape compared to the shotgun approach.

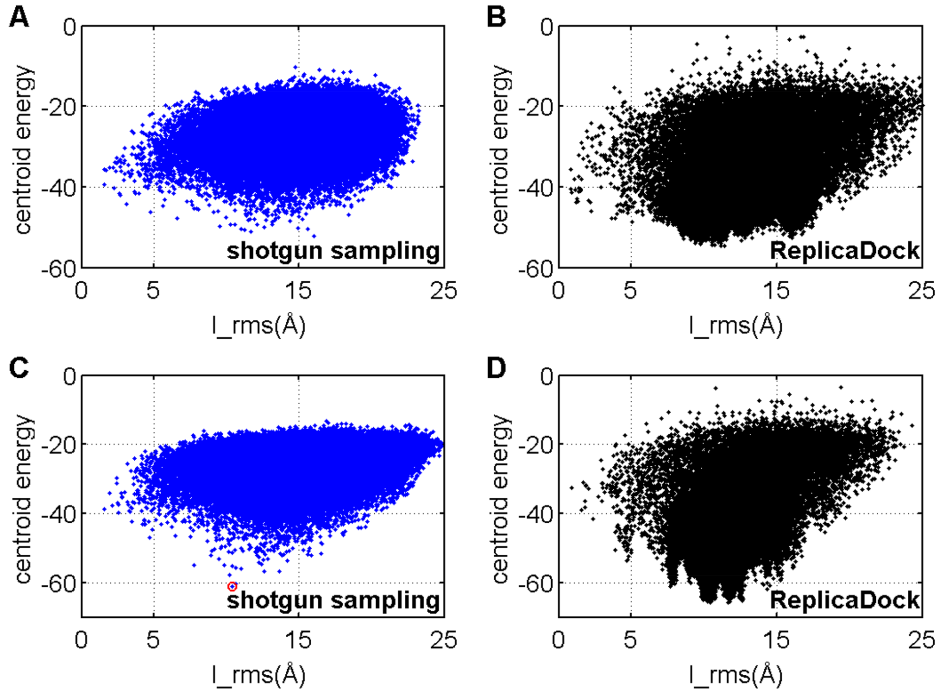

Figure 2.7. The centroid energy function prefers an alternative binding modes for the bound docking target 1emv. A-B) shotgun and ReplicaDock sample the low-resolution docking stage with 'capped' centroid energy function. C-D) shotgun and ReplicaDock sample the low-resolution docking stage with centroid energy function with no cap. The structure indicated by the red circle will be shown in Figure 2.8.

2.3.7 Performance in structure prediction / ranking

In previous sections we established that ZDOCK and ReplicaDock dramatically improve sampling in the initial docking stage compared to shotgun sampling, and refined ensembles of ReplicaDock reveal many more distinct funnel like features in the high-resolution energy landscape. However, we also found that the improved probing of the high-resolution energy landscape reveals many non-native energy funnels with lower energies than the native energy. Thus, it is not clear whether the improved sampling also leads to improved structure prediction in RosettaDock for unbound docking with relatively rigid targets as they were selected for our benchmark (Methods).

To address this question, we clustered refined ensembles and ranked clusters by size. From the ten largest clusters we selected the lowest interface energy model and evaluated using CAPRI criteria. As customary in CAPRI assessment we report metrics for the best of the ten models (Table 2.2). As a summary, ZDOCK and

ReplicaDock performed both better than shotgun in structure prediction, but ZDOCK displayed the overall best performance. In CAPRI, 0, 1, 2 or 3 stars are awarded for incorrect, acceptable, medium and high quality models [102]. Here, ZDOCK, ReplicaDock and shotgun generate acceptable or better models for 14, 8 and 5 targets, respectively. Indeed, the consistent appearance of non-native energy funnels in ReplicaDock ensembles obviously impedes concise selection of the native funnel despite its dense sampling.

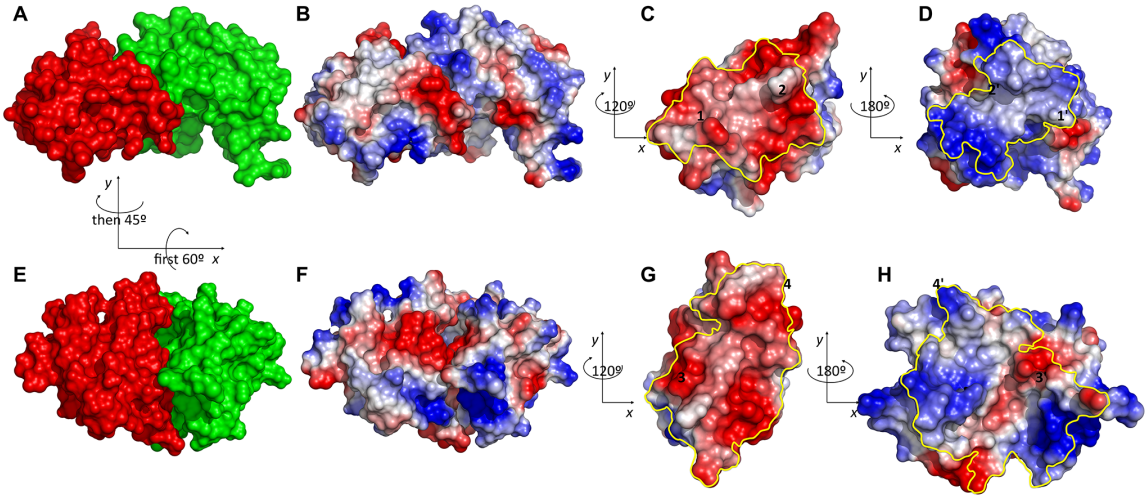

Figure 2.8. Electrostatic surface potential of native (A-D) and *interchain_cen* preferred conformation (E-H) in all-atom representation. A) native complex with receptor in green and ligand in red B) electrostatics map of the native complex C) electrostatics map of native ligand interface D) electrostatics map of native receptor interface E) *interchain_cen* preferred conformation after refinement F) electrostatics map G) electrostatics map of receptor interface H) electrostatics map of ligand interface. The yellow lines in C, D, G and H indicate the respective interface regions, and number pairs (e.g. 1 in C and 1' in D) indicate corresponding contact regions. Relations of viewing angle are given between panels where required.

2.3.8 Targets with insufficient near-native sampling

As shown in Section 2.3.4, for some targets none or only a few near-native decoys are sampled by both approaches. We found that in these cases the centroid energy function fails to stabilize the native conformation and/or over-stabilizes non-native conformations (Figure 2.7). In this section, we explore possible reasons for these failures of the low-resolution energy function.

The low-resolution energies are dominated by the *interchain_contact* term, which awards large contact surfaces (*Appendix* Figure S11). Indeed, all over-stabilized alternative binding modes that we found have significantly larger contact surfaces compared to the native complex (Figure 2.8). For example, 1emv features a spurious binding mode with significantly larger buried surface area of $1926.39 Å^2$ than its

native state 962.19Å^2 (Figure 2.8). Buried surface was calculated from the low-resolution models using the POPSCOMP sever [110,111]. Performing additional docking runs starting from the bound monomers we ruled out the possibility, that lack of backbone flexibility prohibits full stabilization of the native conformation .

A qualitative comparison of the electrostatic potentials at the native and alternative binding interfaces reveals a possible energetic counter-weight to the dramatic difference in buried surface area. Whereas the native state binding mode consists of surface patches of complementary charges, the alternative binding mode would superpose surface patches of equal charge. A more quantitative analysis is necessary, but this observation suggests that the electrostatic interactions are not captured sufficiently well by the empirical *interchain_pair* and *interchain_env* potential terms.

2.4 Conclusions

In this study, we introduced ReplicaDock, which uses temperature replica exchange to switch between bound and unbound thermodynamic states, and benchmarked its performance for sampling the low-resolution stage of protein-protein docking in RosettaDock.

For most targets tested in our benchmark, ReplicaDock reached significantly lower energies and generates drastically higher fraction of near-native decoys than shotgun sampling. After refinement, ReplicaDock-generated decoys reach lower interface energies and reveal funnel-like features of the energy landscape that are hardly visible when shotgun sampling is applied. The new funnel-like features seem to reflect distinct binding modes. Most importantly, the native energy funnels are often exactly matched in shape and position by ReplicaDock.

Enumerative search of rigid body degrees of freedom as carried out by ZDOCK 3.0 yields similar results as ReplicaDock. It also yields a higher fraction of near-native decoys and lower energies after refinement with Rosetta when compared to shotgun sampling. But ZDOCK does not reveal quite as many distinct funnels in the high-resolution Rosetta energy landscape as ReplicaDock. This is an advantage when the goal is to predict complex structures with the correct energy function. However, to improve methods further, it is important to have sampling methods that can identify *all* low-energy regions. We expect that this and other improvements in sampling of the Rosetta energy landscape [112,113] will help to iteratively improve the energy function until non-native conformations will no longer obtain spuriously low energies.

As expected, Importance Sampling, here in the form of replica exchange Monte Carlo, is more susceptible to spurious low energy regions than the shotgun approach. However, our results also show that this disadvantage is far outweighed by the much improved quality of final ensembles. Moreover, the improved sampling of the new method will allow to thoroughly probe the docking energy landscape, and thus to improve energy functions of both low- and high-resolution stage.

Computationally, the original shotgun approach has the advantage that it runs in an embarrassingly parallel fashion and thus can utilize more computer power in the same period of time than ReplicaDock. Indeed, the ReplicaDock trajectories generated here, required 5h-72h of computing on current hardware (using 12 processors), whereas the shotgun sampling can in principle be carried out in a few hours, if thousands of processors are used in parallel. ZDOCK uses an FFT based search of the translational degrees of freedom, rendering it computationally efficient. Only about 1-7 core-hours are required in total, which is <1% of the computational expense for shotgun or ReplicaDock sampling. Additionally, ZDOCK's search could in principle be re-implemented to support an embarrassingly parallel scheme, too. The FFT based search, however, requires grid-based energy functions, and thus is more challenging or even limiting in the possibilities to model interaction energies and to incorporate experimental data.

Despite of the drastic improvement of ReplicaDock to sample low energy structures and to recover near-native basins, there are still a few cases in which ReplicaDock samples very few or even no near-native conformations in the low-resolution stage and thus fails to recover near-native basins after refinement. In these cases, sampling is led astray by alternative binding modes with dramatically increased buried surface area. As it might be difficult or even impossible to ever balance out different contributions to the binding energies, especially at low-resolution and without better treatment of electrostatics (Figure 2.8), it seems advisable to develop energy functions that are globally flat but locally discriminative, in the sense that well contacting conformations are stabilized regardless their overall buried surface area, whereas miss-aligned conformations with bad shape complementarity are dis-favored. Unfortunately, this is not achieved by simply capping the energy function at a certain cutoff, as this quick fix removes local differences, too. Experimental data, or a higher-resolution energy function, can then be used to discriminate native from non-native conformations.

In this study we benchmarked the performance of replica exchange sampling in RosettaDock. RosettaDock performs a rigid-body minimization followed by all-atom

refinement like many other docking programs. The developed method and the conclusions derived from the presented benchmark should thus transfer well to other programs.

2.5 My contribution to this project

My contribution in this work include carrying out all the experiments and analysis, implementing part of the ReplicaDock protocol into Rosetta, implementing part of the automated setup tool for benchmarking into CS-Rosetta toolbox, and participating in the paper writing.

Chapter 3 Application of Enhanced Sampling Monte Carlo Methods for High-Resolution Protein-Protein Docking in Rosetta[§]

3.1 Introduction

Protein-protein interactions are integral to many mechanisms of cellular activities, ranging from enzyme catalysis and inhibition to signal transduction and gene regulation. Atomic-level structural knowledge is essential to understand the function of protein-protein complexes in biological processes. However, experimental structure determination of protein-protein complexes is often difficult and for many interactions the corresponding complex structures are lacking [82,114]. Computational protein-protein docking methods can provide structural models of protein-protein interactions where experimental data is absent, of low-resolution or too sparse. Besides providing valuable structural biology information, high-resolution protein-protein docking can also help to explain binding affinities and specificities, the nature of the binding free energy funnel and effects of mutations. Furthermore, these techniques are essential for computational protein-protein interface design of the design of non-natural complexes [82,114-119]

Docking programs often employ a two-stage protocol [91,120]. First, the space of putative docking geometries is sampled broadly, keeping the partner structures rigid, which corresponds to six degrees of freedom. Second, structures are refined in one or multiple steps, typically employing partner structures at atomic resolution allowing for conformational changes of side chains and possibly also of the protein main chain. For the rigid-body stage, the application of grid-based fast Fourier transformation [24] or geometric hashing [69] can accelerate the search [91,120]. Alternatively, the search can also be performed efficiently at reduced resolution using a coarse-grained model of the protein structures [36,37]. Both, at the rigid-body search stage or during refinement, Monte Carlo (MC) methods can be very helpful [38,39,41,76,121].

For high-resolution refinement, most protocols require a starting configuration that is already fairly close to the native structure. The likely reason for this strong dependence on the starting structure is the energy landscape's ruggedness, which frustrates the sampling and renders the global energy minimum hard to reach [122]. Accordingly, the rationale behind typical two stage docking refinement protocols is

[§] The work presented in this chapter has been published [51]

that the initial docking predictions is likely to generate at least one structure close to the native. A justification of this assumption is given by the general hypothesis that the native conformation coincides with the global energy minimum at the bottom of a broad basin in a rugged energy landscape [123,124]. However, since initial docking stages often use a simplified energy function, the initial docking stage is routinely misled and does not actually produce sufficiently many near-native candidates for the subsequent refinement stage, especially when there is an alternative binding site with larger buried surface [41].

Focus of this work is to improve the sampling for high-resolution docking based on the MC approach. In the MC method, random translational and rotational moves or conformational changes of the partner structures are applied on the configuration in a step-wise manner using the Metropolis criterion for acceptance of a move. Advantages of the MC method are the generation of a physically relevant canonical ensemble of docking configurations, use of arbitrary energy functions that can contain discontinuities and for the possibility to include various levels of structural flexibility. However, an exhaustive high-resolution sampling of the conformational space with the MC method can be computationally demanding. In general, the docking success of MC docking is limited by the sampling of putative complex geometries and by the accuracy of the energy function used for scoring predicted complexes.

Parallel tempering or replica exchange techniques promise to overcome these challenges and have received wide-spread interest in recent years [96,97,125,126]. The general idea of parallel tempering is to simulate the system with M replicas at different temperatures and to frequently exchange configurations between neighboring replicas. The high temperature replicas sample broadly, whereas the low-temperature replicas allow precise exploration of deep energy minima. Due to the frequent exchanges between the (hot) broad sampling regime and the (cold) annealing regime, configurations are less likely to get trapped in local minima. A generalization of temperature replica exchange is to vary the Hamiltonian (H-REMC) among replicas [80], which allows, for instance, to blend between a smoothed van der Waals potential and a realistically hard formulation to allow overcoming of sampling barriers in molecular dynamics simulation [127,128]. For Rosetta, previous studies also showed that softening the Lennard-Jones repulsive term is beneficial and better suited for side-chain modeling and prediction [129,130]. Of course, it is possible to combine variation of temperature and Hamiltonian in multi-dimensional replica exchange approaches [131,132]. A bottleneck in using replica exchange is that to cover the same parameter range (temperature, or smoothness)

the number of replicas required increases quickly with the number of degrees of freedom sampled. This is due to the fact, that to achieve efficient exchange between replicas, a substantial overlap between sampled energy levels of neighboring replicas is required [133,134].

Metadynamics is another popular enhanced sampling method, in which sampling is facilitated by a history-dependent biasing potential. It is constructed as the sum of Gaussian functions deposited along the trajectory in the collective variable space [135,136]. Choosing energy as collective variable gives rise to the so-called well-tempered ensemble (WTE) with much larger fluctuations in the sampled energies than the canonical ensemble [81]. This property of the WTE can be exploited to overcome the major bottleneck of temperature replica exchange discussed above. Since, the overlap of the energy distribution between neighboring replicas controls the exchange efficiency, using WTE drastically reduces the number of replicas required [81,137].

In the present study we have compared the efficiency of a standard MC protocol for high resolution protein-protein docking using *RosettaDock* and various extensions based on advanced sampling techniques. In particular, we tested four different protocols, standard Monte Carlo (MC), Temperature Replica Exchange Monte Carlo (REMC), well-tempered ensemble temperature Replica Exchange Monte Carlo (WTE-REMC), and well-tempered ensemble two dimensional Hamiltonian Replica Exchange Monte Carlo (WTE-H-REMC). The approaches were systematically evaluated on protein-protein complexes using unbound partner structures and starting in each case from the same starting placements. Overall best performance was achieved with the WTE-H-REMC method at the same computational cost compared to the alternative protocols.

3.2 Methods

3.2.1 Energy scoring function and starting structure generation

The standard all-atom energy function for *RosettaDock* as given by the weight-set *docking* [77] was used in all docking protocols. The docking energy function consists of van der Waals attractive and repulsive interactions, an implicit solvation term, hydrogen-bonding energies, a statistical residue-residue pairwise interaction term, a rotamer probability term and a pairwise electrostatic energy term [5,121]. For each target, the different docking simulation protocols were started from the same initial protein partner arrangements. The start geometries were based on unbound partner structures and one partner was initially separated relative to the position in

the complex in a random direction by 15 Å and a random rotation by 60° relative to the bound geometry. Only geometries without steric overlap between partners were accepted. The ligand RMSD (L_rmsd, root mean square deviation of the mobile protein after best superposition of the receptor protein onto native complex structure) from the respective bound complex for all the targets was on average ~18 Å with slight variation depending on the shape and size of the binding partners (Figure 3.1). The initial placement corresponds to a scenario where the binding region is approximately known.

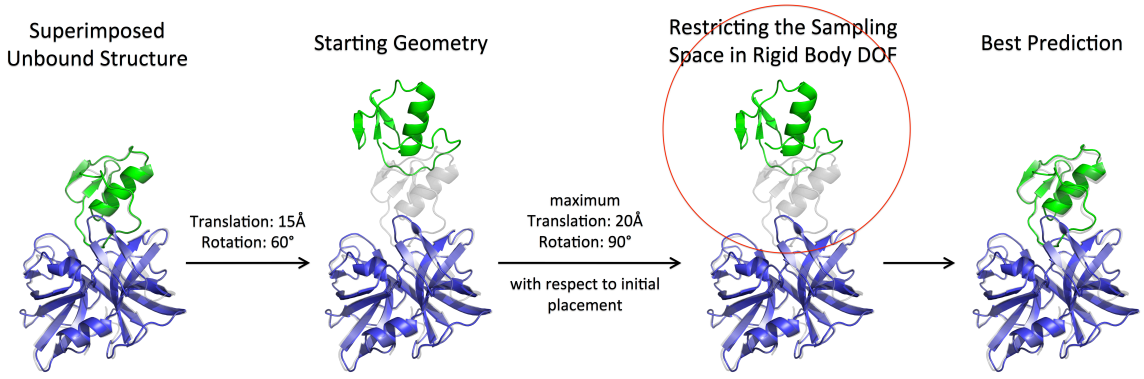

Figure 3.1. Docking refinement conditions. Each docking starting geometry was generated by an initial random translation of one unbound partner from the geometry in the complex by 15Å and random rotation of 60° (compare green displaced and grey cartoon representations). During the docking search translation and rotation of one partner with respect to the other was restricted relative to the starting geometry by 20Å and 90° (indicated by red circle), respectively.

3.2.2 Restricting the sampling space in rigid body degrees of freedom

For rigid body moves, random translations drawn from Gaussian distribution are performed along all three axes, and the axis-angle notation is used to represent rotations [41]. In order to perform a local search in the vicinity of the starting geometry, the sampling space in the rigid body degrees of freedom was restricted with respect to the initial input conformation by a maximum translation of 20Å and maximum rotation of 90° (this exceeds the maximum displacement of the starting structure from the bound configuration, see previous paragraph and illustration in Figure 3.1). To avoid dissociation of the two binding partners, we also applied an encounter constraint, which acts on the distance between the center of mass of the two binding partners and only penalizes the sampled geometries if the two binding partners are too far apart [41].

3.2.3 General Settings

We have combined enhanced sampling techniques with Monte Carlo (MC) method to sample protein-protein docking with atomistic representation, and tested four

protocols within Rosetta including standard MC, Temperature Replica Exchange Monte Carlo (REMC), well-tempered ensemble temperature Replica Exchange Monte Carlo (WTE-REMC), and well-tempered ensemble 2-dimensional Hamiltonian Replica Exchange Monte Carlo (WTE-H-REMC). In those docking approaches, rigid body displacements and side-chain optimization are accomplished by the rigid body mover UnbiasedRigidBodyPerturbNoCenterMover, and sidechain movers including JumpRotamerSidechainMover, PerturbRotamerSidechainMover and PerturbChiSidechainMover. Those movers are applied under the control of the Metropolis-Hastings framework. For each move, the MetropolisHastingsMover randomly applies one out of the four movers based on their sampling weights. Mover step size for UnbiasedRigidBodyPerturbNoCenterMover and PerturbChiSidechainMover are drawn from random Gaussian distributions. In the protocols with replica exchange, the magnitude for mover step size and sampling weight were modulated according to the replica level during initialization such that in the lower levels more frequent side-chain moves and fewer small rigid body moves were applied and in the higher levels less frequent side-chain moves and larger rigid-body moves. The magnitude of the step size and sampling weight were, however, kept fixed along the simulation in each replica. All the settings for the reference replica were made exactly the same as used in the standard MC protocol, and we denote these settings as reference settings. If not indicated otherwise, for all the protocols and on each target $2x10^6$ MC steps were employed. Snapshots are taken and stored every 1,000 steps. In the REMC protocols exchanges were attempted every 1,000 MC steps.

3.2.4 Monte Carlo and Asynchronous Parallel Tempering protocol

For the standard MC docking protocol, 25 trajectories are run with temperature set to 0.15. At the end, about 25 x 2,000 sampled structures were collected for each target. The step size of translation and rotation for rigid-body moves are drawn from normal distributions with small mean value of 0.1Å and 1°. The sampling weights for UnbiasedRigidBodyPerturbNocenter, JumpRotamerSidechain, PerturbRotamerSidechain and PerturbChiSidechain are set to 0.5, 4, 6 and 10, respectively.

For the parallel tempering replica exchange (REMC) protocol, 13 temperature levels were drawn from geometric progressions ranging from 0.15 (reference) to 0.31. Two trajectories with the 13 replicas are run for each target. Exchanges were attempted between neighbor replicas every 1,000 steps. For all targets, good exchange rates (between 25% and 69% with median value 49%) are achieved and

no further target dependent optimization was required. In replica exchange, it is common that the speed of each replica is not exactly the same. To avoid that the faster replica wait for the slower partner, we used an asynchronous exchange scheme. That is the faster replicas can perform more steps instead of waiting for its partner to reach the pre-defined exchange stride. Finally, the simulation will finish as soon as the slowest replica has reached the required step number.

3.2.5 Well-Tempered Parallel Tempering and Hamiltonian replica exchange protocol

We applied the well-tempered ensemble (WTE) technique with parallel tempering replica exchange Monte Carlo using a value of 5 for the tunable factor γ and reduced the temperature levels from 13 to 5 with the same range. The bin size for well-tempered ensemble technique to collect the history-dependent bias energy is set to two units of Rosetta docking energy. The resulting exchange rates are between 20% and 55% with median value 37%. For the WTE-H-REMC protocol we took advantage of the splitting of the van der Waals interactions into attractive and repulsive components in *RosettaDock*. It is represented with a modified Lennard-Jones 6-12 potential which includes a linear extrapolation in the repulsive part below the threshold of $0.6\sigma_{ij}$, where σ_{ij} is the sum of the atomic radii of atoms i and j. The atomic radii and energy well depth are taken from the CHARMM19 parameter set [36,121,138], and we denote this as "hard-rep". For the standard "soft-rep" in Rosetta, the atomic radii were either held fixed or scaled by a factor of 1.07 (typically for non-polar atoms) from the hard-rep radii, and the "switch point" for the linear extrapolation was selected empirically [129]. In the WTE-H-REMC protocol, we applied a 2-dimensional replica exchange, with the temperature as variable in the first dimension, and used the scaling factor for the soft Lennard-Jones repulsive term as the second dimension. The scaling factor allows linear interpolation of atomic radii and switch point between the hard-rep and soft-rep potentials (see above). In the Hamiltonian scaling dimension, we used five levels: hard_rep, soft 50%, soft 55%, soft 60% and soft 65%. In the temperature dimension five temperatures between 0.15 and 0.3 were used (in arbitrary units depending on the scaling of the Rosetta score), yielding a total of 25 replicas. Exchange between neighboring replicas is attempted every 1,000 steps along the two dimensions. Well tempered ensemble technique was applied again to improve the exchange rate with tunable factor γ=5. The bin size for well-tempered ensemble technique to collect the history-dependent bias energy is set to two units of the Rosetta docking energy. Note, that each replica accumulated separate history-dependent biasing potentials depending on the individual sampling history. Replica exchange rate ranged between 14% and 56% in the temperature dimension with median value 32%. In

the scaling dimension, it increased along the shifting from soft to hard repulsive interaction, and ranged between 12% and 99% with median value 39%.

3.2.6 Implementation in Rosetta

Previously, we implemented replica exchange within the general Metropolis-Hastings framework of the Rosetta3 software package [41]. The replica exchange module is accessible through the RosettaScripts interface and can be combined with any conformational moves that are implemented as children of the ThermodynamicMover class. For rigid-body docking refinement, we have applied a rigid-body mover UnbiasedRigidBodyPerturbNoCenterMover, and sidechain movers including JumpRotamerSidechainMover, PerturbRotamerSidechainMover and PerturbChiSidechainMover. For detailed balance, UnbiasedRigidBodyPerturbNoCenterMover performs unbiased rotational and translational perturbations in the restricted space as described in the previous section, and sidechain movers provide the proposal density of the perturbation through implementation of the abstract virtual function compute_proposal_density() in the ThermodynamicMover interface. The acceptance of a move is decided by the Metropolis criterion [99].

Side-chain motion is applied on one randomly selected residue (among all the residues and all the residue types but Proline) each time. Continuous sampling of side chain chi angles are used instead of fixed rotamers in all the three sidechain movers. The angles are chosen according to the Dunbrack rotamer library 2010 probabilities [139]. PerturbChiSidechainMover does a perturbation on the side chain *chi* angles, either uniformly distributed or Gaussian distributed with a given magnitude. For JumpRotamerSidechainMover and PerturbRotamerSidechainMover, a rotamer is first selected randomly or selected such that it has the highest probability of proposing the old *chi* angles according to the Dunbrack rotamer library probabilities, respectively, then individual *chi* angles are chosen using Gaussian distributed random angles with the means and standard deviations from the Dunbrack rotamer library.

The well-tempered ensemble technique is implemented into the framework of MetropolisHastings as ThermodynamicObserver. It is applied with a certain time interval (here in the test stride=1) and deposits the Gaussians to the bias energy with height of

$$W = \omega e^{-[V(s,t)/\Delta T]}\tau_G$$

where τ_G is the time interval or stride number, $V(s,t)$ is the old bias energy in the energy bin where the current decoy's energy has dropped into, ω represents the initial bias deposition rate and $\Delta T = (1-\gamma)T$, in which γ represents the tunable factor and T is the temperature in the simulation [81,135,137,140]. When well-tempered ensemble technique is applied, acceptance of a move or an exchange attempt is decided based on the total energy, which is the sum of the force field energy and the bias energy. For final analysis, only the force field scoring energy was used.

Table 3.1. Test complex structures and partner structures

Complex	Cat.	Difficulty	Partner I	Nres 1	Partner II	Nres 2	RMSD (Å)	DASA (Å)
1EAW_A:B	E	rigid	1EAX_A	241	9PTI_	56	0.54	1866
1GCQ_B:C	O	rigid	1GRI_B	66	1GCP_B	56	0.92	1208
1KTZ_A:B	O	rigid	1TGK_	105	1M9Z_A	82	0.39	989
1PPE_E:I	E	rigid	1BTP_	223	1LU0_A	29	0.44	1688
1S1Q_A:B	O	rigid	2F0R_A	141	1YJ1_A	69	0.98	1288
2AYO_A:B	O	rigid	2AYN_A	337	2FCN_A	72	1.39	3027
2SNI_E:I	E	rigid	1UBN_A	274	2CI2_I	64	0.35	1628
3D5S_A:C	O	rigid	1C3D_A	294	2GOM_A	61	0.56	1620
3SGQ_E:I	E	rigid	2QA9_E	185	2OVO_A	51	0.39	1211
7CEI_A:B	E	rigid	1UNK_D	127	1M08_B	87	0.7	1384
1AY7_A:B	E	rigid	1RGH_B	96	1A19_B	89	0.54	1237
1H9D_A:B	O	rigid	1EAN_A	125	1ILF_A(1)	114	1.32	2121
1HE1_C:A	O	rigid	1MH1_	176	1HE9_A	128	0.93	2113
1JK9_A:B	O	difficult	1QUP_A	219	2JCW_A	153	2.51	2130
1MQ8_A:B	O	medium	1IAM_A	184	1MQ9_A	171	1.76	1253
1RV6_VW:X	O	rigid	1FZV_AB	189	1QSZ_A	92	1.09	1626
1YVB_A:I	E	rigid	2CHU_A	241	1CEW_I	108	0.51	1743
2CFH_A:C	O	medium	1SZ7_A	156	2BJN_A	137	1.55	2384
2OUL_A:B	E	rigid	3BPF_A	236	2NNR_A	107	0.53	1933
2SIC_E:I	E	rigid	1SUP_	275	3SSI_	107	0.36	1617

cat: Complex category labels: E = Enzyme/Inhibitor or Enzyme/Substrate O = Others
RMSD: RMSD of Ca atoms of interface residues calculated after finding the best superposition of bound and unbound interfaces.
DASA: Change in Accessible Surface Area upon complex formation .

3.2.7 Construction of a Benchmark

The four protocols were first tested on 10 unbound targets (Table 3.1) from the benchmak4.0 set [141,142] with reasonable energy funnels using the RosettaDock scoring force field. This was checked by generating 1000 decoys with standard RosettaDock full protocol starting from the bound docking geometry (using

unbound structures). These 10 targets belong to the group of "rigid body" docking cases with small changes of side chains associated with complex formation (according to the classification of the protein docking benchmark4.0 collection [141,142]). In addition, the standard MC protocol and the WTE_H_REMC protocols were also tested on another 10 unbound targets including one "difficult" (1JK9) and two "medium difficulty" (1MQ8 and 2CFH) targets (Table 3.1). The number of residues of the 20 targets range between 122 and 409.

3.2.8 Analysis of docking results and computational efficiency

The sampled docked complex were analyzed according to ligand RMSD (L_rmsd) and fraction of native contacts (f_{nat}), as defined in CAPRI [102] using the bound complexes as references. To evaluate the capacity of the methods to sample near-native decoys, we calculated the fraction of CAPRI medium (**, $f_{nat} \geq 0.5$ & L_rmsd $> 1\mathring{A}$ or $0.3 \leq f_{nat} < 0.5$ & L_rmsd ≤ 5) or high (***, $f_{nat} \geq 0.5$ & L_rmsd $\leq 1\mathring{A}$) quality decoys in the collected results [65]. To evaluate the agreement between generated complexes with closest L_rmsd and best f_{nat} compared to the bound complex. To evaluate the scoring energies the interaction score (I_sc) was used which is computed by subtracting the all-atom energy of non-interacting partners from the all-atom energy of the interacting partners in the complex. To compute the energy of non-interacting partners the two binding partners are moved far away from each other while keeping all internal degrees of freedom fixed. To investigate the efficiency of optimizing the scoring energies, we calculated for a given MC step number the average difference of the sampled best score (up to the selected MC step number) and the final most favorable score.

3.3 Results and Discussion

3.3.1 Results and Discussion

Monte Carlo docking simulations are frequently used to perform protein-protein docking searches or for the refinement of predicted complexes at atomic resolution including limited conformational changes of the partner structures [38,39,41,76]. In recent years, enhanced sampling methods to improve the MC search efficiency have been developed. In order to test the performance of such improvements we compare the application of standard MD, parallel tempering REMC, well tempered replica exchange (WTE-REMC), and well tempered ensemble combined with 2-dimensional temperature and Hamiltonian replica exchange (WTE-H-REMC) to a set of protein-protein complexes in unbound partner conformation. In each case the MC moves included rigid body translation and rotation as well as side-chain moves

(illustrated in Figure 3.2). For each protocol the same start configuration was used corresponding to a random arrangement of one mobile partner placed approximately 15 Å away from the bound complex geometry (see Methods for details). In case of the replica exchange methods only configurations in the reference replica were retained, resulting in approximately 25x2,000 decoys for standard MC protocol (with $2x10^6$ MC steps), 2x2,000 for REMC protocol, 5x2,000 for WTE-REMC protocol and 1x2,000 for WTE-H-REMC protocol, respectively. On a 2.6 GHz AMD Opteron Processor (12 cores), $2x10^6$ MC steps take between 4.5-20 hours. The sampled docking solutions were analyzed in terms of deviation from the known complex geometry (using the root mean square backbone deviation of the mobile ligand partner protein from the bound complex after best superposition of the receptor protein onto the bound complex: L_rmsd) and interaction score (I_sc).

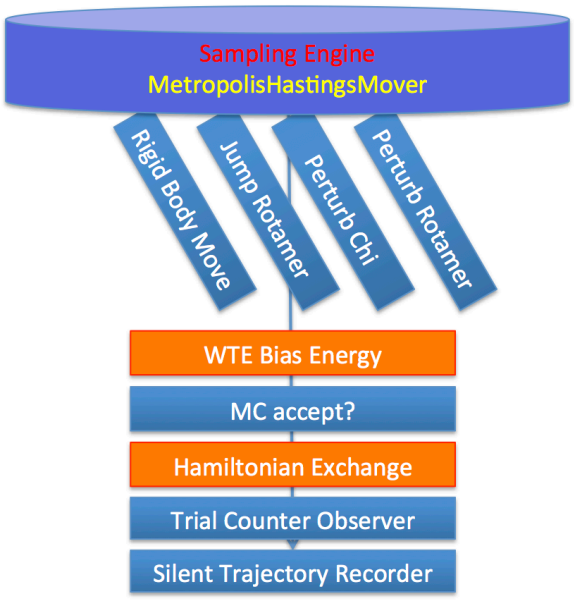

Figure 3.2. Workflow represented by combination of Rosetta modules and setup of the four docking protocols. The modules in orange, representing the enhanced sampling techniques of replica exchange and well-tempered ensemble, are only applied in the combined protocols.

The protocols were first tested on 10 benchmark targets of the "rigid body" category (with only limited side chain changes upon complex formation, see Methods for details). We consider the sampling of medium or high-quality solutions (CAPRI **/*** solutions, defined in Methods section) according to the CAPRI criteria as successful docking refinement. The evolution of the sampling in terms of L_rmsd and I_sc scoring is showcased for two representative targets (pdb1EAW and pdb3SGQ) in Figure 3.3. For the pdb1EAW-target all methods sample progressively lower (more favorable) force field scores with increasing number of MC steps. However, for the first example only the WTE-H-REMC protocol samples docking

solutions with L_rmsd < 5 Å after $2x10^6$ MC steps. Only after 10^7 MC steps all techniques except the standard MC technique sample near-native solutions (Figure 3.3A). For the second example (pdb3SGQ) the MC technique successfully samples solutions with Lrmsd < 5 Å only after 10^7 MC steps whereas all three advanced sampling techniques reach near-native solutions already after $3x10^5$ MC steps (Figure 3.3B). Qualitatively similar trends were observed for all other test cases (see *Appendix* Figures S12-S14).

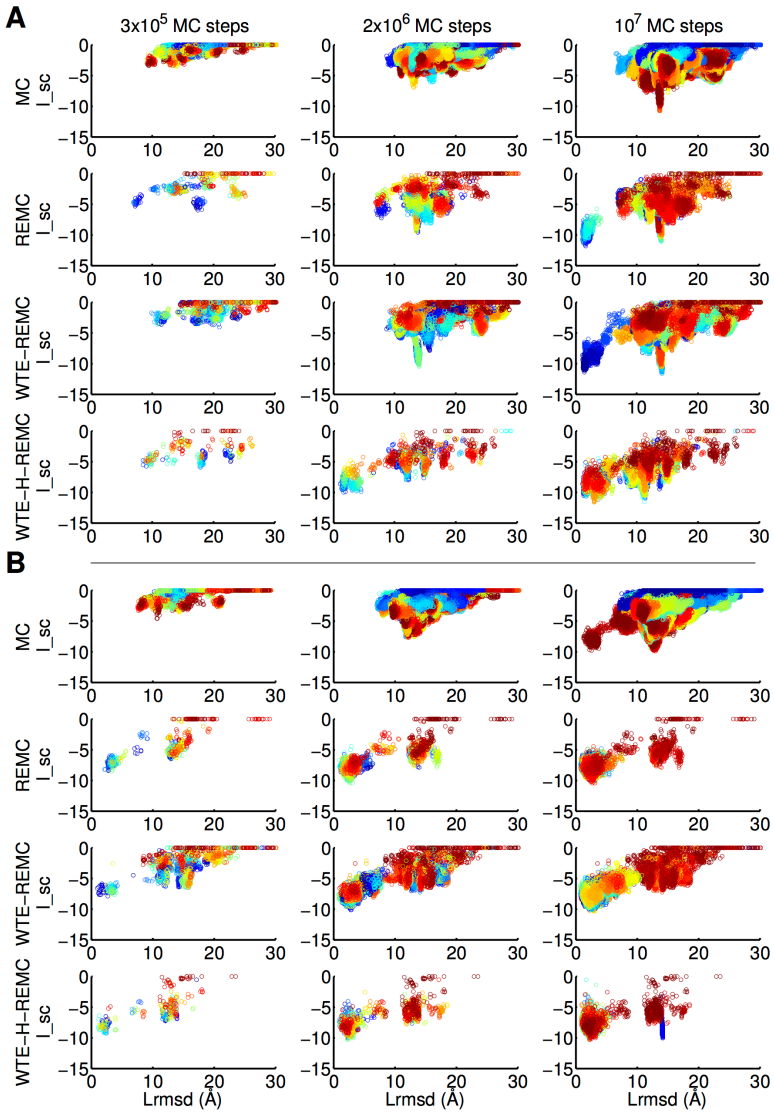

Figure 3.3. Scatter plot of interaction score I_sc (Rosetta units) vs. L_rmsd (Å) for the four docking refinement protocols and two representative targets, 1EAW (A) and 3SGQ (B). The protocol is indicated on the left for each row of plots. The snapshots number is color-coded, that means blue and red dots corresponding to decoys sampled at the beginning and the end of the docking searches in each panel, respectively. The three columns of plots indicate the result after different simulation lengths (indicated on top of each column).

A quantitative comparison of the docking refinement solutions in terms of the fraction of native contacts (f_{nat}) indicates that the WTE-H-REMC method succeeded in all 10 cases in sampling near-native docking solutions with very high quality (Figure 3.4A). In contrast, the other protocols succeeded only in 3 (MC) or 7 (REMC and WTE-REMC) of the first 10 cases (Figure 3.4A). Note, that near-native docking solutions are also the best scoring solutions in several but not all docking test cases (e.g. for targets 1KTZ, 3D5S and 7CEI docking solutions with Lrmsd > 10 Å give I_sc scores lower than the solutions closest to the bound docking geometry, *Appendix* Figures S13 and S14). The results on the first 10 test cases indicate that the WTE-H-REMC enhanced sampling protocol showed the best performance. For a second test set of 10 targets (including also targets of the "medium" and "difficult" category, see Table 3.1), only the standard MC and the WTE-H-REMC protocols were compared. Again, the WTE-H-REMC protocol gave better docking results in 5 cases (IJK9, 1MQ8, 2CFH, 2OUL, 2SIC) with lower final I_sc scores and Lrmsd (Figure 3.4A, *Appendix* Figures S15-S17) compared to the standard MC-method. However, in two cases (1H9D and 1HE1) the standard MC–method reached configurations closer to the bound form compared to the WTE-H-REMC technique. Note, that especially in these cases the score of near-native docking solutions was higher (less favorable) than for alternative docking geometries (*Appendix* Figures S16 and S17). Since the search techniques optimize the score (and not deviation from bound structure) it may explain the failure of the WTE-H-REMC technique in these cases.

To check if slower convergence to reach low energy docking solutions was the main reason for the failure of some protocols to reach near-native docking solutions, we increased the trajectory length to 10^7 MC steps. Indeed, the success of MC, REMC, and WTE-REMC protocols to reach near-native docking solutions increased to 8, 10 and 10 out of the first 10 targets, respectively (Figure 3.4B, see also *Appendix* Figures S14 and S17). The results were further analyzed with respect to fraction of native contacts (f_{nat}) of near –native docking solutions and the maximum quality of predicted docking geometries (Figure 3.4). Also for these measures and in case of the protocol with $2x10^6$ MC steps the WTE-H-REMC protocol achieves overall the best performance (Figure 3.4A). For the extended protocol with 10^7 MC steps the quality of solutions in terms of f_{nat} is more similar for all 4 protocols (Figure 3.4B), indicating that indeed the standard MC technique requires longer searches to achieve convergence compared to the WTE-H-REMC method. The best f_{nat} for protocols MC, REMC and WTE-REMC all increased on average around 17%, getting close to that of the WTE-H-REMC protocol. The best f_{nat} for protocol WTE-H-REMC also increased slightly (~7%, Figure 3.4B). Figure 3.3 presents two representative

examples on targets 1EAW and 3SGQ of ligand RMSD (L_rmsd) versus interaction score (I_sc). Comparing these data with that from the trajectory of $2x10^6$ MC steps, it shows that the details of the energy landscape sampled by WTE-H-REMC remains similar indicating reasonable convergence within $2x10^6$ MC steps for most cases. Meanwhile, the rigid-body space reached for target 1EAW by protocols MC, REMC and WTE-REMC, and on target 3SGQ by protocol MC, drastically improved (Figure 3.3).

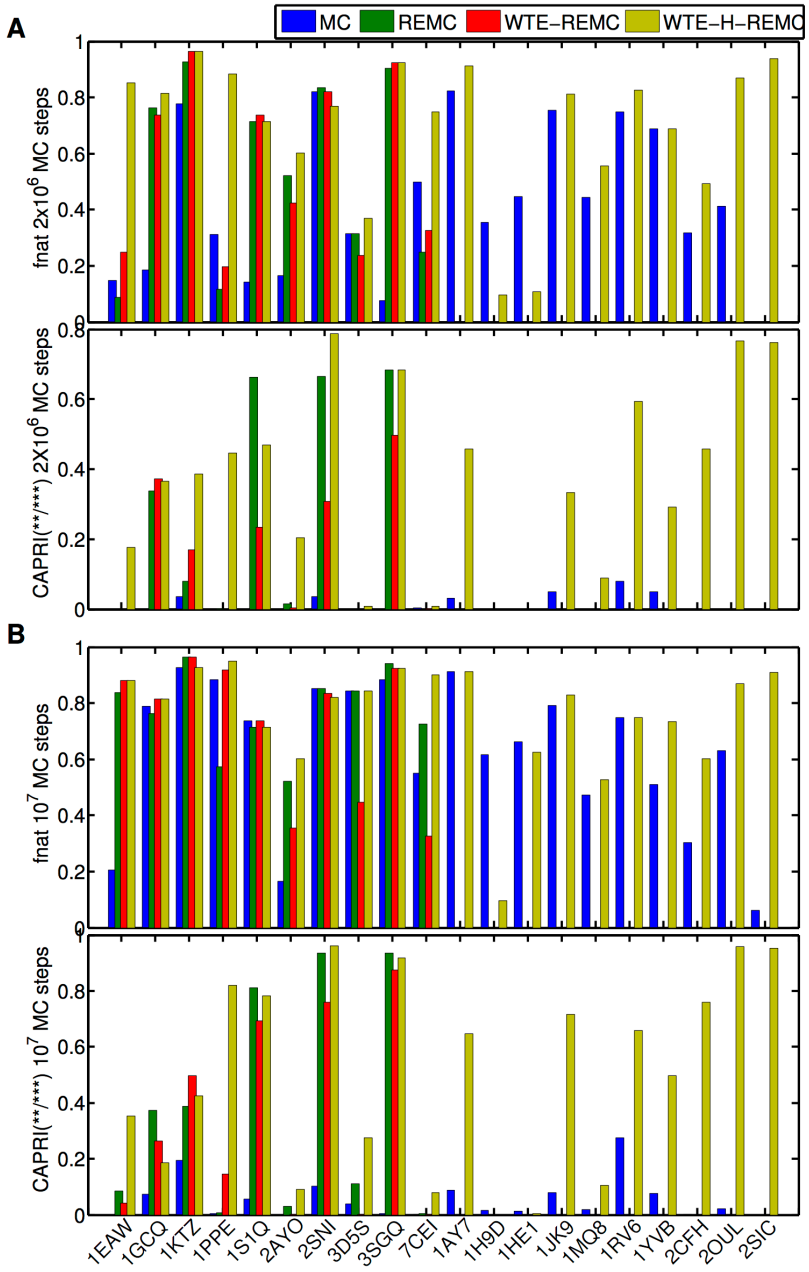

Figure 3.4. Agreement between sampled docking geometries and the corresponding bound complex. (A, upper panel) Highest fraction of native contacts (f_{nat}) found in the top 10 decoys (according to

L_rmsd) sampled in each protocol ($2x10^6$ MC steps). (A, lower panel) Fraction of CAPRI medium and high quality complexes found for each target and each protocol (the protocols MC, REMC, WTE-REMC and WTE-H-REMC are indicated by different colors). (B) same as in (A) but for the docking refinement runs with 10^7 MC steps.

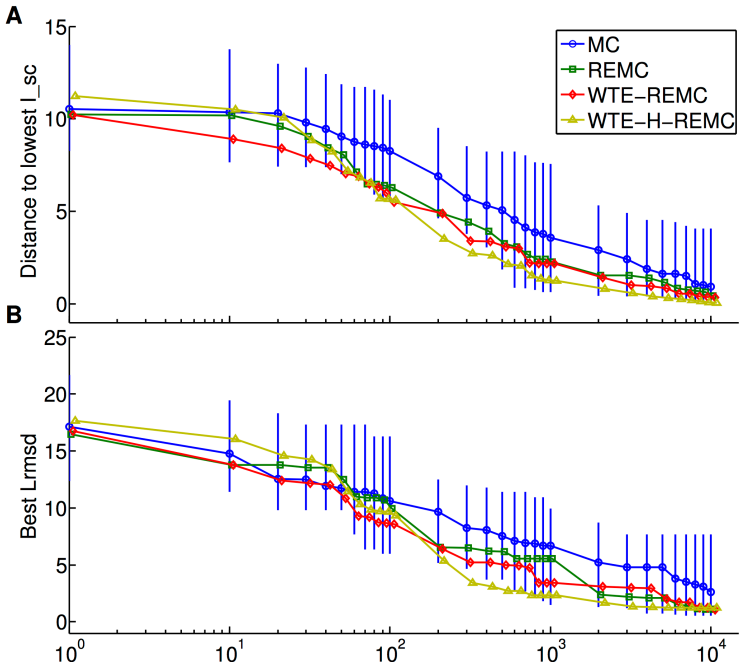

Figure 3.5. Evolution of I_sc docking interaction score (A) and best sampled Lrmsd (B) vs. MC step number. The MC step number is scaled by x1000. For the interaction score I_sc the smallest difference (sampled up to the selected step number) relative to the lowest scoring complex sampled in the entire docking search is plotted. The variance in sampled scores (up to the considered number of MC steps) is indicated by error bars for the MC protocol. It is of similar magnitude for the other protocols (not shown). For (B) the smallest sampled L_rmsd up to the step number indicated in the x-axis is shown.

Since the force field score is the quantity which is optimized during the docking searches (and not the agreement with the bound structure) it is of interest to compare the protocols in terms of the efficiency to optimize the force field score. For each target, the average of the lowest 10 interaction scores (lowest I_sc) sampled up to a given MC step number was considered (using the extended trajectories) and the difference relative to the lowest score found during the search was recorded. The average of this quantity for all 20 cases was calculated and plotted in Figure 3.5. The enhanced sampling techniques REMC, WTE-REMC and WTE-H-REMC, consistently, reached lower interaction scores than the standard MC method for a given number of MC step (Figure 3.5A). The WTE-H-REMC technique reached on average lower I_sc than the other three protocols already after ~$3x10^5$ MC steps. Interestingly, the same analysis using the L_rmsd instead of the I_sc yields the same trend, indicating that on average the I_sc score correlates with the L_rmsd (Figure 3.5B). Low L_rmsd

of sampled geometries gives on average (but not for all targets) also a favorable score.

3.3.2 Protocol Testing

A subset of three complexes (pdb-entries: 1PPE, 2OUL and 2SIC) was chosen as an independent protocol test set. For two of these three complexes, only the MC and the WTE_H_REMC protocol had been tested in the main work. The four protocols were run using the RosettaScripts interface with parameters as described above, trajectory length of $2x10^6$ MC steps and a newly created starting conformation. In contrast to the previous docking runs, only a single starting conformation was generated for each complex. All the protocols could be successfully executed. A summary of the test results can be found in Table 3.2. For all three test cases, the enhanced sampling methods yielded structures of lower L_rmsd and higher f_{nat} than standard MC sampling. Enhanced sampling methods generated near-native docking models, whereas standard MC sampling did not yield any structures of CAPRI one star quality or better. The WTE_H_REMC technique was the only method to generate CAPRI three star quality structures for all three cases and thus yielded the best performance on the test set. The results match the previously presented data for these three complexes and thus confirm that the sampling does not depend on the choice of the starting conformation. The randomly generated starting conformations sometimes contained clashes, but the enhanced sampling methods were able to refine them to high-quality solutions. Hence, it might be possible to use enhanced sampling methods also for refinement of docked complexes using other methods than *RosettaDock*.

Table 3.2 Results for protocol testing on a subset of three complexes.

PDB ID	MC	REMC	WTE-REMC	WTE-H-REMC
1PPE	5.6 Å	1.5 Å	2.2 Å	1.0 Å
	14.5 h	14 h	12.3 h	17.8 h
2OUL	4.2 Å	1.9 Å	2.2 Å	0.7 Å
	21.8 h	18.3 h	18 h	22 h
2SIC	4.9 Å	0.4 Å	2.8 Å	0.5 Å
	25.5 h	22 h	19 h	20 h

For all complexes, the best sampled L_rmsd and the execution time on 27 threads are listed.

3.4 Conclusions

In this work, four different Monte Carlo advanced sampling protocols implemented in *RosettaDock* to predict the geometry of protein-protein complexes have been compared. For all the protocols (on each target) the same initial protein-protein docking start configurations were used with 15 Å translational displacements and 60° rotation of one partner from the native complex structure. This situation corresponds to the frequent scenario that the interaction region between proteins is approximately known and start configurations are placed close to the approximately known binding region. It is also very useful for directly comparing the docking performance of different approaches at the computational demanding atomistic high resolution level. Note, that the protocol is computationally too demanding for routine applications that require to search over the entire surface of two protein partners. If complete protein surfaces are considered it is also very likely that the scoring function is not accurate enough to pick out near-native solutions as lowest energy complexes. Our results on docking refinement show indeed that the application of advanced sampling schemes improves the docking refinement performance yielding final configurations in better agreement with the bound structure and yield also a much larger fraction of near-native structures compared to regular MC searches. The WTE-H-REMC consistently gave the best performance since it explores the phase space more efficiently due to larger energy fluctuation and due to the added biasing potential that effectively smoothes the landscape and increases the replica exchange rates. An increase of the number of MC steps to 10^7 in the standard MC protocol resulted in improved performance achieving a similar fraction of native contacts of best sampled solutions and similar final docking scores compared to WTE-H-REMC with $2x10^6$ steps. Hence, in most cases a standard MC protocol requires roughly 5 times larger computational demand to achieve the same final docking prediction performance. It should be emphasized that this reflects only a general trend. For some test cases even 10^7 MC steps still gave inferior docking results compared to WTE-H-REMC and still the fraction of the best solutions relative to the total number of sampled geometries is much smaller than for the advanced sampling method. Further improvement might be possible by an adjustment of the bin size in the WTE to collect the history dependent bias energy. However, an even larger gain could be achieved by an improvement of the docking scoring function to increase the gap between ranking near-native and non-native solutions.

3.5 My contributions to this project

In this paper "Application of Enhanced Sampling Monte Carlo Methods for High-Resolution Protein-Protein Docking in Rosetta", I have conceived and designed the experiments together with Prof. Dr. Martin Zacharias, implemented the protocol into Rosetta together with Dr. Oliver Lange, carried out all the calculations, performed all the analysis and written the paper.

Chapter 4 Monte Carlo Replica Exchange based Ensemble Docking of Protein Conformations[§]

4.1 Introduction

Protein-protein interactions are fundamental for all living systems and involved in almost all biological processes, ranging from enzyme activation and inhibition to signal transduction and gene regulation. Knowledge of the structure of complexes is essential for the understanding of the function and mechanism of protein-protein interactions. However, the number of protein-protein complexes is much larger than the number of individually folded proteins [16,54,143]. Especially for transient interactions the experimental structure determination, for example using X-ray crystallography or nuclear magnetic resonance (NMR), can be very costly and technically difficult or even impossible. Hence, computational structure prediction by protein-protein docking methods plays an increasing role to generate at least models of protein-protein complexes, guiding experimental approaches and may help in structure determination if combined with low-resolution or sparse experimental data [16,52,64,144-146].

Protein-protein docking methods to predict the three-dimensional structure of complexes from its known constituents have gained steady progress over the past years. The community-wide blind docking challenge CAPRI (Critical Assessment of Predicted Interactions) has been monitoring the performance of docking methods since its establishment in 2001 [65,66]. Despite progress, accurate complex prediction remains challenging especially when coming to targets in which conformational changes take place upon binding (even for backbone RMSD changes as small as 1-2Å) [55-57,65,66,144,145,147]. In general, docking methods are much more successful when starting from the bound than from the unbound partner structures, even in cases with minor conformational changes [16,64,144,145]. This is suggesting that proper treatment of structural flexibility is an important factor for protein docking generally for most complexes, not just those exhibiting large changes upon complex formation [55,57,147-149].

Properly dealing with flexibility and conformational changes thus becomes one of the major challenges in the docking field. The difficulties are rooted in the large number of degrees of freedom for protein flexibility, which makes exhaustive sampling unfeasible, but also challenge the existing scoring functions and result in a higher rate of false-positive solutions. Currently, many docking programs employ a two-stage protocol, first

[§] The work presented in this chapter is in press on journal Proteins: Structure, Function, and Bioinformatics

broad sampling while keeping partner structures rigid, then followed by refinement in one or multiple steps, which normally allow conformational changes of side chains and sometimes also of selected backbone regions [144,145]. In this kind of two-stage docking approaches, conformational flexibility can also be accounted for in the rigid-body docking stage implicitly by soft docking [150,151] or a coarse-grained force field [37], which allows a certain amount of steric clashes to deal with side chain flexibility and sometimes also small scale backbone movements. Then steric clashes can be resolved in the proceeding refinement stage. Alternatively, relaxation in soft collective normal modes of the proteins during docking can be helpful if global motions of the protein partners are involved in the binding process [152,153]. In the refinement, flexibility is typically explicitly expressed by rotamer libraries or copies of side chains [36,154] and dihedral torsional angles for backbone possibly combined with side chain mobility [36,48,75,154-156]. Besides, soft docking can also be included in the refinement stage through Hamiltonian replica exchange [51]. The strategy behind those docking methods mimics the induced-fit model for protein-protein interaction, reflects the popularity of the model in the past decades, and also mirrors our current understanding of the recognition mechanism.

An alternative mechanism, termed conformational selection generalized and adapted to protein interactions [52,54,146], postulates that all binding relevant protein conformations pre-exist (to a significant extend), and the partner selects the best fitting conformation to bind. Following binding the ensemble undergoes a population shift, redistributing the conformational states. Recently, Csermely *et al* has extended the conformational selection to describe the general scenario, where both selection- and adjustment-type steps act simultaneously [54]. Following this model of molecular recognition, conformational flexibility can be represented by pre-calculated ensembles of structures. The term ensemble docking typically refers to a one by one docking of each ensemble member for partner 1 to each ensemble member of partner 2 [34,58,59,157] which results in a dramatic increase of the computing time with respect to the size of the conformational ensembles and becomes prohibitively expensive for large ensembles. Thus studies investigating the use of ensembles with cross docking of all ensemble members have only been performed on relatively small benchmarks or single cases, which have shown improvement over single conformation unbound docking in terms of the number of near native solutions [34,57,58]. The representation of the receptor binding region by multiple conformations has been found beneficial for the docking success also in the area of docking of drug molecule candidates to protein binding sites [59-63]. In particular, it has been found that docking in combination with stochastically switching between different receptor structures is both computationally efficient and improves the accuracy of the docking results [60,61]. Monte Carlo type switching

between receptor binding site conformations (ordered along the structural similarity) during docking avoids the rapid increase of the computational demand in case of docking ligands separately to each member of the ensemble [58]. The stochastic search rapidly settles towards the regime of receptor binding site conformations compatible with favorable binding of a ligand [60,61]. In the present study we follow this approach for docking of protein partners but employ ensembles for both partners and also combine it with replica exchange Monte Carlo (REMC) to further improve the search performance for favorable binding geometries. The implementation employs the protein-protein docking program ATTRACT [37]. It uses a coarse-grained protein model which is intermediate between a residue-level and full atomistic description, and represents each amino acid by several (up to four) pseudo atoms. The coarse-grained representation results in a relatively smooth energy landscape which contains fewer docking energy minima [37,158-160]. This feature can further facilitate the use of Monte Carlo methods. We have thus in this work applied Monte Carlo combined with temperature replica exchange (REMC) to sample the docking ensemble space as well as the rigid-body space, making ensemble docking computationally feasible. Several docking protocols have been compared which include minimization for unbound docking (MIN-unbound), replica exchange Monte Carlo for unbound docking (REMC-unbound), minimization for ensemble docking (MIN-ensemble), and replica exchange Monte Carlo for ensemble docking (REMC-ensemble). Our results indicate that even though the generated ensembles are not optimal, ensemble docking still outperforms unbound docking using single conformations for each partner and REMC-ensemble docking consistently performs better in terms of the number and quality of near-native solutions and also reaches significantly lower scoring energy in most cases compared to other protocols.

4.2 Methods

4.2.1 Generating conformational ensemble

Rather than by a single static structure, proteins are better described by an ensemble of conformations [54,55,58,59,147,161]. Ensemble representations have been found useful for improving applications of molecular modeling such as protein-small molecule and protein-protein docking as well as in protein design. Several computational methods have been developed to model conformational ensembles. In this study, we have used normal mode deformation [162], Rosetta backrub deformation [163,164] and MD simulation snapshots to generate conformational ensemble starting from the free unbound structure. All easy and medium difficulty targets except 1N2C (too big), 1QFW (HL:AB) and 1OYV (B:I) in docking benchmark 4.0 have been tested. For the ligand of 1H9D, 2VIS, 2OUL and 2FJU and the receptor of 1YVB and 2OUL, normal modes protocol does not

run, we have used Rosetta backrub generated models to substitute the normal modes models.

In order to generate globally deformed variants of unbound protein structures soft normal modes based on an elastic network model developed by Hinsen [162] was employed. Since most of the global deformability is encoded in the softest normal modes we performed deformations in the two possible directions along a normal mode corresponding to an excitation with an energy of 1RT (R: gas constant, T: room temperature). Typically, this leads to a root mean square deviation (RMSD) of the structures from the unbound reference structure of up to 0.5-2 Å. To limit the number of possible global deformations only the four softest normal modes were used (no combined deformations were considered). The normal modes were calculated for the protein backbone and all side chain atoms were displaced like the corresponding backbone atoms during normal mode deformation. Thus, this procedure gives mainly structures with collective backbone displacements but little side chain changes. To remove any steric strain, structures were finally energy minimized using the sander program of the Amber14 [165] in 2000 minimization steps using the parm14SB force field [166].

The Backrub model is inspired by alternative conformations observed in high-resolution crystal structures of proteins, and has been suggested to capture a significant fraction of small conformational changes proteins undergo in solution [163,164]. We have run backrub application implemented in the program Rosetta with RT value for temperature 0.35 and number of Monte Carlo trials of 10000. The generated models have RMSDs from the starting unbound structure between 0 and 3 Å mostly populated around 1 Å. This procedure yields structures that include both small (mostly localized) backbone changes and side chain changes.

As a third method small deformations of the unbound structure were also generated using restraint MD simulations with the unbound structure as reference conformation. Distances between Cα atoms in the unbound structure were used as restraints to limit the backbone. This included all Cα-Cα distances within the distance interval of 8-16 Å of the unbound structure and allowing distance changes of +/- 0.5 Å (further deviations were penalized with a force constant of 1.5 kcal·mol^{-1} Å^{-2} for the quadratic distance penalty potential). All side chain atoms were completely mobile. The unbound structure was first energy minimized (2000 steps) followed by short (10 ps, 300K) MD simulations using an implicit Generalized Born (GB) solvation model and the Cα-Cα distance restraints using the Amber14 pmemd.cuda program [165] using parm14SB force field [166]. For each conformational variant a different initial seed for the random velocities was assigned. Final structures were again energy-minimized (2000 steps). Note, that the short range distance restraints allow to some degree small global and local backbone changes but

keep the local secondary structure close to the structure in the unbound reference. This procedure produces models with RMSDs from the starting unbound structure ranging between 0.5 and 3 Å including only small backbone but more significant side chain changes (in contrast to the normal mode displacement procedure).

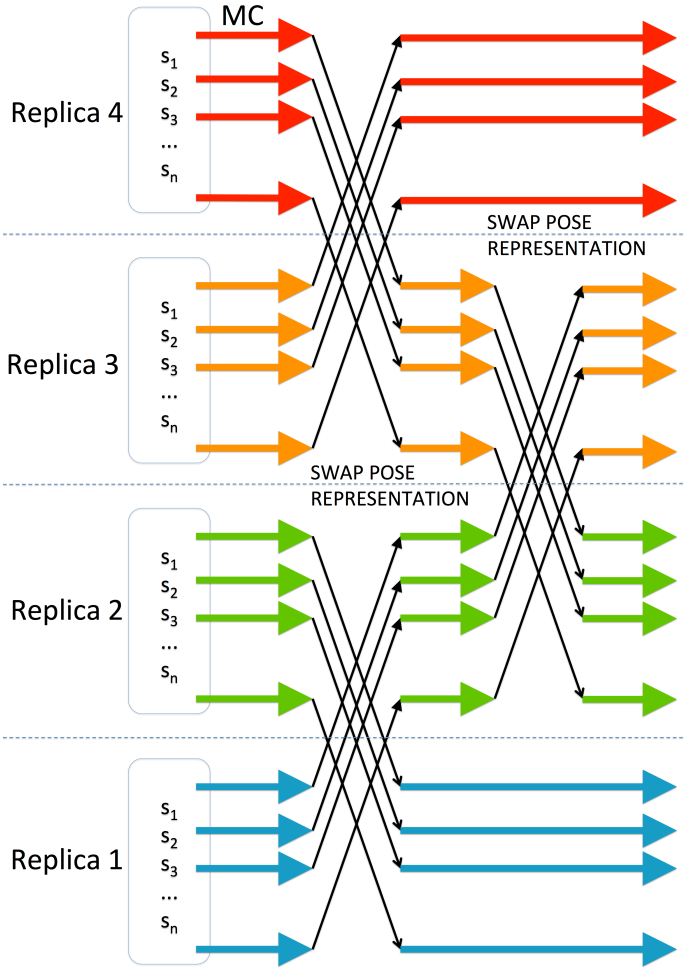

Figure 4.1. Simultaneous Monte Carlo replica exchange scheme. In the box, s_1, s_2, ... s_n refer to the n starting conformations. All of the four replicas start from the same set of conformations.

4.2.2 Randomized ensemble docking in ATTRACT with minimization

The ATTRACT CPU version supports randomized ensemble docking with minimization. For this, ATTRACT first generates initial placements for ligand on the surface of the receptor. Then for each initial placement, the model indices of both binding partners are assigned randomly within the respective range of the model numbers. During the minimization process, the respective model remains fixed and is also used for scoring and docking metric evaluation at the end.

4.2.3 Monte Carlo in ATTRACT explores docking ensemble space

Monte Carlo is frequently used to sample the rigid-body degrees of freedom or rotamers in protein-protein docking and docking refinement [5,36,39,41,51,154,167]. Inspired by the 'conformational selection' theory, we applied Monte Carlo also to sample the docking ensemble space. Rigid body move and/or ensemble move is randomly applied based on a pre-defined probability (25% rigid-body standalone move, 25% ensemble standalone move, 50% rigid-body + ensemble combined move), independently or simultaneously. In ATTRACT, a pose is represented by a translation vector and an Euler angle vector, plus an additional model number for ensemble docking, for each binding partner. For rigid body move, unbiased rotational and translational perturbations are applied. For ensemble move, a conformer index is randomly suggested with probability inversely proportional to its similarity (all-atom RMSD) to the current conformer. This RMSD is 0 for the current conformer. To avoid singularity, it is made to have a mean probability ($1/N_{conformer}$) to stay at the current one. Each standalone or combined move are judged by the Metropolis-Hastings criteria to decide if it will be accepted. To improve the sampling efficiency, we combined Monte Carlo with temperature replica exchange (REMC) [41].

Table 4.1. Summary of CAPRI classification.

CAPRI quality	Conditions based on CAPRI metrics
***** (High)**	$f_{nat} \geq 0.5$ & $Lrmsd \leq 1$ & $Irmsd \leq 1$
***** / ** (at least Medium)**	$f_{nat} \geq 0.5$ or $0.3 \leq f_{nat} < 0.5$ & $(Lrmsd \leq 5$ *or* $Irmsd \leq 2)$
***** / ** / * (at least acceptable)**	$f_{nat} \geq 0.3$ or $0.1 \leq f_{nat} < 0.3$ & $(Lrmsd \leq 10$ *or* $Irmsd \leq 4)$

The coarse grained representation in ATTRACT makes the docking energy landscape smoother and containing fewer docking energy minima. Benefiting from this feature, we only need a small number of replicas. In the test, four replicas are used and the values of RT for the temperature are initialized as 0.5 0.6 0.8 and 2.5 (RT units). For each replica, 200000 MC steps are run. Snapshots as results for final analysis are taken every 2000 MC steps. Replica exchange is attempted every 2000 MC steps (100 times), between the decoys started originally from the same initial placement and done for all. The RT value of 0.5 is fixed for the reference replica, the other temperatures are adjusted on the fly for sampling efficiency according to the collective exchange rate on all the decoys, to maintain an exchange rate between 30% and 60%. The temperature of the lowest three replicas are generally very stable, the high temperature of the highest replica in most cases converges after 80 or less

swaps (*Appendix* Figure S18). In replica exchange using an ATTRACT GPU CUDA version, we swap simply the pose representation instead of the temperature, and then in each CUDA device, the temperature remains always the same (Figure 4.1).

4.2.4 Measuring the quality of docking solutions

The docked complex structures are analyzed according to the fraction of native contacts (f_{nat}), interface root mean square deviation (Irmsd) and ligand root mean square deviation (Lrmsd). For the calculation of fraction of native contacts f_{nat}, a contact was defined as two residues from the two binding partner having at least one atom within a distance of 5 Å. For Irmsd, the interface residues in the target were defined as those having at least one atom within 10 Å from the other binding partner. Ligand RMSD (Lrmsd) was calculated on backbone atoms after the receptor of the prediction superimposed onto that of the bound conformation. The CAPRI classification of predictions as "high", "medium", and "acceptable" accuracy is summarized in Table 4.1.

To check how the generated structures score with respect to CAPRI criteria, we calculated the success rate within selected top N decoys by ATTRACT score as well as success rate in top clusters after clustering. The clustering was carried out on 4% selected decoys from all. For minimization, the decoys are selected by ATTRACT score. For REMC, we used every fifth snapshot from the first replica and kept the 80% low ATTRACT score decoys (1/4/5*80%=4%). We clustered the structures by pair-wise ligand-mass-center-RMSD for computational efficiency with a cutoff of 1.0Å and a minimum cluster size of 5. The clusters were then ranked by the average score of the cluster center and top four members. A cluster is designated as of one star CAPRI quality (*) if any of the cluster center or top four members is of one star CAPRI quality (same for medium (**) and high (***) quality clusters.

4.3 Results and Discussion

4.3.1 Ensemble generation

The methods to generate ensembles of partner structures starting from the unbound structures included normal mode deformation in the softest four normal modes based on an elastic network protein model, restraint MD simulations and Rosetta backrub simulations (see Methods). In test simulations we found that none of the methods performed on average particularly well in producing structures closer to the bound structure (*Appendix* Figure S19, and see below). For example in case of global collective motions normal mode displacements move the structure closer to the bound form but for more local changes MD simulations or backrub approaches are better suited. On average the Rosetta backrub approach gave slightly better results than the two other methods.

Hence, we decided to include models generated by all three approaches (as appropriate fractions) in each ensemble. To investigate a proper model number per ensemble, we first tested on 1T6B (830 residues) and 1J2J (203 residues) the use of several ensembles of different sizes (in steps of 10 models). Application of the REMC ensemble docking procedure revealed that beyond an ensemble size of 40 the Irmsd of the best models we sample does not further improve (not shown).

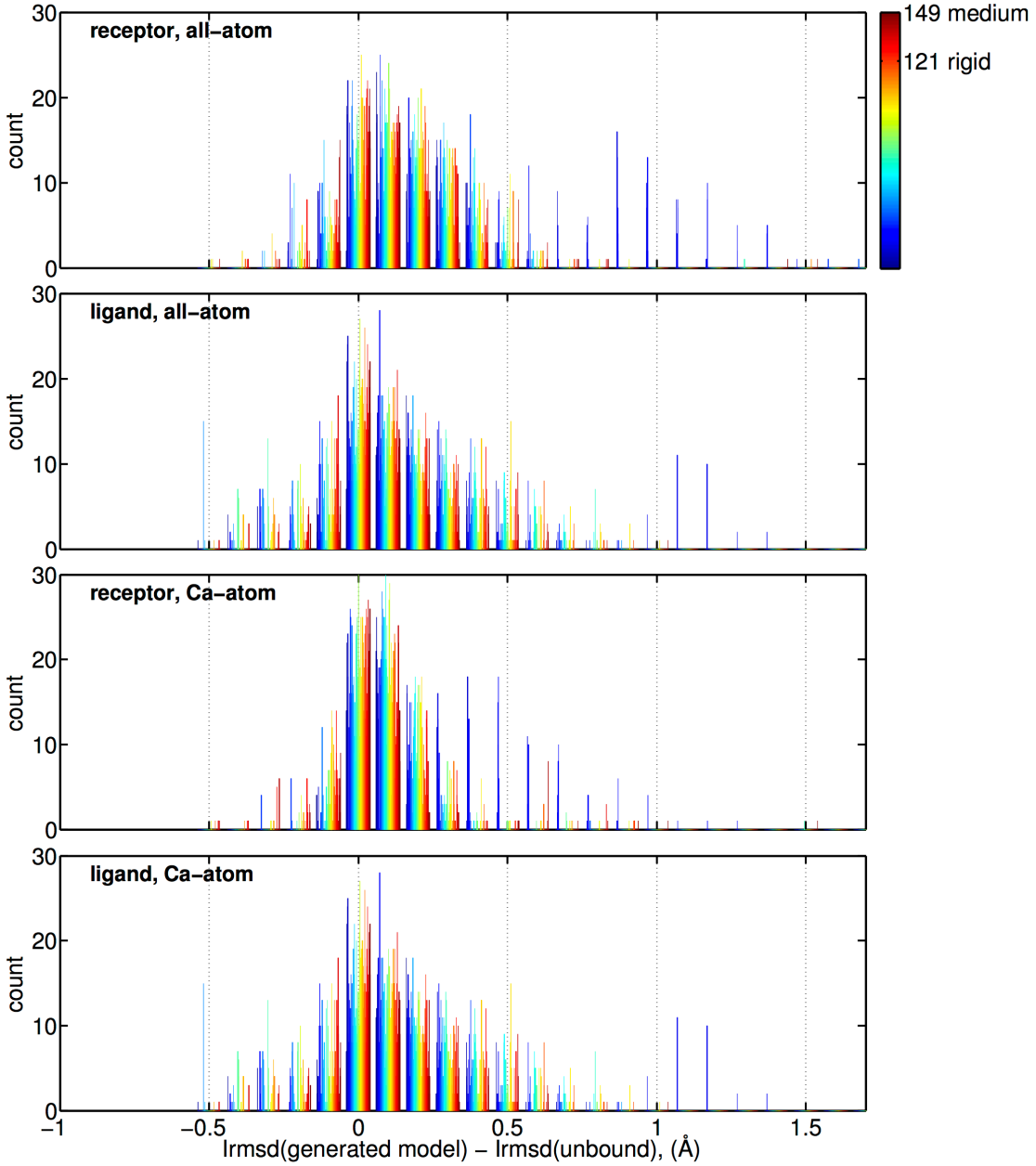

Figure 4.2. Distribution of differences in Irmsd of generated protein conformers after superposition on the native complex relative to the Irmsd of the unbound protein structure after superposition on the complex. The Irmsds are calculated for receptor (larger partner) and ligand proteins considering all-heavy atoms or Cα-atoms, respectively. The target index (1-121 are rigid body targets, 122-149

are medium difficult targets) is color coded. A negative difference (x-axis) indicates that a generated ensemble protein member fits better to the bound structure than the unbound structure at the protein-protein interface.

Thus for the systematic evaluation of the ensemble REMC docking method we generated for each binding partner 38 models starting from the unbound structure for the 149 targets in the benchmark, including 8 from normal mode deformation, 20 from Rosetta backrub calculations and 10 snapshots from MD simulations. Since most bound and unbound structures deviate by less than 2 Å [141] the maximum deviation of the generated structures from the starting unbound structure was set to ~2Å. Most of the generated models have Cα-atom Irmsd from the starting unbound structure below 1Å and all-atom Irmsd below 2Å (*Appendix* Figure S19). We then calculated the all-atom and Cα-atom Irmsd from the respective partner in the bound complex for the generated models and compared with the RMSDs for the unbound partners (*Appendix* Figure S20). For the 149x38 models generated for the receptor (larger partner), 14.6% (by all-atom Irmsd) and 17.4% (by Cα-atom Irmsd) moved closer to the bound form than the corresponding unbound structure at the known interface. In case of the 149x38 models generated for the ligand proteins, 22.3%/24.5% (all-atom / Cα-atom Irmsd) moved on average closer to the bound form (Figure 4.2). In total, 135 by all-atom Irmsd or 136 by Cα-atom Irmsd out of 149 targets in the benchmark showed at least some geometric improvement compared to the unbound form on at least one binding partner (*Appendix* Figure S20). The movement of the structure closer to the bound form (compared to the corresponding unbound form) reached, however, only up to 0.5Å.

4.3.2 Comparison of REMC ensemble and minimization docking approach

Our REMC docking approach consists of translational and rotational moves of the binding partners relative to each other as well as of random selection of conformers from the ensemble pool. To enhance the sampling, replica exchanges between four parallel runs at four different temperatures were performed (see Methods).

This differs from a straightforward and standard ensemble docking approach that corresponds to a systematic cross docking of all the possible receptor-ligand pairs of the ensembles. The latter method is computationally very expensive, and also difficult in the analysis stage because of the huge number of generated decoys. We first evaluated our REMC ensemble docking by comparing to a complete cross docking with minimization (MIN-cross-docking) using default settings from ATTRACT easy online [160] on all 39x39 receptor-ligand combinations on two of the test cases pdb1T6B and pdb1J2J. This resulted in 76,589,760 and 30,782,964 decoys for 1T6B and 1J2J, and spent 28.5 hours and 12 hours with MIN-cross-docking, respectively, in sampling with one GPU-CUDA device (NVIDIA-GTX980). The ATTRACT score vs. interface RMSD (Irmsd) from

MIN-cross-docking indicates dense sampling along the Irmsd coordinate and reveals the location of several docking funnels with few points indicating favorable docking sites (Figure 4.3a).

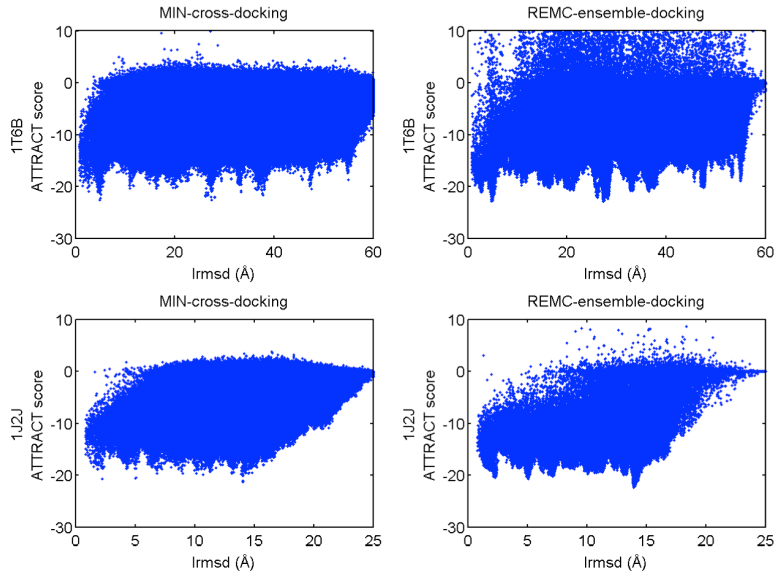

Figure 4.3. Distribution of sampled energy scores vs. Irmsd for a complete MIN-cross-docking and REMC-ensemble docking on two test cases pdb1T6B and pdb1J2J.

Basically, the same energy landscape along the Irmsd coordinate with the same binding funnels can be obtained using the REMC ensemble docking approach (Figure 4.3b) based on the same ensemble as used for the exhaustive MIN-cross-docking approach. Since an MC approach gives the highest density of states near the lowest energy solutions the density of sampled configurations is much higher in the regime of the favorable docking sites than in case of the exhaustive MIN-cross-docking approach (Figure 4.3). This can be achieved within a smaller computer time of 8 h and 2.2 h (on a single GPU) for the two 1T6B and 1J2J test targets, respectively.

In the following, we compare four different sampling protocols for global docking on 149 targets from the docking benchmark 4.0 [141]: minimization for unbound docking (MIN-unbound), REMC for unbound docking (REMC-unbound), minimization for ensemble docking (MIN-ensemble) and REMC for ensemble docking (REMC-ensemble). Ensemble docking has been run with 38 generated models and 1 unbound structure for both binding partners (ensemble-39). Note, that in case of the MIN-ensemble docking before the starting of the minimization one pair of structures is taken at random from the ensemble and minimized.

For the minimization approach, the program was run on 2.6 GHZ AMD CPU clusters using 48 cores for each target with running time ranging from 2 minutes to 3 hours

depending on the size of the target. For REMC approach, the program was run with GPU accelerated version using 4 CUDA devices for each target, with slightly longer runtime of 20 minutes to 4 hours. The number of generated decoys for each target is the same across the four protocols, and ranges between 106,000 to 656,000 depending on the target size.

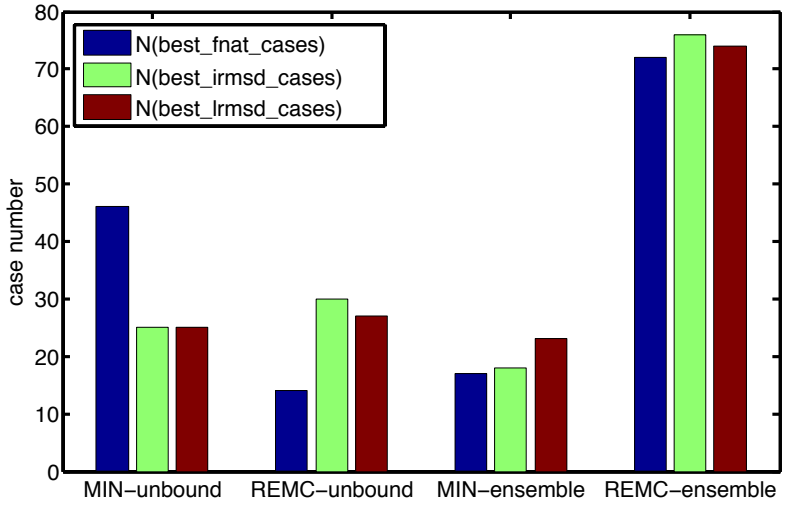

Figure 4.4. Number of best docking solutions in terms of fnat, Irmsd and Lrmsd for Min-unbound, REMC-unbound, MIN-ensemble and REMC-ensemble docking not considering score ranking of solutions. The sum of the bars with the same color equals to 149 (the total number of target).

We first compare the best fnat, best Irmsd and best Lrmsd (with respect to the bound structure) of the sampled decoys for each of the 149 targets (without looking at score ranking). We count the number of cases for each protocol that it has the best solution in terms of fnat, Irmsd and Lrmsd cross the four protocols as shown in Figure 4.4. Strikingly, REMC-ensemble docking almost always reaches better quality solutions with regard to all the three metrics for over twice the targets comparing to the other three protocols. For minimization approaches, ensemble docking on average reaches better solutions for about twice the number of targets as in unbound docking except for fnat (MIN-ensemble 17 vs. MIN-unbound 21). If comparing ensemble docking and unbound docking collectively, ensemble docking has reached better fnat, Irmsd and Lrmsd on 89, 94 and 97 targets comparing to 60, 55 and 52 for unbound docking. This is suggesting that even though the generated models are not optimal, they do help in improving docking predictions in terms of generating better quality solutions.

We then plotted the distribution of top-100 fnat of each target across the whole benchmark for the four docking protocols (Figure 4.5). We can see that the distribution is sampling method related. REMC sampling, especially combined with ensemble docking, always recovers more native contacts in a lot of more decoys. It is especially of interest in the range of higher fnat. In the selected 149x100 top fnat structures, REMC sampling has

around 70% for ensemble docking and 50% for unbound docking with over 50% native contacts recovered compared to about 38% structures for minimization approaches. Furthermore, with a stricter criterion checking only for fnat solutions above 80%, the improvement of REMC sampling compared to the minimization approach is even more prominent. It is surprising that REMC sampling still has over 3% structures when we consider over 90% native contacts recovery, comparing to below 0.1% for minimization approach. These results show that the REMC method samples more high quality decoys, at least in terms of the fnat metric (see also Figure 4.6).

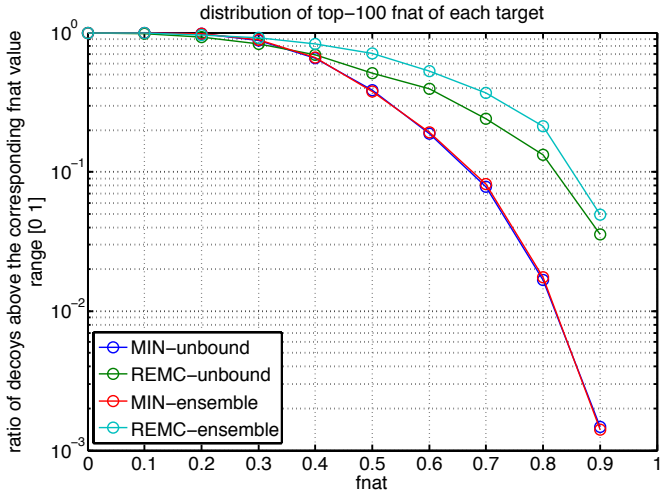

Figure 4.5. Distribution of top-100 fnat across the whole benchmark for the four docking protocols. Each circle represents the ratio of decoys above the corresponding fnat value. The fnat (x-axis) range is [0 1]. The maximum of y-axis is 1.

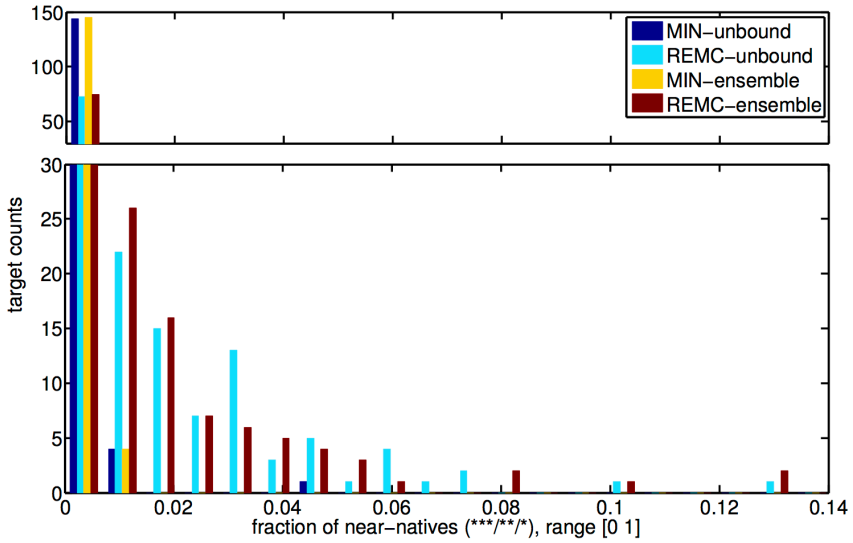

Figure 4.6. Distribution of the fraction of near-native (at least one CAPRI star) decoys for each protocol.

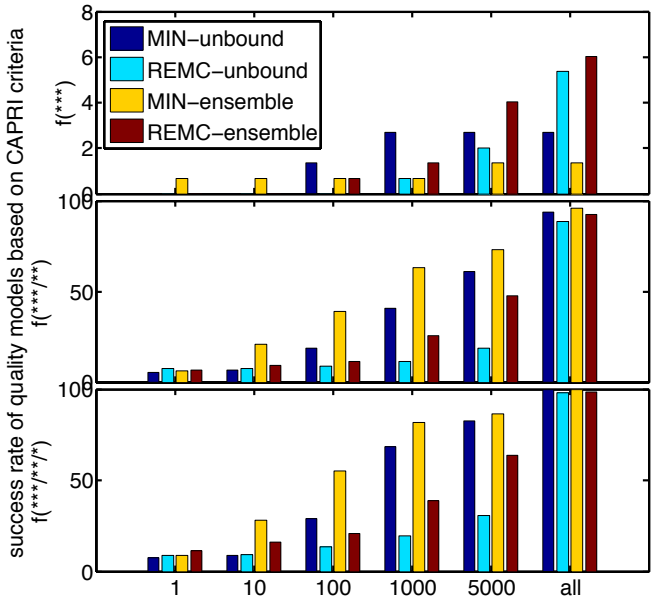

Figure 4.7. Success rate of CAPRI quality decoys in the top N decoys selected by ATTRACT score.

Table 4.2. Summary of the CAPRI star case number each protocol has best reached.

CAPRI quality	MIN-unbound	REMC-unbound	MIN-ensemble	REMC-ensemble
***	4	8	2	9
**	136	124	141	131
*	9	14	6	7
Failed	0	3	0	2

Considering the success rate for sampling medium (**) or high quality (***) solutions REMC-ensemble docking also performs slightly better than the other protocols (Table 4.2) but it failed on two cases, 2A5T and 2J7P (1BGX, 1IB1 and 2J7P for REMC-unbound docking). This is reasonable because in those cases, the ATTRACT energy in the near native regions is far too high comparing to alternative binding regions. We further exploited the success rate within selected top N decoys by ATTRACT score to check the selectivity of the energy function (Figure 4.7). Here, the ranking of quality solution in the top predictions is slightly better for the minimization vs. the REMC approaches. Apparently, if a high quality solution is found, it is slightly better ranked by energy minimization than by REMC because in the same energy funnel, energy minimization will always reach lower energies than a MC approach. However, as an importance sampling method, REMC spends more time in low-energy regions not necessarily close to near-native (see also *Appendix* Figure S21). When we compare ensemble docking and unbound docking within each sampling method, minimization or REMC, with the same number of selected top-

energy decoys, the success rate for ensemble docking is almost always higher than single model unbound docking.

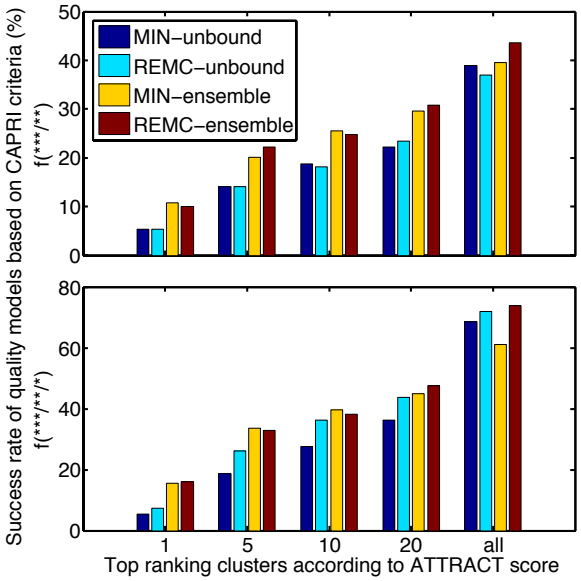

Figure 4.8. Success rate of CAPRI quality clusters for the top scoring clusters of docking solutions. Same color scheme as in Figure 4.7.

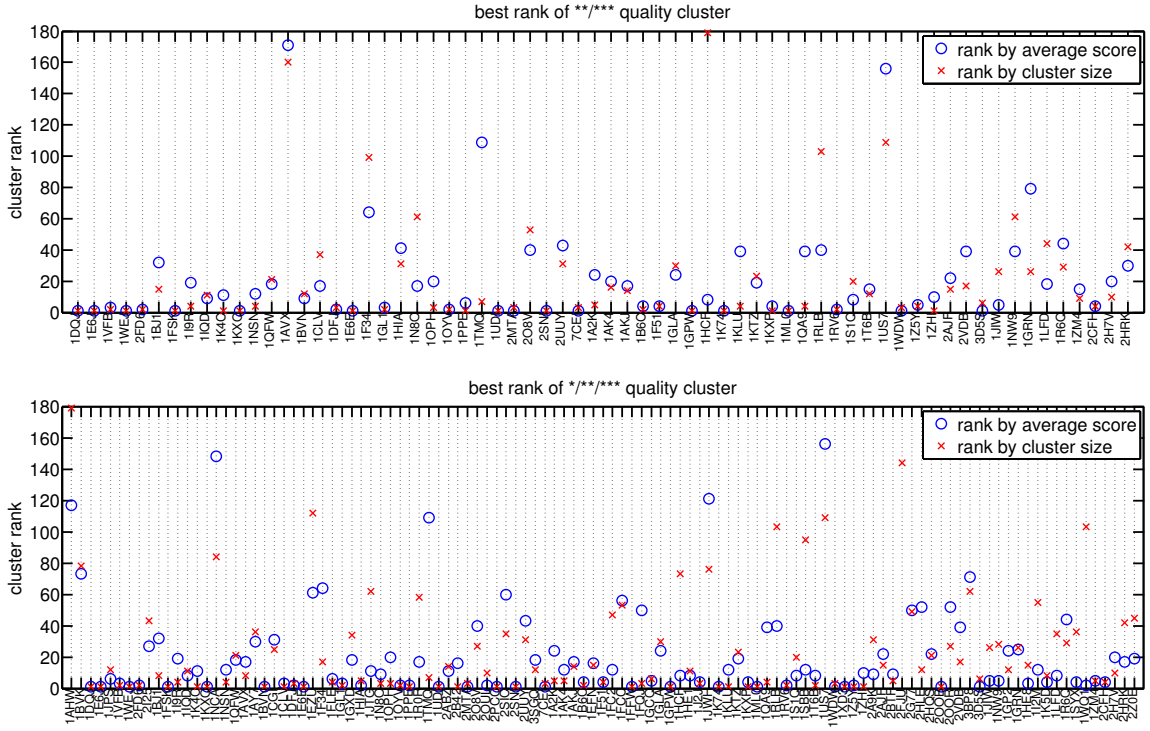

Figure 4.9. Comparison of cluster ranking by average score and by cluster size on the REMC-ensemble-39 docking results.

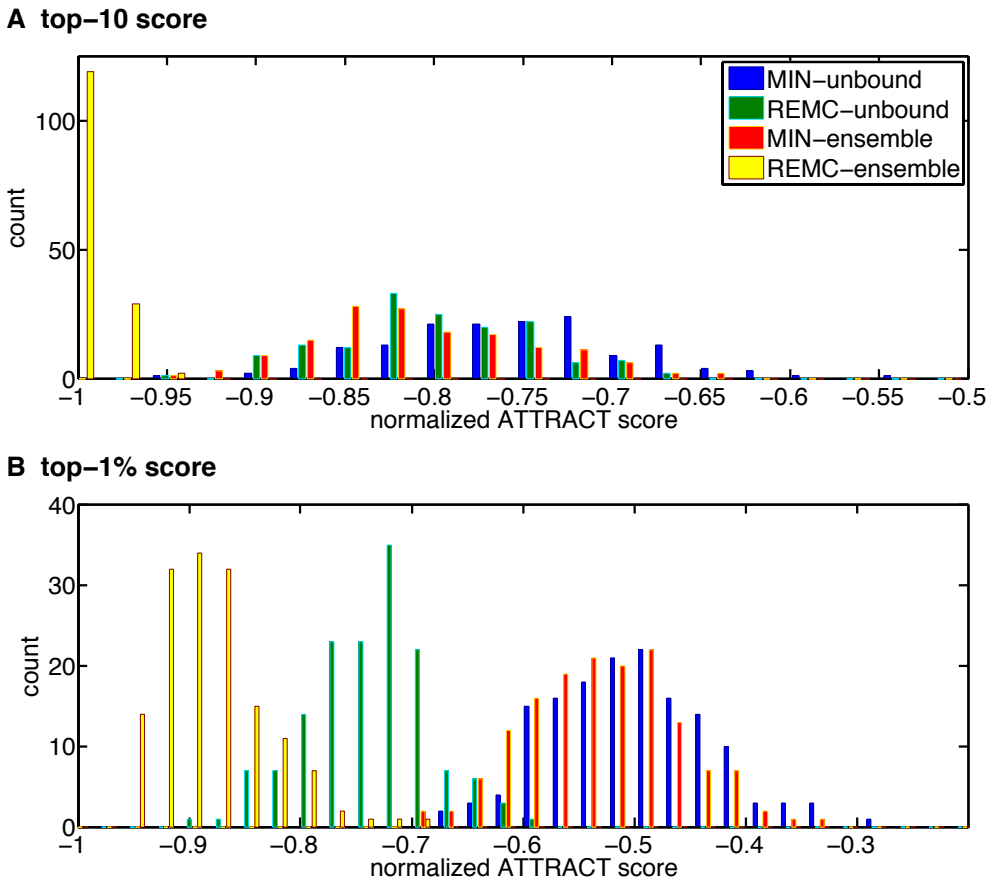

Figure 4.10. ATTRACT score energy distribution obtained with four simulation protocols. The energies for each target and method have been normalized by the absolute value of the best energy observed for the respective target across all of the four protocols. Shown are the distribution of A) top-10 energies and B) top-1% tile energies across the whole benchmark.

In Figure 4.8 we plot the success rate in the top ranking clusters by ATTRACT score to exploit if clustering will reveal an improvement of scoring upon REMC sampling. The score of a cluster is the average value of the five representing decoys: the center of the cluster and the top-four score decoys. After clustering, only one target from one protocol (1PPE from MIN-ensemble docking) has still high (***) quality clusters. If we count high/medium (***/**) and high/medium/acceptable (***/**/*) quality clusters, ensemble docking is more successful regard to the success rate compared to the single body unbound docking (Figure 4.8).

Based on this impression on REMC docking results, we investigated if ranking the clusters by the cluster size alone allows already the identification of near native docking geometries. Comparison with ranking using the scores of clusters indicates that indeed the cluster size criterion alone is in most cases as useful as the best rank scoring (Figure 4.9). Apart from the ability to sample near-native regions it is also interesting to explore the energy optimization efficiency of the protocols. It is possible to normalize the

ATTRACT scores by the absolute value of the best energy observed for the respective target across all of the four methods (Figure 4.10). When considering both top 10 and top 1 %-tile energies, REMC-ensemble docking reaches much lower ATTRACT energies in most cases compared to the other three protocols.

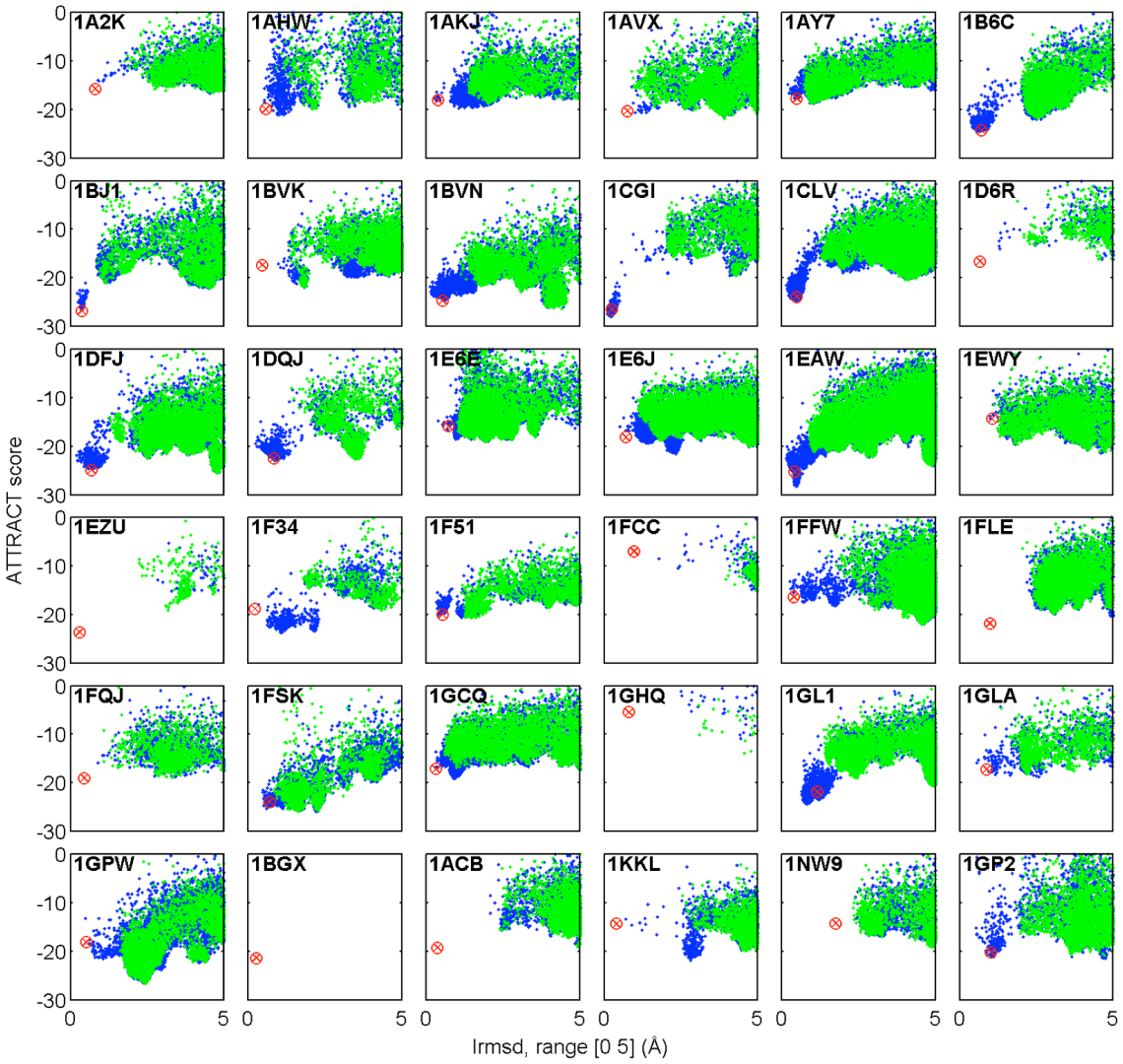

Figure 4.11. ATTRACT score versus interface RMSD (Irmsd) in the near-native (below 5 Å) region on the REMC ensemble docking application to 36 targets. The range of x-axis [0 5] Å and y-axis [-40 0] are all the same for all panels. Results from REMC-ensemble-40 (included bound form partner) and REMC-ensemble-39 are shown in blue and green, respectively. The Irmsd and scoring energy of the energy minimized bound complex is shown as red circle with cross in each panel.

It is also of interest to the test the performance of the REMC ensemble docking in case of including the bound structures of each partner in the ensemble. In this case the ensembles contained 40 models (included the respective bound partner, denoted as REMC-ensemble-40) and we considered a subset of benchmark cases (36 cases). In the majority (but not all) of cases (26 of 36) inclusion of the bound structures resulted in

improved docking results and docking geometries in closer agreement with the native structure were sampled (Figure 4.11). However, in a subset of cases (pdb-entries: 1BVK, 1D6R, 1EZU, 1FCC, 1FLE, 1FQJ, 1GHQ, 1BGX, 1ACB, 1NW9) inclusion of the bound partner structures did not always improve the docking results because the near native sampling region scored much worse than alternative regions in the energy landscape (even for the bound partner pair, see *Appendix* Figure S22). This result indicates that in these cases the sampling quality is not limiting but limitations in the docking scoring function are decisive and even an improved ensemble (with conformations closer to native) may not improve the docking performance.

4.4 Conclusion

A REMC protein-protein docking protocol has been designed that includes partner flexibility by picking partner structures in an MC approach from pre-generated conformational ensembles. The MC switching between conformers is such that at each trial step conformers with close similarity to the current conformer are selected. The approach is similar to methods used in small-ligand docking to an ensemble of receptor conformers [60-63] but has so far not been applied to protein-protein docking. It outperforms a standard all-against-all docking of each ensemble pair of partner structures in terms of computational speed and accumulates more low energy structures and also on average more structures close to the native docking solution. Previous protein-protein docking efforts based on ensemble representation of protein partners have also found improved performance compared to docking just the unbound partner proteins [34,57,58,157]. Although the structural ensemble used for docking contained only few conformers in better agreement with the bound form than the unbound partner structure the REMC ensemble docking gave overall better results in terms of scoring and number of native contacts or Irmsd of the final solutions.

A possible reason is the use of a coarse-grained model in ATTRACT that may allow for improved fit of docking partners in near-native geometry even if the structures still deviate from the bound form (e.g. toleration of alternative packing of residues at the interface). However, also non-native docking geometries can profit from the use of an ensemble of partner conformations resulting in more low-energy predictions that deviate from the native structure. Indeed, *Appendix* Figure S20 demonstrates this effect on two examples. Inclusion of the bound partner structures in the docking ensemble gives in the majority of cases docking solutions closer to native with favorable score. This indicates that the ensemble of protein conformers is nevertheless critical for the success of the approach and ensemble members in closer agreement with the bound form can further improve the performance.

The focus of the present study is not on designing an optimal ensemble but to demonstrate that even a non-optimal ensemble combined with the REMC method can already improve the protein-protein docking performance. Designing an optimal ensemble with more structures closer to the bound form will be subject of future research. Our current experience and limited success on employing several standard methods indicates that this is a difficult task that may also involve systematic force field improvements and inclusion of explicit solvent molecules during the ensemble generation step.

4.5 My contributions to this project

I have conceived and designed the research work together with Prof. Dr. Martin Zacharias. I have implemented the protocol together with Uwe Ehmann. All the calculations and analysis are carried out by me. And I have wrote the paper.

Chapter 5 Perspectives on Bayesian inference in Monte Carlo sampling data driven protein-protein docking

5.1 Introduction

Studies have shown that integrating experimental or predicted information about the interaction to bias the sampling and/or scoring of protein-protein docking can consequently increased the accuracy of the docking predictions [16,168]. NOE data from NMR experiment gives specific short interacting distances; Förster resonance energy transfer, chemical cross linking and electron paramagnetic resonance spectroscopy are able to provide unambiguous distances over larger distances; NMR chemical shift perturbation allows to narrowing down the interface surface as been employed in Haddock; SAXS, small angle neutron scattering (SANS) as well as cryo-EM can provide shape information about the complex. All the prior information could be used one way or another. The most simple way is implemented as a post filter. That is to evaluate on the generated decoys of its agreement with the prior information. This imposes no influence upon the generation of the decoys. Another more comprehensive way is through a restraint energy term explicitly included in the sampling, as being used in Haddock, RosettaDock and so on. In Haddock [38], the prior information is translated into a set of ambiguous interaction restraints (AIRs) and calculated with a harmonic potential to penalize the violation against the upper limit of the effective distance which is formulated as follows:

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N_{atoms}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}=1}^{N_{atoms}} \frac{1}{d_{m_{iA}n_{kB}}^{6}} \right)^{-\frac{1}{6}}$$

where N_{atoms} indicates all atoms of a given residue and N_{res} the sum of active and passive residues for a given protein. The atom-atom distances contribute inversely to this effective distance. As long as a residue comes in close with any active or passive residue of the partner molecule, the restraint will be satisfied. Besides, Haddock in each trial by default uses only 50% randomly selected AIR restraints to counteract possible false positives in the prior information.

In our canonical Monte Carlo sampling framework, we could integrate the prior information in a Bayesian inference fashion. This suites better in the context when the data is sparse, of low resolution or might be ambiguous, erroneous or even conflicting.

With Markov chain Monte Carlo (MCMC) we can construct an ensemble with a defined distribution as $\{x_i\} \sim P(X|Y)$ where X is the conformation and Y is the prior knowledge. According to Bayes' theorem, the probability changing from state x_i to x_j given prior data Y can be reformulated as follows:

$$\frac{p(x_j|Y)}{p(x_i|Y)} = \frac{p(Y|x_j)p(x_j)/p(Y)}{p(Y|x_i)p(x_i)/p(Y)} = \frac{p(Y|x_j)}{p(Y|x_i)} * \frac{p(x_j)}{p(x_i)}$$

That is:

$$p(x|Y) \sim p(Y|x)p(x)$$

This could be explained as the posterior probability of conformation x given prior information Y is the product of prior probability and the likelihood of the conformation [169-171]. In standard Monte Carlo simulation, the acceptance of a state is given by:

$$p = min(1, e^{(-\Delta E/kT)})$$

Thus the posterior probability could be written as:

$$p(x|Y) \sim e^{ln(p(Y|x)} * e^{(-E(x)/kT)}$$

That is:

$$p(x|Y) \sim e^{\left(-\frac{E(x)-kT*ln(p(Y|x))}{kT}\right)}$$

For implementation, $-kT * ln(p(Y|x)$ makes a new term into the energy function and

$$E_{total}(x) = E(x) - kT * ln(p(Y|x))$$

For the likelihood $p(Y|x)$, an unfounded but convenient assumption of amino acid pair independence leads to the expression:

$$p(Y|x) = \prod_{i=1}^{n} p(y_i|x)$$

For application, assume we have a set of data of contacting or no-contacting atom pairs. But we are not 100% confident with the accuracy and correctness of the data. Based on the common sense, atom pairs within a certain distance cutoff have a

higher probability (denotes as λ) to be observed in experiment in contact, while pairs beyond this cutoff have a higher probability (denotes as μ) to be observed not in contact. Then we have:

$$p(y_{contact}|x) = \begin{cases} \lambda & d < d_{cutoff} \\ 1-\mu & else \end{cases}$$

$$p(y_{no\ contact}|x) = \begin{cases} 1-\lambda & d < d_{cutoff} \\ \mu & else \end{cases}$$

5.2 Preliminary Results and Discussion

This has been implemented into Rosetta and tested by running the low-resolution replica exchange docking protocol [41] using artificial data on unbound docking target 1BVN and 2OOB. We have used only the Ca atom for each residue. In the test, I have used 0.7 for λ, 0.95 for μ and 8Å for Ca atom distance cutoff d_{cutoff}. For example for 2OOB, we assume that we have prior information from experiment or evolutionary analysis about interacting / non-interacting residue pairs shown in Figure 5.1. For the blue dashed line connected pairs, we give true information about if they are in contact. For the pair with red dashed line, residue 4A and B39, we suppose the prior information suggesting they are in contact to simulate error in experiment observations. Last but not least, the prior information is not limited to pairwise information. It can also be in the form of in or not in binding site from only one side of the binding partners.

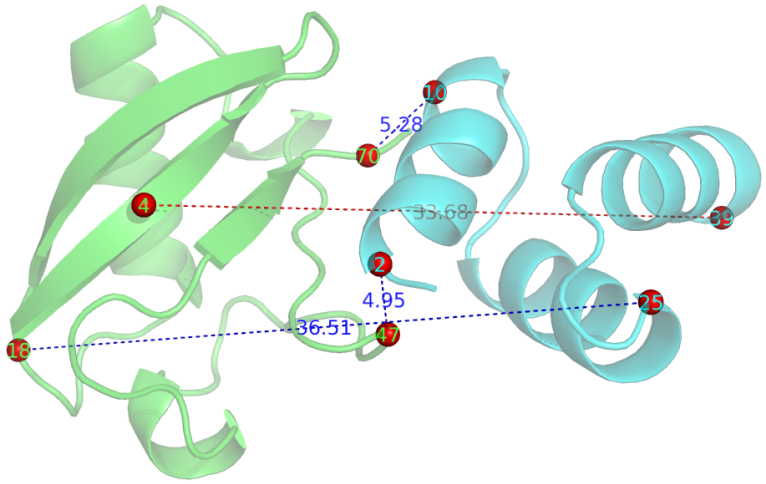

Figure 5.1. The native structure of 2OOB with Ca-atom distances labeled for the residue pairs which have been used for testing Bayesian inference approach for integrative docking.

Bayes inference approach for prior data

λ	0.7		
μ	0.95		
d_{cutoff}	8		
70A	10B	in contact	# true
47A	2B	in contact	# true
4A	39B	in contact	# simulate error in observation
18A	25B	no contact	# true

AtomPair Constraint

AtomPair	CA	70A	CA	10B	BOUNDED	0	8	5	CST
AtomPair	CA	47A	CA	2B	BOUNDED	0	8	5	CST
AtomPair	CA	4A	CA	39B	BOUNDED	0	8	5	CST
SiteConstraint		CA	18A	B	SIGMOID	5	-2	#discourage contact	

Ambiguous AtomPair Constraint

AmbiguousConstraint									
AtomPair	CA	70A	CA	10B	BOUNDED	0	8	5	CST
AtomPair	CA	47A	CA	2B	BOUNDED	0	8	5	CST
AtomPair	CA	4A	CA	39B	BOUNDED	0	8	5	CST
END_AMBIGUOUS									
SiteConstraint		CA	18A	B	SIGMOID	5	-2	#discourage contact	

Ambiguous AtomPair Constraint

AmbiguousConstraint									
AtomPair	CA	70A	CA	10B	SIGMOID	0	8	5	CST
AtomPair	CA	47A	CA	2B	SIGMOID	0	8	5	CST
AtomPair	CA	4A	CA	39B	SIGMOID	0	8	5	CST
END_AMBIGUOUS									
SiteConstraint		CA	18A	B	SIGMOID	5	-2	#discourage contact	

For comparison, I have also run the same low-resolution replica exchange docking protocol, but with the prior information used as Rosetta standard constraints. The profile of the constraint functions are shown in Figure 5.2.

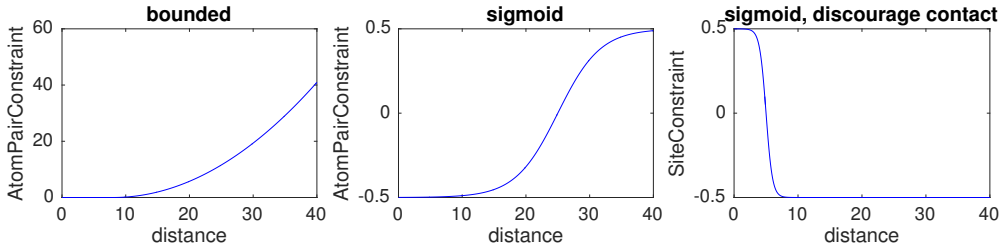

Figure 5.2. Profile of constraint functions from Rosetta, which are used for comparison with Bayes inference approach.

In our simulation, the weight for likelihood in Bayes approach, AtomPairConstraint and SiteConstraint are 5, 5 and 50, respectively. In Figure 5.3, we show the scatter plot from replica exchange low-resolution docking with no prior information. In stead of near-native region, the low-resolution energy function favors structures of about 7 and 10 Å away from the native. In Figure 5.4, we show in the first column the scatter plot of total score versus the Ca-RMSD from the native structure, in the second column score for evaluating the prior information versus Ca-RMSD. We can see that the Bayes inference approach is most solid and tolerant with the misleading prior information. Further more, the Bayes inference approach has also benefited greatly from the included correct prior information and leads the docking procedure to the correct direction with lower scores for the low Ca-RMSD structures. However, for the standard constraints, when all pairs applied as shown in the second row, the constraint between residue 4A and B39 has strongly biased the docking result to a completely wrong region. When using ambiguous constraint as shown in the third and fourth rows, the bottom of the scatter plot of the score evaluating the prior information versus Ca-RMSD is rather flat. The bottom of the scatter plot of total score versus Ca-RMSD is also flattened. This is better than that in the case with no prior information as shown in Figure 5.3 the decoys below 5 Å have much higher energies than binding modes with about 7 and 10 Å away from native. This is suggesting that using the prior information as the traditional constraint can be helpful, but with limited effect, and not as helpful as in Bayes inference approach.

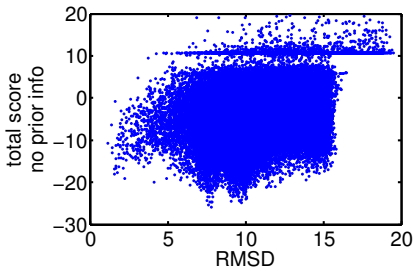

Figure 5.3. Total score vs. Ca-RMSD of unbound docking on 2OOB from replica exchange Monte Carlo sampling low-resolution docking in Rosetta with no prior information applied.

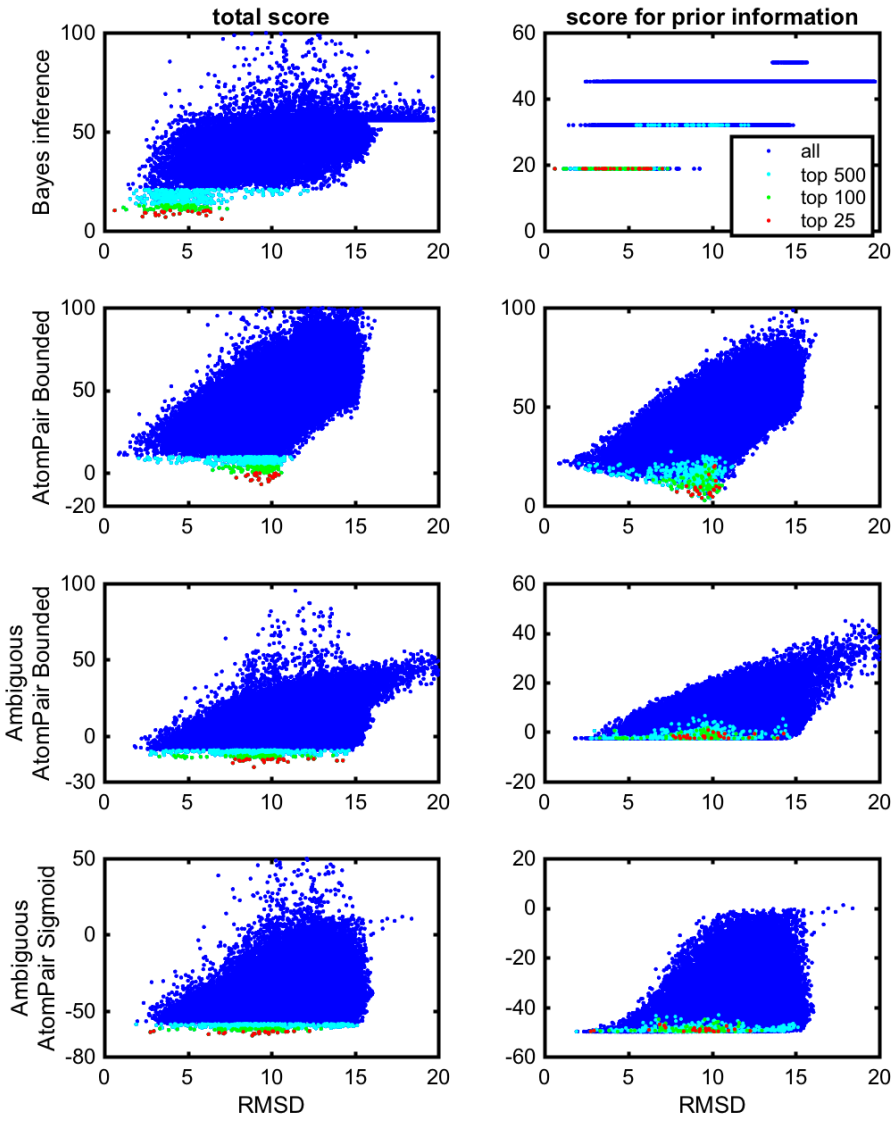

Figure 5.4. Docking results with prior information applied in Bayesian inference fashion and in the form Rosetta constraints. The first column shows the total score versus Ca-RMSD. The second column is the score evaluating only the prior information versus Ca-RMSD. Red, green and cyan dots correspond to top 25, top 100, top 500 decoys by energy in each panel.

Chapter 6 Conclusion

Proteins are virtually involved in all biological events that taking place within or between cells, ranging from enzyme catalysis and inhibition to signal transduction and regulation. In carrying out these functions, proteins rarely work alone, but rather in close association with other biomolecules. In fact, many biological processes are carried out by large molecular machines whose action is coordinated through intricate regulatory networks of transient macromolecular interactions, among which protein-protein interaction counts the largest proportion. To fully understand the mechanism underlying these biological processes and further serve for pharmaceutical purposes, high resolution structural is extremely valuable. However, the classical experimental high resolution structure determination methods, X-ray crystallography and Nuclear Magnetic Resonance, are surrounded by the experimental bottlenecks when coming to the structure determination of complexes.

This thesis studies the recognition of macromolecules, to be more specific, recognition between protein and proteins, through computational method — protein-protein docking. The focus of this work is investigating the application of Monte Carlo and Monte Carlo based sampling methods in addressing the computational protein-protein docking problem. Following the three theoretical model of protein-protein interaction: key and lock, induced fit and conformational selection, I have developed three Monte Carlo method based docking protocols, with each corresponding to a different scenario in docking practice.

Rigid-body docking, stemmed from the key and lock model, commonly applies a systematic search (enumeration) as used in ZDOCK, or shotgun approach, in which thousands of independent random-start trajectories minimise the rigid-body degrees of freedom. I have in **Chapter 2** introduced an alternative strategy, ReplicaDock, using a small number of long trajectories of replica exchange Monte Carlo to sample putative interaction geometries. I have compared ReplicaDock sampling the low-resolution stage of RosettaDock with RosettaDock's original shotgun sampling as well as with ZDOCK. Better performance is observed for ReplicaDock and ZDOCK comparing to shotgun sampling while benchmarking. ReplicaDock and ZDOCK consistently reach lower energies and generate significantly more near-native conformations than shotgun sampling. Accordingly, they both improve the CAPRI quality metrics of complex structures also after refinement. Additionally, I have shown that the refined ReplicaDock ensembles

reach significantly lower interface energies and many previously hidden features of the docking energy landscape become visible when ReplicaDock is applied.

Induced fit model justifies the many docking refinement protocols developed in the community. They carry out minor yet non-trivial adjustments from the primary predictions from rigid-body docking, achieving optimisation of energy and further facilitating the discrimination between near native and non-near native conformations. In Chapter 3 I have explored the efficiency of Monte Carlo based methods sampling the docking refinement stage with atomistic representation in Rosetta. Long trajectories are started from a random position with defined deviation from native conformation and explored with the rigid-body space fairly restricted to simulate the docking scenario that the binding site is approximately known in docking practice. Several enhanced sampling techniques, including temperature or Hamiltonian replica exchange and well-tempered ensemble approaches have been combined with the Monte Carlo method and were evaluated on 20 protein complexes using unbound partner structures. The well-tempered ensemble method combined with a two dimensional temperature and Hamiltonian replica exchange scheme (WTE-H-REMC) was identified as the most efficient search strategy. Comparison with prolonged standard MC searches indicates that the WTE-H-REMC approach requires approximately 5 times fewer MC steps to identify near native docking geometries compared to conventional MC searches.

In **Chapter 4** I have developed a replica exchange Monte Carlo (REMC) ensemble docking approach in docking engine ATTRACT that allows efficient exploration of protein-protein docking geometries. In addition to Monte Carlo steps in translation and orientation of binding partners, possible conformational changes upon binding are included based on Monte Carlo selection of protein conformations stored as ordered pre-generated conformational ensembles. The conformational ensembles of each binding partner protein were generated by three different approaches starting from the unbound partner protein structure with a range spanning a root mean square deviation of 1-2.5 Å with respect to the unbound structure. Since MC sampling is performed to select appropriate partner conformations on the fly, the approach is not limited by the number of conformations in the ensemble compared to ensemble docking of each conformer pair in ensemble cross docking. Although only a fraction of generated conformers was in closer agreement with the bound structure, the REMC ensemble docking approach achieved improved docking results compared to REMC docking with only the unbound partner structures or using docking energy minimisation methods. The approach has significant potential for

further improvement in combination with more realistic structural ensembles and better docking scoring functions.

In **Chapter 5**, I have integrated the prior interacting or non-interacting information in our canonical Monte Carlo sampling framework in a Bayesian inference fashion. It suites especially better in the context when the prior data is sparse, of low resolution or might be ambiguous, erroneous or even conflicting. For comparison, I have run the same low-resolution replica exchange docking protocol on unbound docking target 2OOB, with artificial prior data, which includes both correct and incorrect information about interacting/non-interacting, used in Bayesian inference fashion and Rosetta standard constraints. Simulation results demonstrate that the Bayesian inference approach is most solid and tolerant with the misleading prior information. Using the prior information as the traditional constraints can be helpful, but with limited effect when the data is erroneous. The Bayesian inference approach can nevertheless benefit the most from the correct one among the mixed information and lead the docking procedure to the correct direction with lower scores for lower RMSD structures.

Appendix

Supplementary Figures

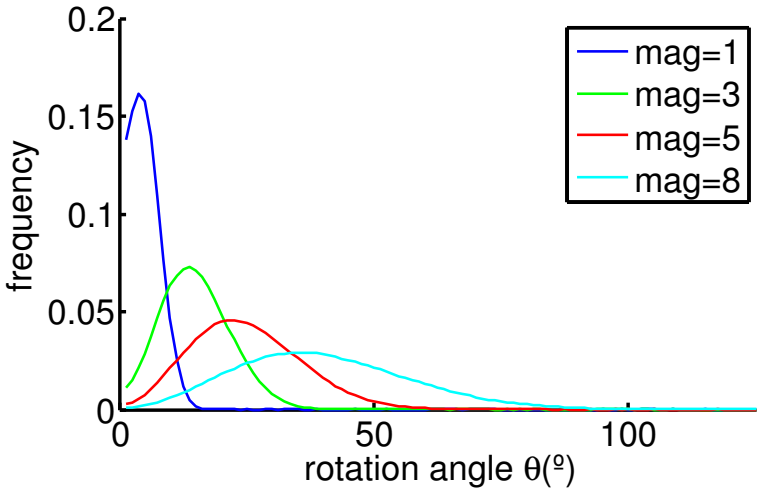

Figure S1: The distribution of rotation angles generated by Rosetta's RigidBodyPerturbNoCenterMover for various "magnitude" parameters. Blue, green, red and cyan correspond to different input parameters of the magnitude.

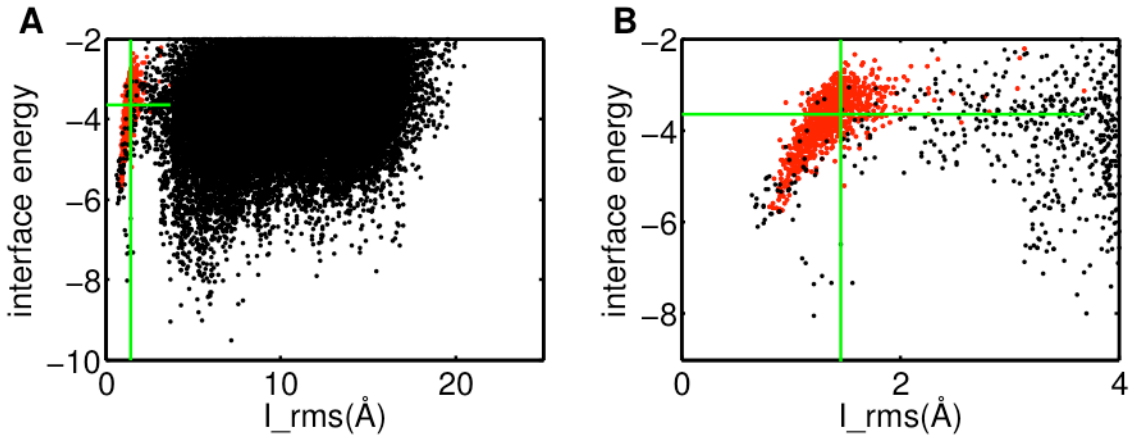

Figure S2: Example of refined decoys overlap with the RelaxedNative ensemble. A) red dots represent the RelaxedNative ensemble, black dots represent decoys refined from centroid decoys generated with one of the three methods shotgun, ZDOCK or ReplicaDock. The green lines define the lower left region and correspond to 50%-tile interface energy and 75%-tile I_rms of RelaxedNatives, respectively. B) zoomed figure of A).

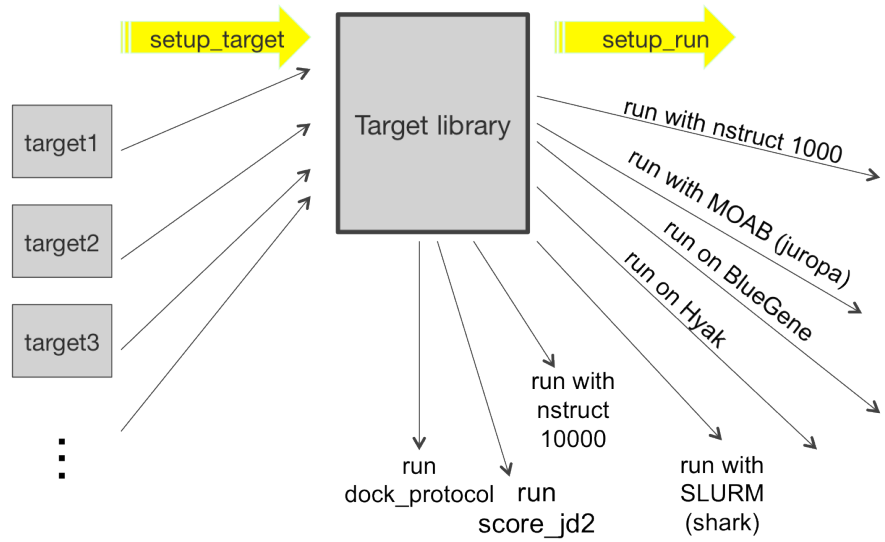

Figure S3: Automated setup work flow.

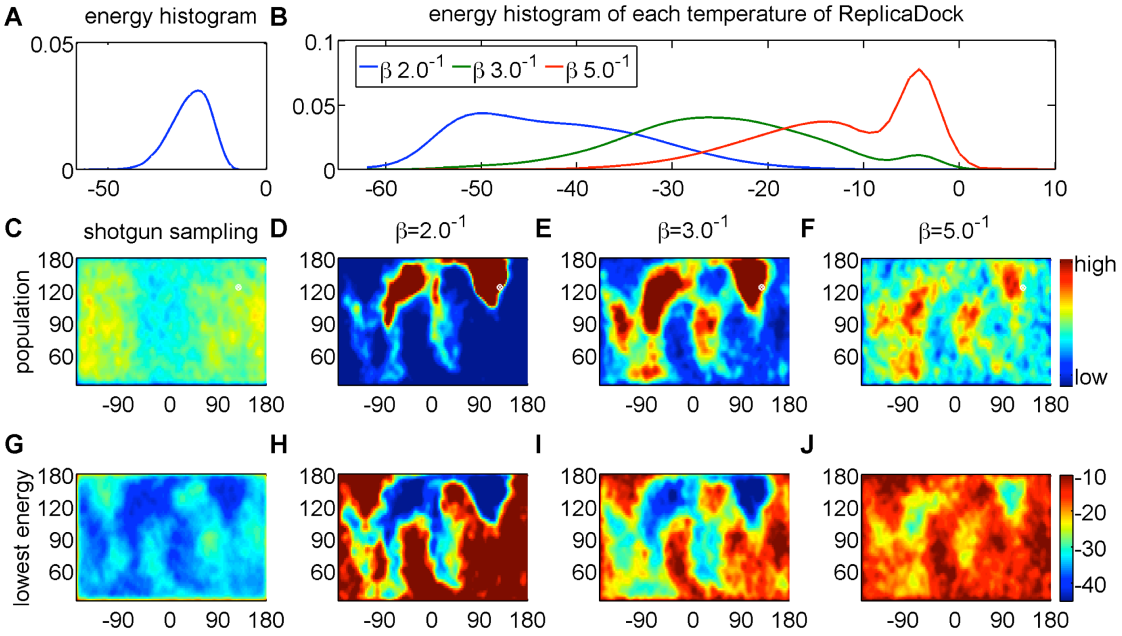

Figure S4: Detailed analysis of shotgun and ReplicaDock sampling on a bound target 1sq2. A) energy distribution of shotgun sampling generated low-resolution decoys. B) energy distribution of conformations sampled by ReplicaDock at respective temperature levels. C-F) Population of sampled conformations in spherical coordinates. Partner A is fixed at the center and the position of Partner B with respect to an idealized spherical surface around Partner A is recorded. The native structure is labeled as white dot (arrow in C). G-J) Conformations are assigned to grid-cells as in C-F, but shown is the lowest energy of all conformations assigned to the respective grid cell. The same color-scale is used for each plot of a row, and the colorbars are attached to the rightmost panel.

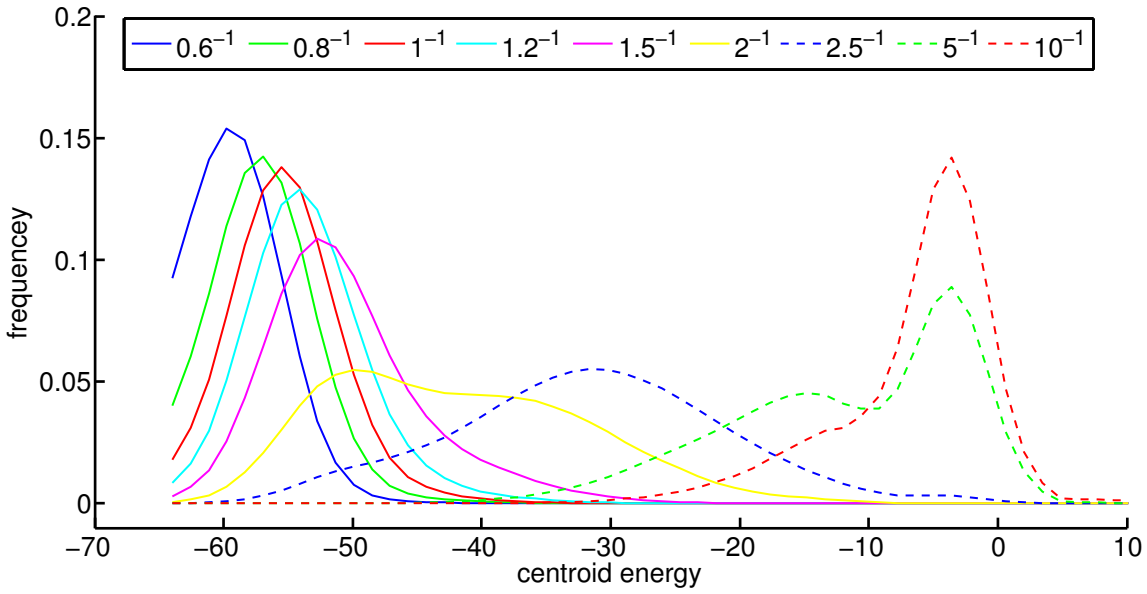

Figure S5: Initial survey of temperature selection on bound target 1sq2. Energy distribution of conformations sampled by ReplicaDock at respective temperature levels.

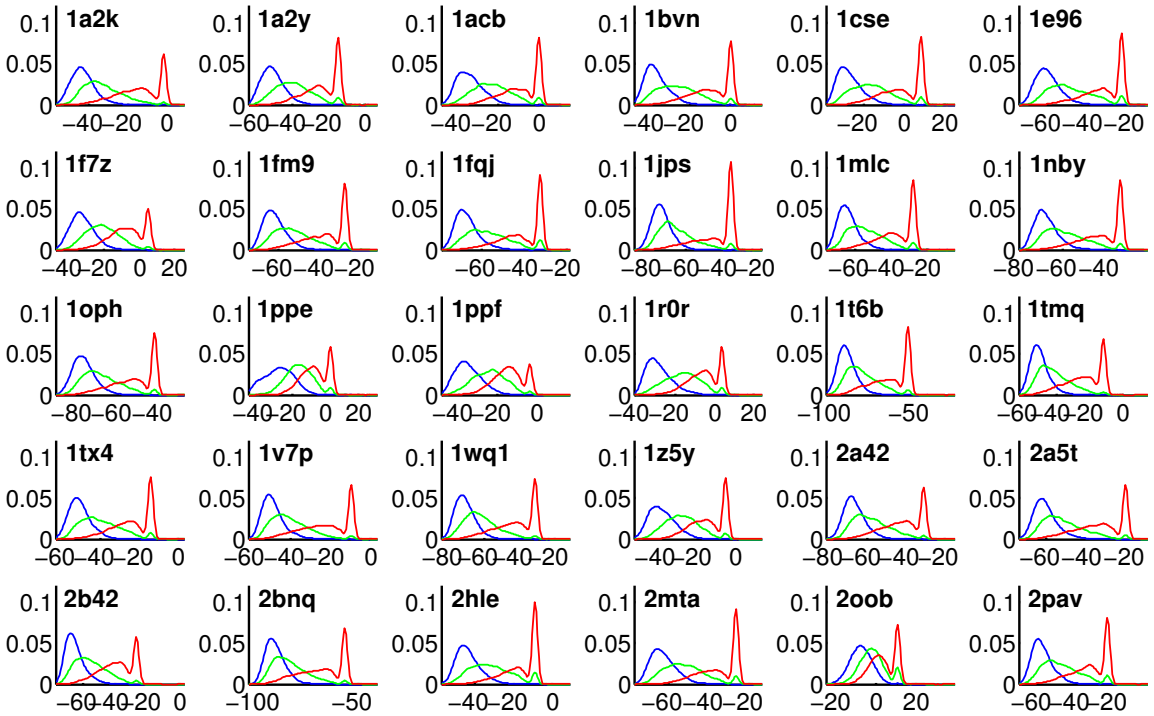

Figure S6: Energy distribution of the three temperatures in ReplicaDock for all the benchmark targets. Blue, green and red correspond to the inverse temperatures 2^{-1}kcal^{-1}·mol, 3^{-1}kcal^{-1}. mol and 5^{-1}kcal^{-1}. mol. Note that the overall value of the *interchain_env* term is highly target dependent (mainly due to system size), such that the position of the energy-peak reflecting non-contacting conformations changes drastically between targets.

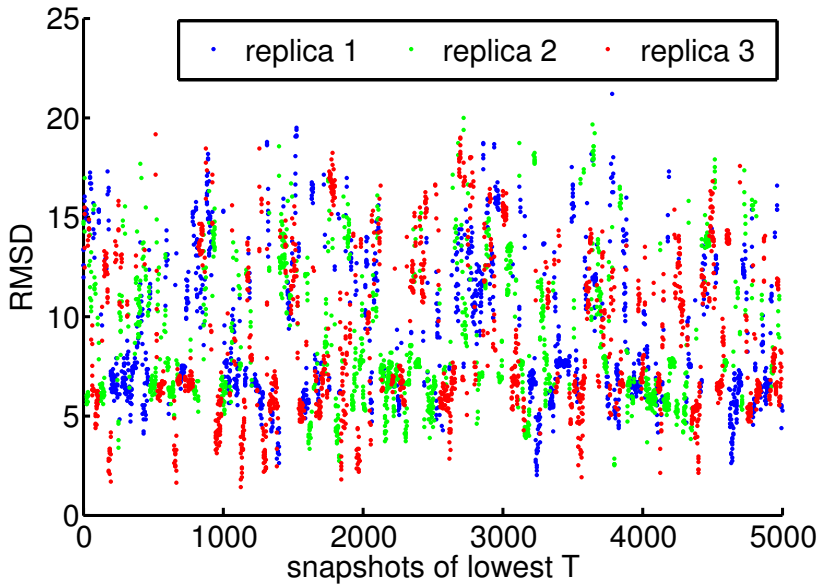

Figure S7: Frequent exchange between bound and unbound state on target 1ppf. RMSD of the snapshots of lowest temperature conformations from the sampled trajectory. 3 colors corresponding to the 3 replicas.

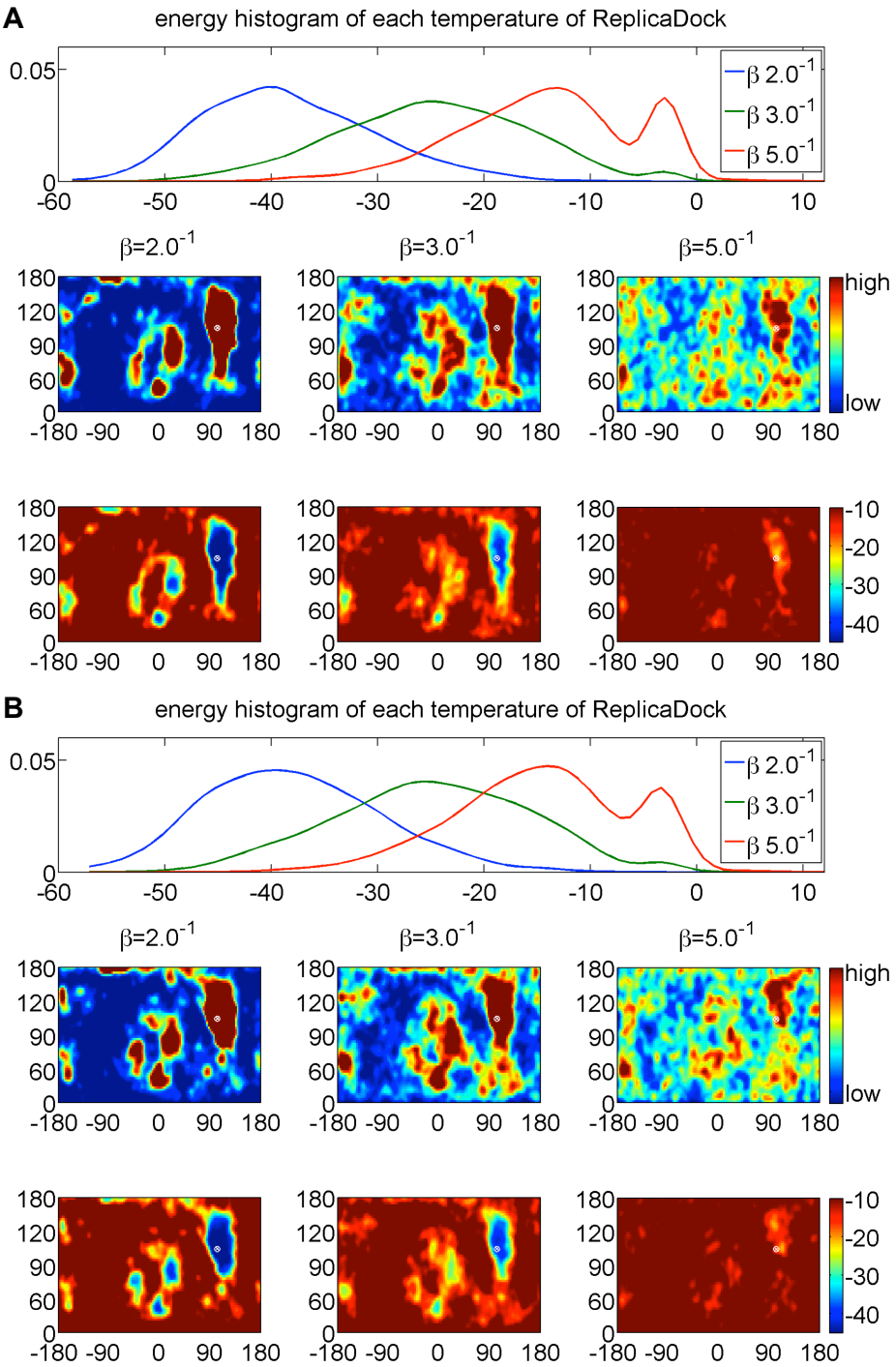

Figure S8: Different start conformations converge to the same populations. Panel A) and B) correspond to ReplicaDock trajectories with different start conformations. Each panel is equivalent to Figure 2.2 of the main text. (first row) energy distribution of conformations sampled by ReplicaDock at respective temperature level; (second row) population of sampled conformations in spherical coordinates in respective temperature level; (third row) lowest energy of the conformations assigned to each grid.

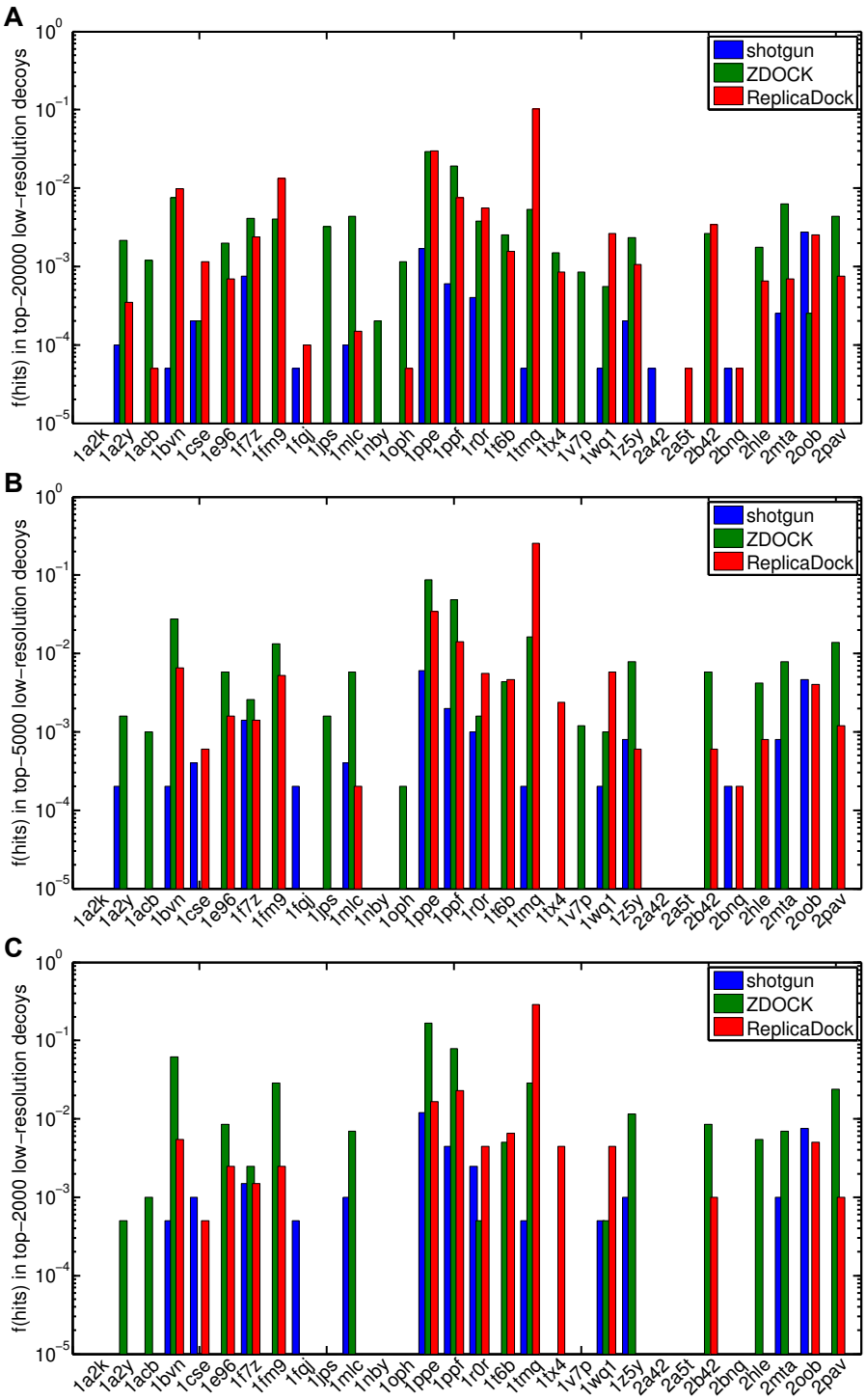

Figure S9: Fraction of hits in low-resolution decoys with different cutoffs. A) Fraction of hits in top-2000 low-resolution decoys, B) Fraction of hits in top-5000 decoys, C) fraction of hits in top-20000 decoys. Blue, green and red correspond to shotgun, ZDOCK and ReplicaDock. Shotgun and ReplicaDock decoys are selected by RosettaDock centroid energy, ZDOCK decoys are selected by ZDOCK score.

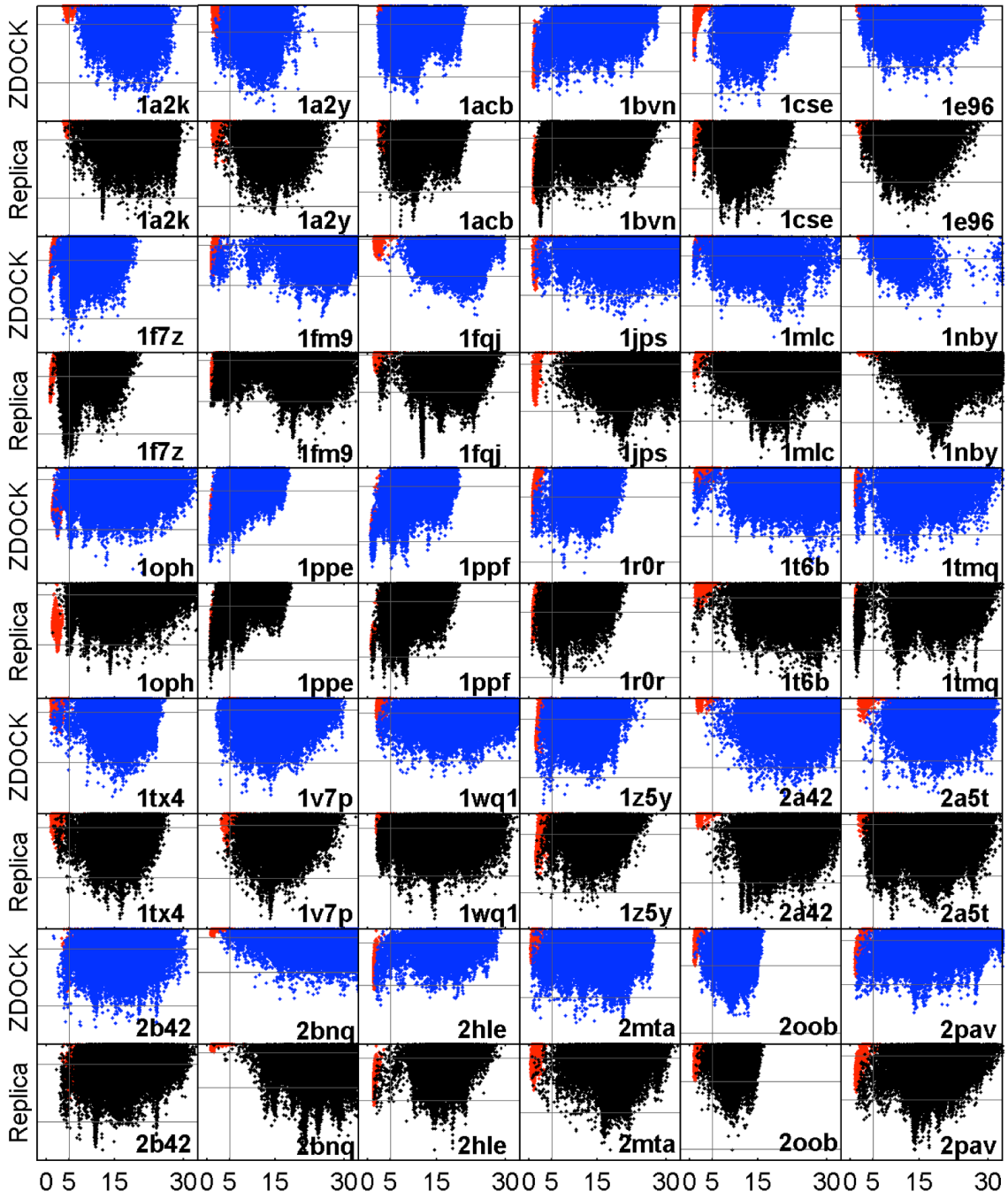

Figure S10: Interface RMSD vs. Interface Energy after refinement of ZDOCK and ReplicaDock ensembles. The red dots represent the RelaxedNative ensembles(Results). The interface RMSD is shown on the x-axis, the interface energy on the y-axis. The same energy range is used for displaying both, ZDOCK (blue) and ReplicaDock (black), results of each target, respectively. The vertical gray lines correspond to I_rms of 5.0 Å, and the two horizontal gray lines correspond to interface energy -4 and -8 Rosetta Energy Units.

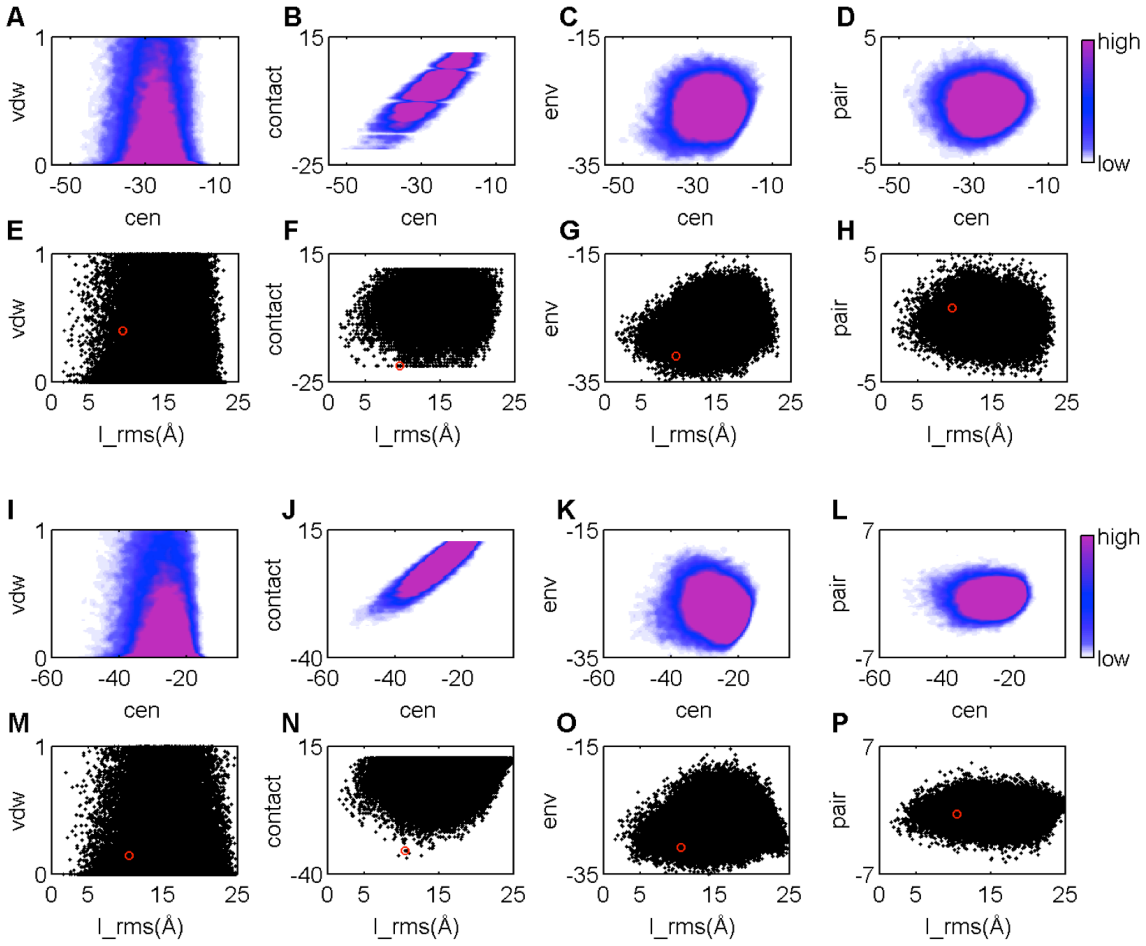

Figure S11: *Interchain_contact* dominates docking centroid energy. A-H) shotgun sampling with capped low-resolution energy function on bound target 1emv, I-P) shotgun sampling without energy capping (Section 2.2.1). A-D, I-L) total centroid energy versus each individual centroid energy term, i.e. *interchain_vdw, interchain_contact, interchain_env and interchain_pair* in order. E-H, M-P) I_rms versus each individual centroid energy term. The red circle in M-P) corresponding to the denoted decoy in Figure 2.7 of the main text. The decoy in E-H) denoted as red circle has similar structure to the denoted decoy in Figure 2.7 of the main text.

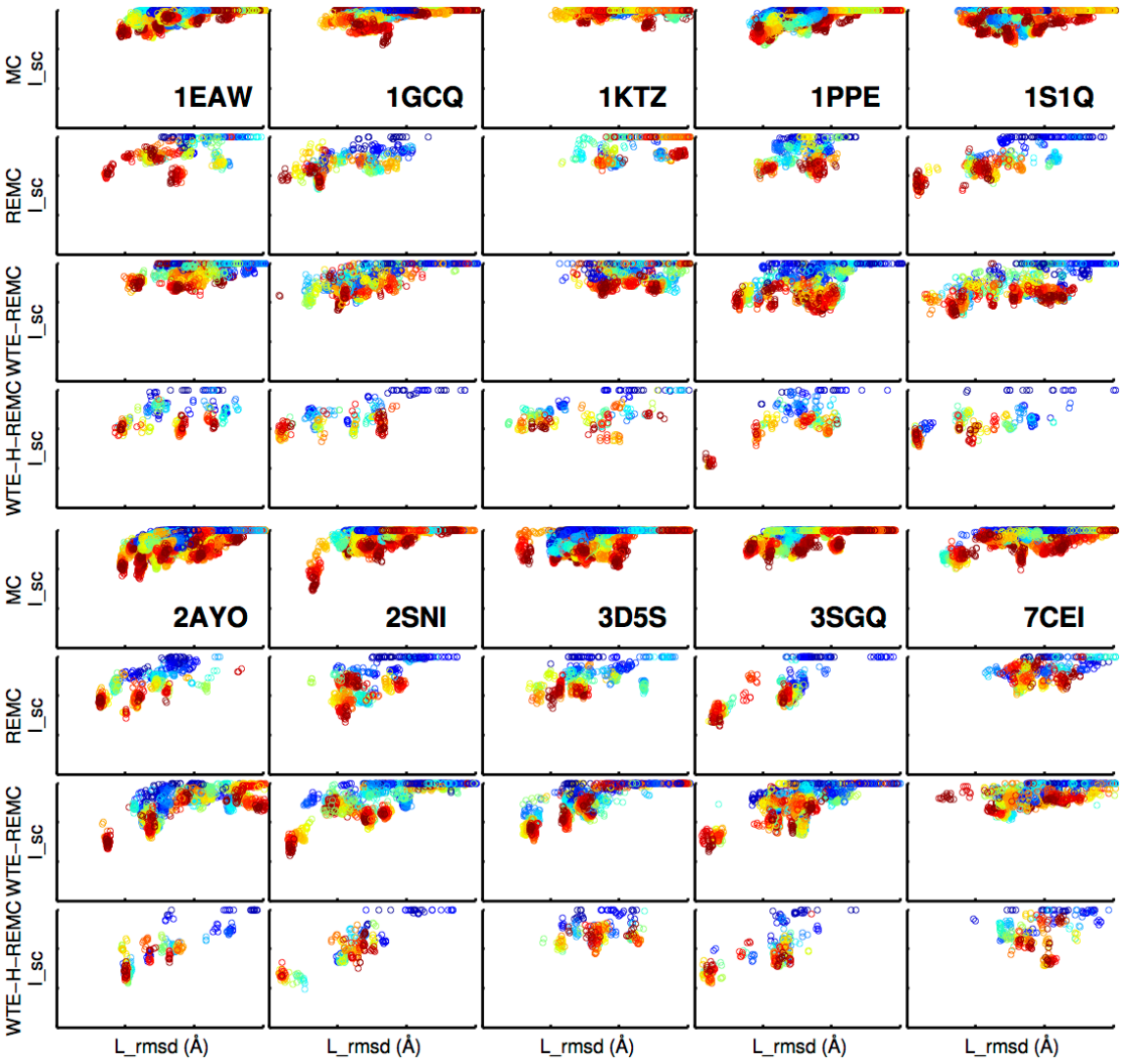

Figure S12. Scatter plot of interaction score (I_sc) vs. ligand RMSD (L_rmsd) for the first 10 targets after $3x10^5$ MC steps. All the panels have the same L_rmsd range of [0..30], and the same I_sc range of [-15..0 Rosetta score units]. For each target, the tested protocols are grouped together and the corresponding protocol is indicated in the score-axis label on the left side. The snapshot number is color-coded, with dark blue and dark red dots corresponding to decoys sampled at the beginning and towards the end of the sampling interval, respectively.

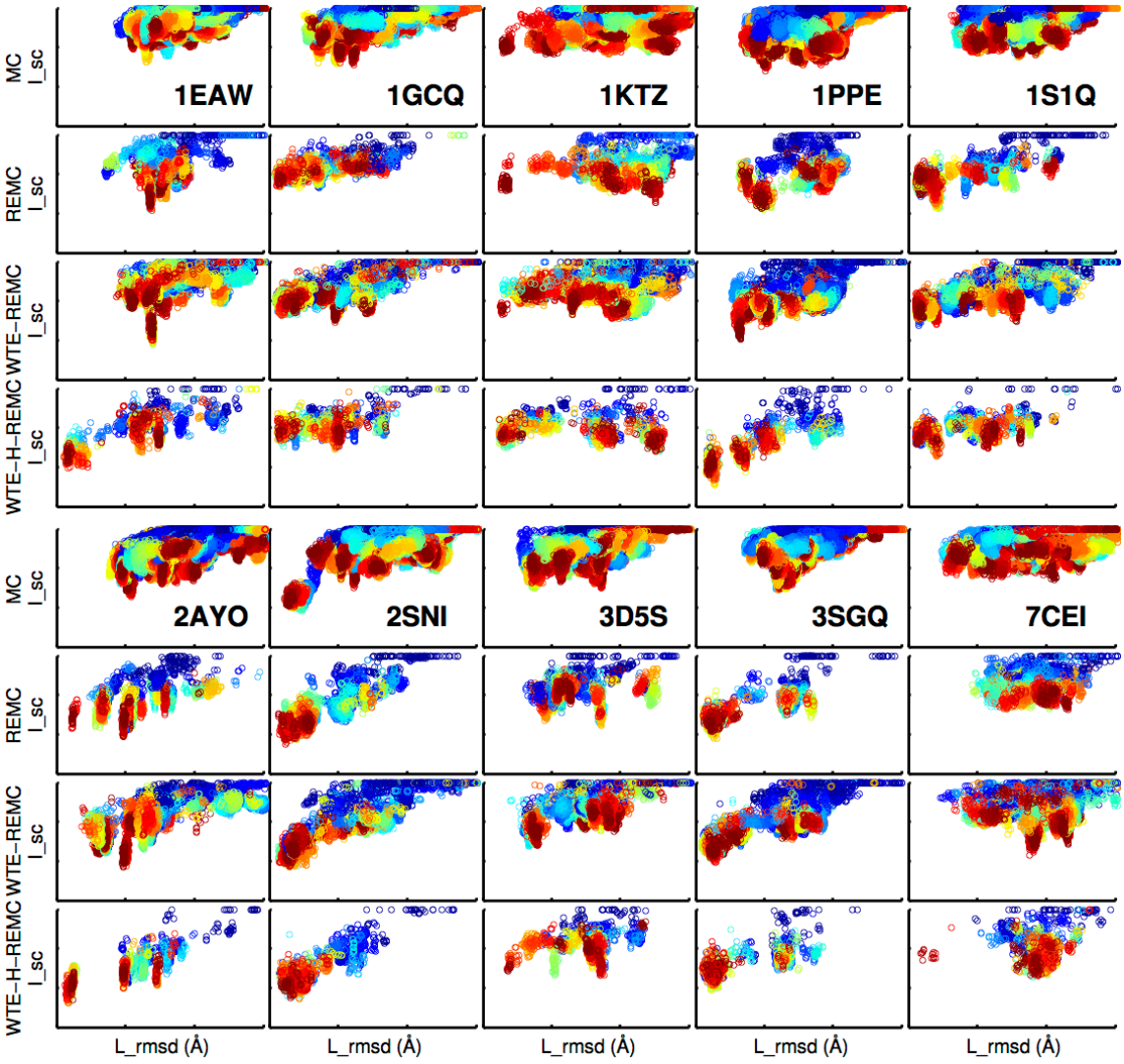

Figure S13. Same as Figure S12 but for the docking searches up to 2×10^6 MC steps.

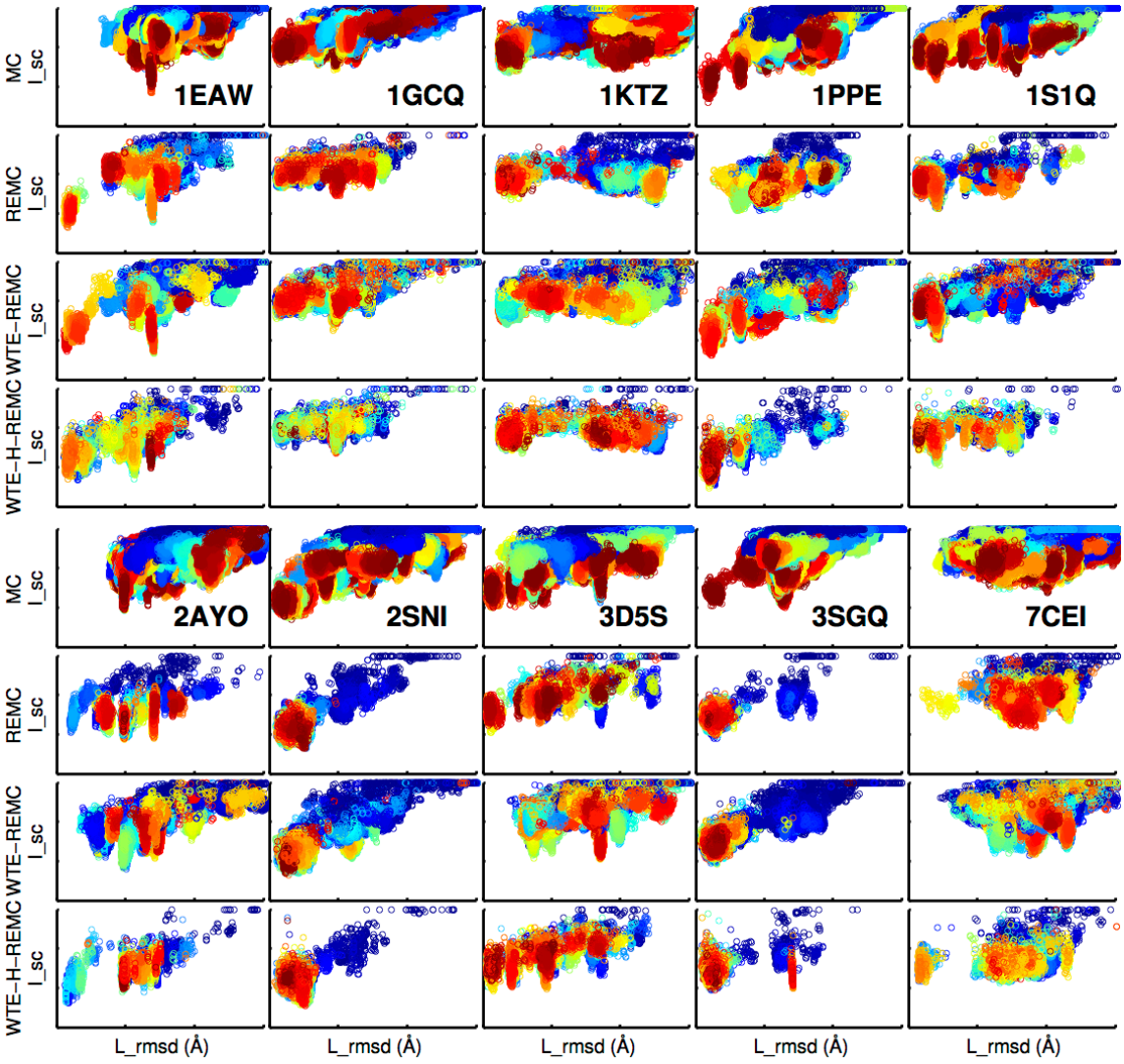

Figure S14. Same as Figure S12 but for the docking refinement simulation with 10^7 MC steps.

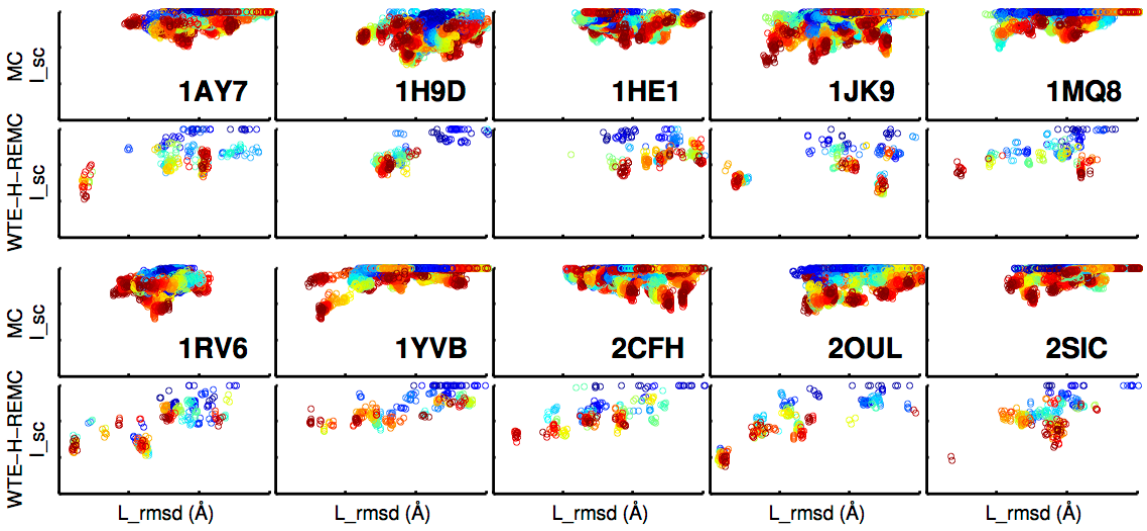

Figure S15. Scatter plot of interaction score (I_sc) vs. ligand RMSD (L_rmsd) for the additional 10 targets with $3x10^5$ MC steps. For each target, the tested protocols are grouped together and the corresponding protocol is indicated in the score-axis label on the left side. The snapshots number is color-coded, with dark blue and dark red dots corresponding to decoys sampled at the beginning and the end, respectively.

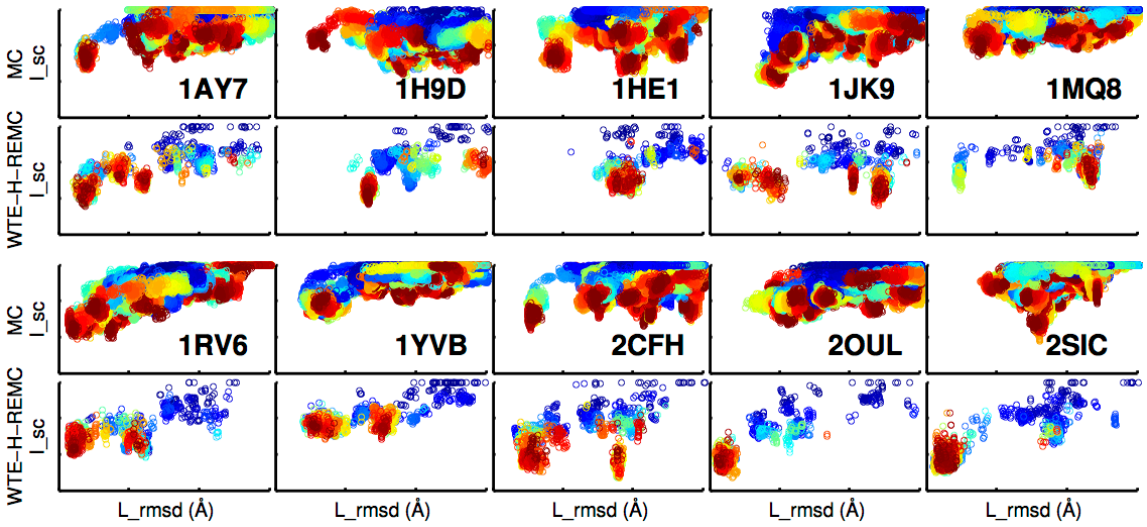

Figure S16. Same as Figure S15 but for docking searches up to $2x10^6$ MC steps.

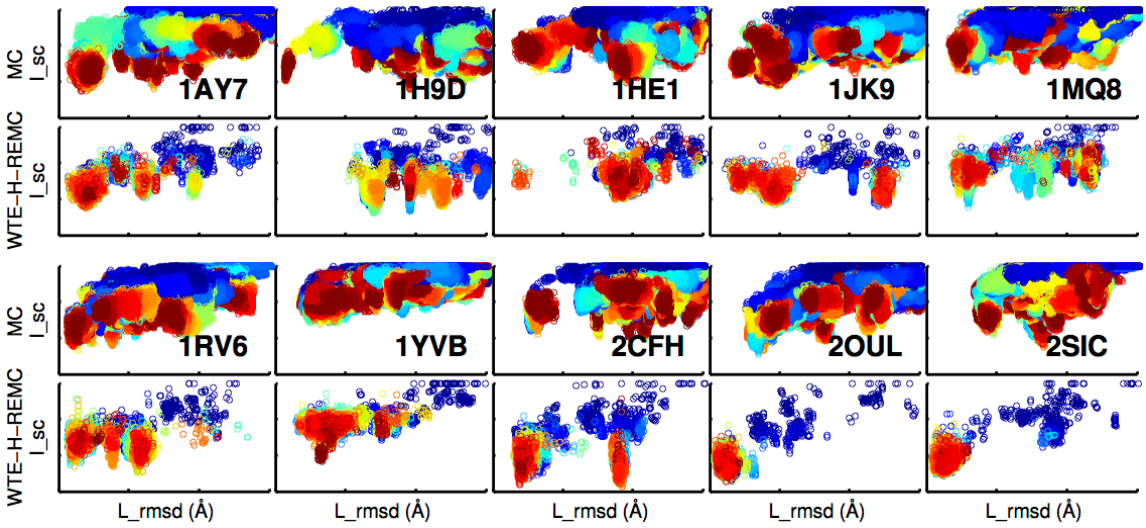

Figure S17. Same as Figure S15 but for docking searches with 10^7 MC steps.

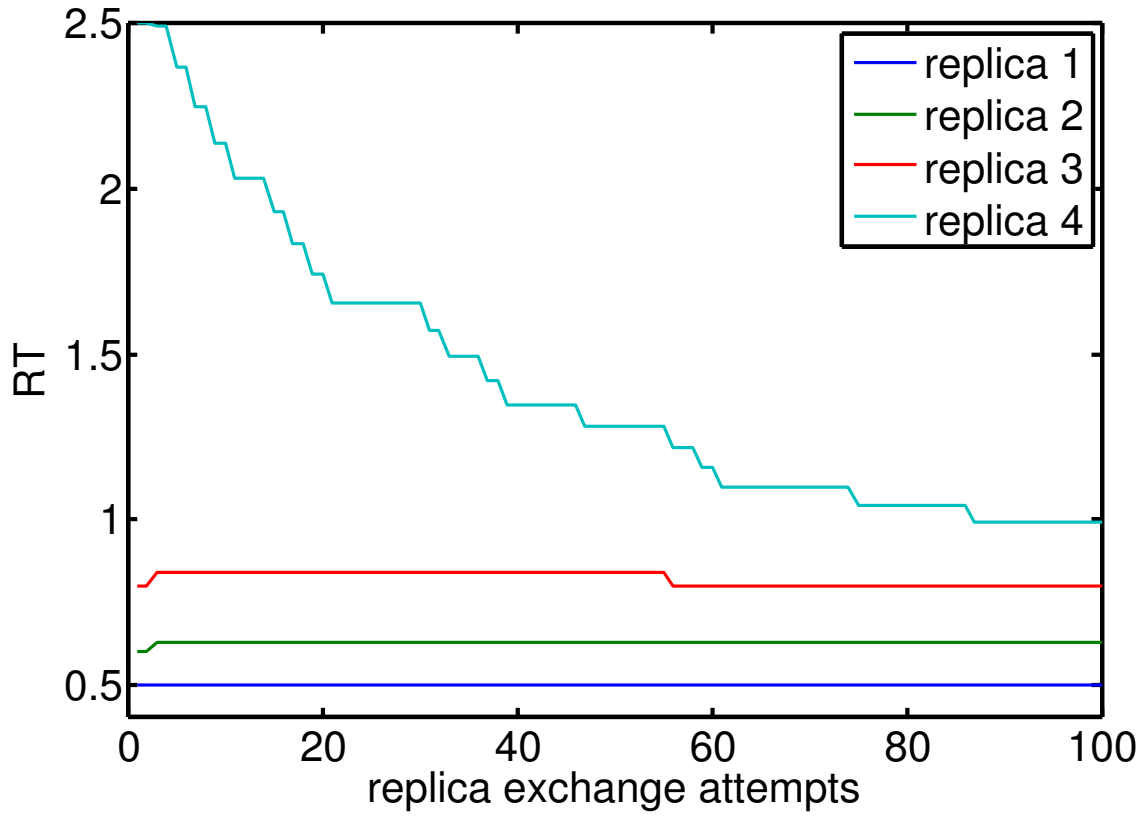

Figure S18. The evolution of the RT values for the temperature on one example (REMC-ensemble docking 1A2K). The RT values of the lowest three replicas are generally very stable, the value of the highest replica normally converges after 80 or less replica exchange attempts. The value of RT of the highest replica converging to varies from target to target.

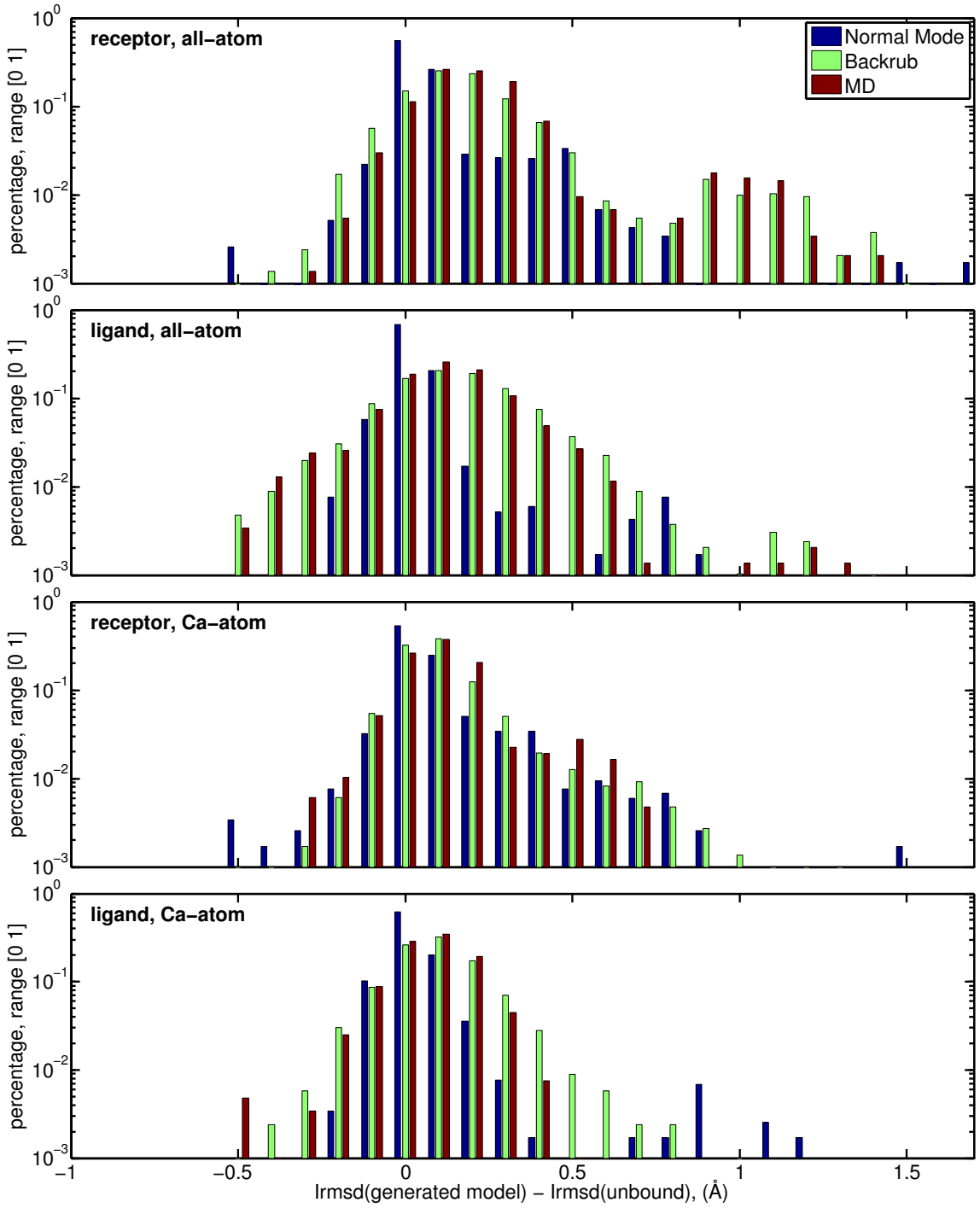

Figure S19. Distribution of interface RMSD (Irmsd) differences between generated models and the respective starting unbound structure. Irmsd is calculated with native bound structure as reference, using all-atom and only Ca-atom respectively.

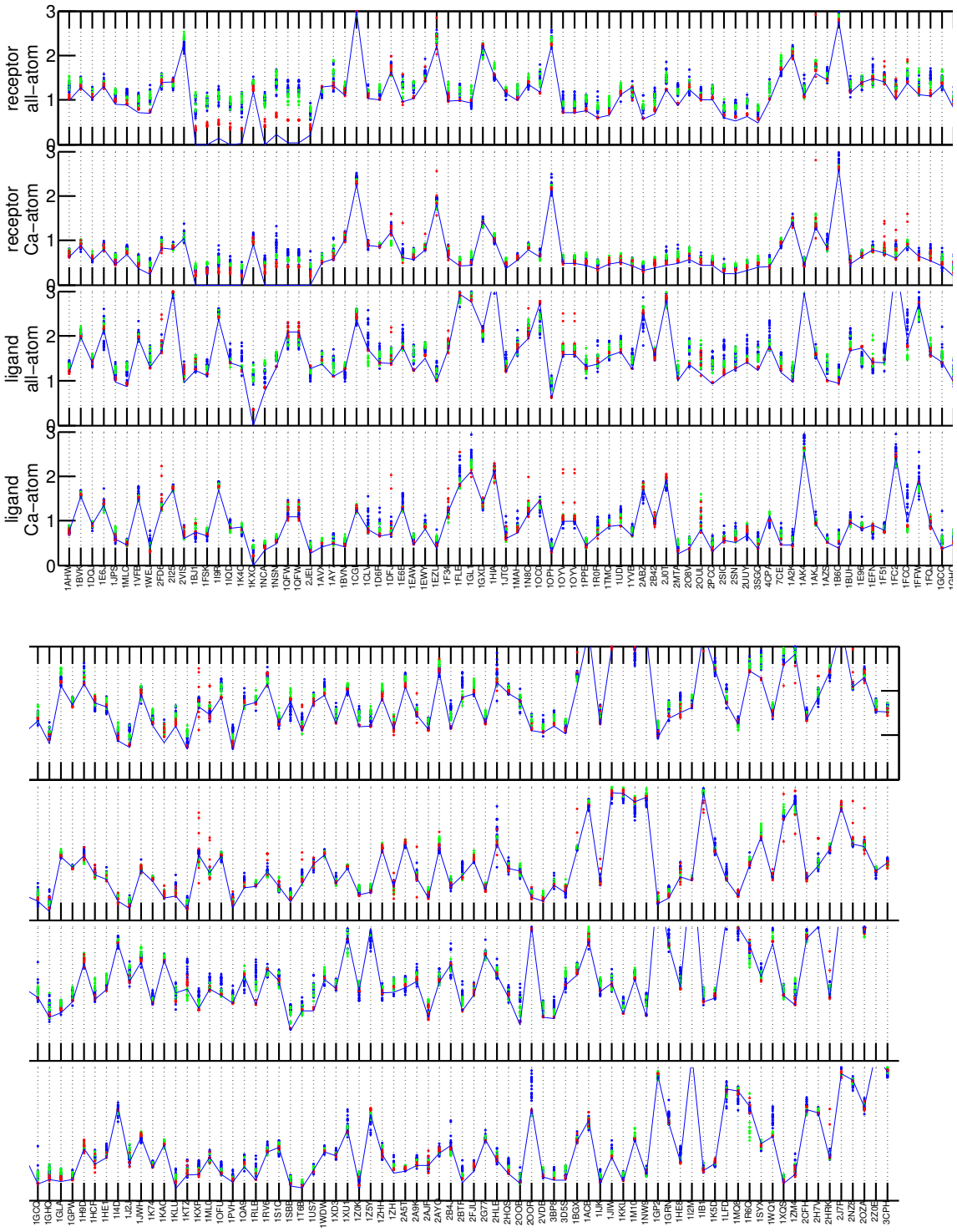

Figure S20. Irmsd of the generated models from the respective bound partner. The black line is the Irmsd of the unbound structure from the respective bound partner. Data of models generated for ROSETTA backrub are shown in blue, for MD simulation are in green, and for normal mode deformation are in red.

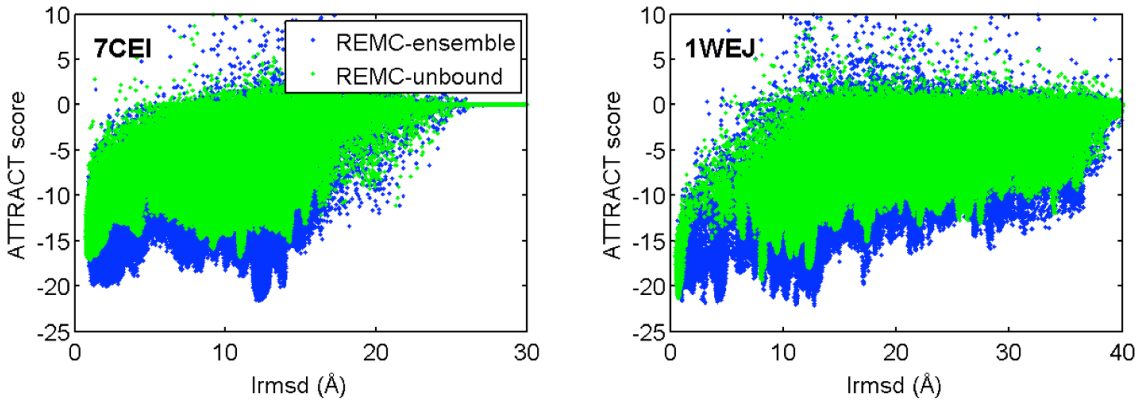

Figure S21. ATTRACT score versus interface RMSD (Irmsd) for REMC sampled results on targets 7CEI and 1WEJ.

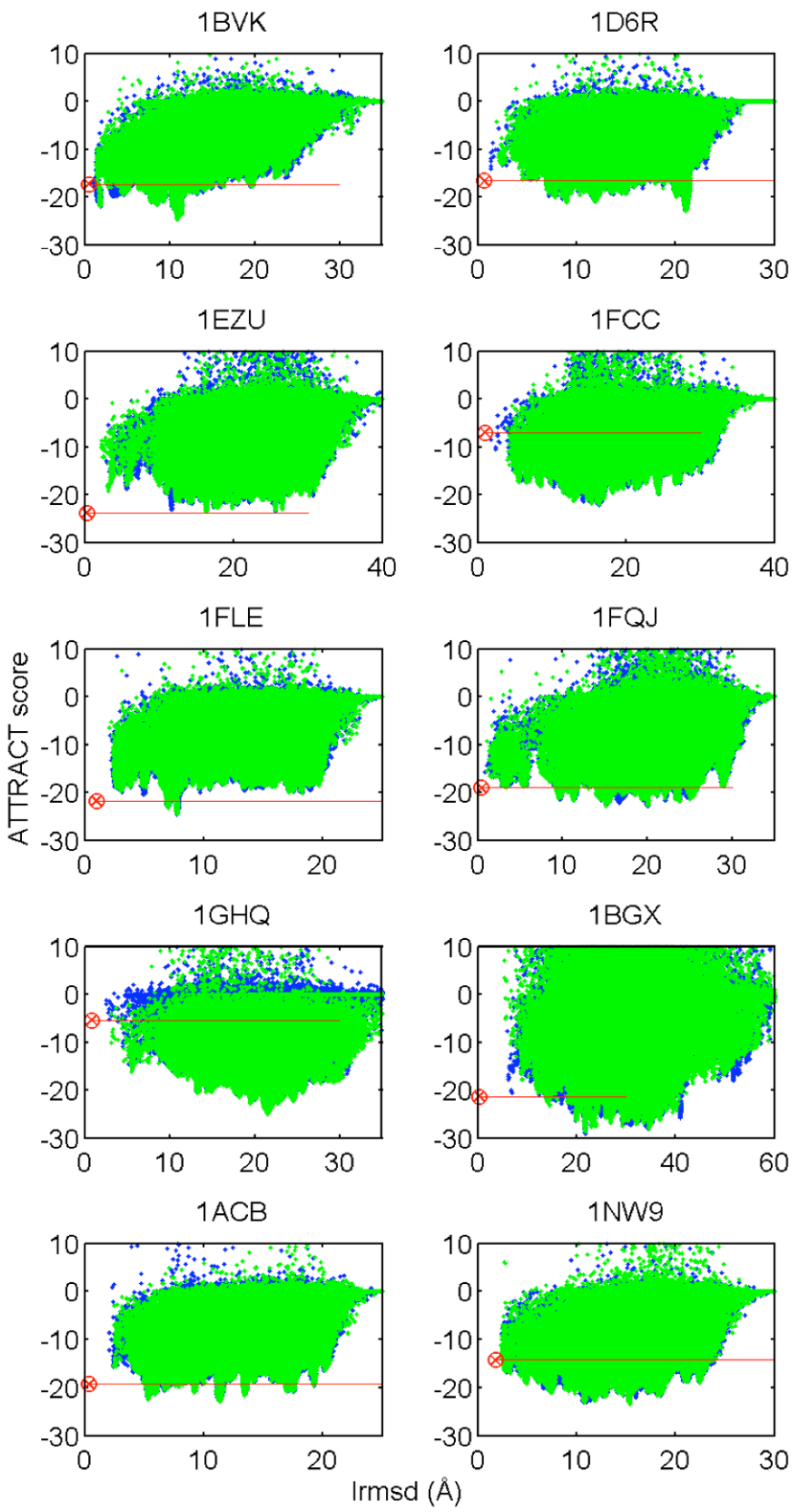

Figure S22. Full scatter plot of Irmsd against ATTRACT score on the 10 cases which has failed to identify the native bound position when native bound partner included in REMC-ensemble docking. Results from REMC-ensemble-40 and REMC-ensemble-39 are shown in blue and green. The Min-

native-bound are shown as red circle with cross. The red horizontal line indicates the energy of the Min-native-bound.

Supplementary Tables

1	2	3	4	5	6	7	8	9	10	11	12
1a2k	AB:C	1gy6	AB:AB	100	1.32	1.6	3ran	A:C	99.54	1.32	2.15
1a2y	AB:C	1vfa	AB:AB	100	0.68	1.8	3lzt	A:C	99.22	1.3	0.93
1acb	E:I	1gct	A:E	98.37	0.33	1.6	1egl	A:I	97.78	1.31	2
1bvn	P:T	1hx0	A:P	98.19	0.64	1.38	1hoe	A:T	100	0.45	2
1cse	E:I	1scn	E:E	100	0.3	1.9	1egl	A:I	100	1.21	2
1e96	A:B	1mh1	A:A	98.91	0.72	1.38	1hh8	A:B	99.51	0.63	1.8
1f7z	A:I	1dpo	A:A	99.55	0.33	1.59	5pti	A:I	100	0.32	1
1fm9	D:A	1zgy	A:D	100	1.13	1.8	2p1t	A:A	100	0.74	1.8
1fqj	A:B	1tnd	C:A	91.4	0.54	2.02	1fqi	A:B	100	1.45	1.94
1jps	LH:T	1jpt	LH:LH	100	1.02	1.85	1boy	A:T	100	1.33	2.2
1mlc	AB:E	1mlb	AB:AB	100	0.93	2.1	3lzt	A:E	100	0.75	0.93
1nby	AB:C	1dqq	CD:AB	100	0.73	1.8	3lzt	A:C	99.22	1.44	0.93
1oph	A:B	1qlp	A:A	97.72	1.19	2	2ayw	A:B	99.55	0.33	0.97
1ppe	E:I	1btp	A:E	97.4	0.43	2.2	1lu0	A:I	96.6	0.44	1.03
1ppf	E:I	2rg3	A:E	100	0.39	1.8	2gkr	I:I	100	0.5	1.16
1r0r	E:I	1scn	E:E	100	0.32	1.9	2gkr	I:I	100	0.62	1.16
1t6b	X:Y	1acc	A:X	99.86	1.43	2.1	1shu	X:Y	100	0.59	1.5
1tmq	A:B	1jae	A:A	100	0.39	1.65	1b1u	A:B	99.15	0.98	2.2
1tx4	A:B	1rgp	A:A	99.49	0.69	2	1kmq	A:B	99.43	0.43	1.55
1v7p	AB:C	1ukm	AB:AB	100	0.64	1.9	1dzi	A:C	99.46	0.67	2.1
1wq1	G:R	1wer	A:G	100	0.97	1.6	2ce2	X:R	98.8	1.23	1
1z5y	D:E	1l6p	A:D	99.2	1.15	1.65	2b1k	A:E	99.3	1.06	1.9
2a42	A:B	2fxu	A:A	100	0.82	1.35	2dnj	A:B	100	0.4	2
2a5t	A:B	1pb7	A:A	100	0.59	1.35	2a5s	A:B	100	1.06	1.7
2b42	A:B	1t6e	X:A	99.48	0.56	1.7	1xnb	A:B	100	1.02	1.49
2bnq	DE:AB	2bnu	AB:DE	99.76	1.13	1.4	1i4f	AB:AB	100	1.14	1.4
2hle	A:B	2bba	A:A	100	1.48	1.65	1iko	P:B	100	0.86	1.92
2mta	HL:A	2bbk	JM:HL	97.5	0.37	1.75	2rac	A:A	100	0.62	1.3
2oob	B:A	1yj1	A:B	100	0.87	1.3	2ooa	A:A	100	0.6	1.56
2pav	A:P	2fxu	A:A	100	1.07	1.35	1fil	A:P	100	0.81	2

Table S1: Targets in the benchmark set. Columns 1-2 are co-crystallized structure. Columns 3-7 are unbound structure of the first binding partner, and columns 8-12 are unbound structure of the second binding partner.
Column1: pdb code of co-crystallized structure.
Column 2: interacting chains of co-crystallized structure, separated by colon.
Column 3: pdb code of unbound structure of the first binding partner.
Column 4: chains before colon are in unbound structure, chains after colon are in co-crystallized structure. Column 5: sequence indentity between unbound and co-crystallized structure of the first binding partner.
Column 6: C_α-rmsd of unbound and co-crystallized structure of the first binding partner.
Column 7: crystal structure resolution.
Column 8: pdb code of unbound structure of second binding partner.
Column 9: chains before colon are in unbound structure; chains after colon are in co-crystallized structure.
Column 10: sequence indentity between unbound and co-crystallized structure of second binding partner.
Column 11: C_α-rmsd of the unbound and co-crystallized structure of second binding partner.
Column 12: crystal structure resolution.

	shotgun refined					ZDOCK refined					ReplicaDock refined				
Target	CQ[1]	L_rms	I_rms	f_{nat}	f_{nonnat}	CQ[1]	L_rms	I_rms	f_{nat}	f_{nonnat}	CQ[1]	L_rms	I_rms	f_{nat}	f_{nonnat}
1a2k	0	34.0	14.2	0.021	0.77	0	20.3	11.1	0.000	1.00	0	30.0	12.3	0.146	1.72
1a2y	0	13.9	8.0	0.023	1.02	0	22.9	6.9	0.068	1.45	0	18.1	10.7	0.068	1.04
1acb	0	17.5	7.6	0.014	0.66	*	9.81	3.8	0.145	0.89	**	4.87	2.70	0.333	0.58
1bvn	0	17.8	7.6	0.068	0.75	0	14.7	6.1	0.110	1.01	*	4.72	2.48	0.274	0.78
1cse	0	17.6	7.9	0.013	1.22	0	18.9	7.8	0.000	0.82	*	12.4	3.90	0.241	0.54
1e96	0	31.8	17.7	0.000	1.17	0	27.52	14.5	0.000	1.95	0	27.7	7.01	0.000	1.95
1f7z	0	15.6	5.3	0.098	0.62	0	14.4	3.73	0.066	1.06	0	17.7	4.16	0.197	0.83
1fm9	0	47.9	19.0	0.000	0.68	0	55.1	20.0	0.000	0.72	0	38.0	18.4	0.000	1.13
1fqj	0	31.7	15.7	0.000	1.21	0	31.5	14.6	0.000	1.15	0	27.6	11.8	0.000	2.25
1jps	0	37.2	10.3	0.000	0.97	0	33.5	15.8	0.000	0.98	0	45.3	16.8	0.000	1.15
1mlc	0	22.2	8.8	0.036	0.73	0	26.8	6.88	0.054	0.96	0	56.4	15.5	0.000	2.19
1nby	0	25.5	14.2	0.000	0.81	0	19.9	10.6	0.041	0.75	0	48.0	18.6	0.000	1.31
1oph	0	37.9	12.8	0.000	1.01	0	25.3	5.16	0.032	0.98	0	19.4	4.60	0.145	0.93
1ppe	***	2.65	0.95	0.746	0.08	***	1.76	0.79	0.761	0.19	***	2.11	0.88	0.775	0.11
1ppf	***	2.69	0.94	0.824	0.19	**	3.57	1.12	0.824	0.27	***	2.89	0.99	0.745	0.25
1r0r	0	17.5	6.3	0.000	0.80	*	10.3	2.76	0.394	0.42	0	14.6	4.81	0.167	0.66
1t6b	0	18.2	8.7	0.046	0.96	0	26.3	14.1	0.077	0.61	0	42.4	14.5	0.000	1.16
1tmq	0	19.8	11.8	0.026	0.77	**	4.94	1.43	0.579	0.44	**	4.45	1.36	0.618	0.51
1tx4	0	30.8	13.7	0.000	0.73	0	28.3	12.4	0.031	0.95	0	22.1	11.1	0.062	1.20
1v7p	0	23.2	11.3	0.048	1.04	0	23.9	8.62	0.145	1.08	0	23.5	11.8	0.048	1.29
1wq1	0	22.4	9.6	0.011	0.69	0	16.2	7.67	0.165	0.62	0	11.5	6.87	0.011	1.01
1z5y	*	4.12	1.92	0.273	0.42	*	6.18	3.22	0.333	0.50	0	10.0	4.83	0.288	0.51
2a42	0	37.6	10.9	0.021	1.18	0	43.3	11.6	0.042	1.16	0	43.4	14.9	0.000	1.95
2a5t	0	31.0	16.0	0.034	1.27	0	16.3	8.48	0.017	1.28	0	21.3	10.2	0.000	1.28
2b42	0	29.0	12.7	0.000	0.61	*	9.28	4.15	0.157	0.74	0	17.6	7.31	0.079	0.73
2bnq	0	36.7	14.9	0.000	1.31	0	65.5	24.1	0.000	1.75	0	62.5	24.2	0.000	2.90
2hle	0	24.9	10.7	0.000	0.77	*	6.63	2.91	0.464	0.40	0	36.6	10.28	0.083	0.61
2mta	*	7.25	3.34	0.457	0.63	**	10.1	2.52	0.543	0.56	0	21.0	8.97	0.000	1.17
2oob	0	10.6	5.7	0.111	0.66	0	20.3	5.27	0.074	0.77	0	15.3	9.22	0.000	0.88
2pav	0	18.6	10.4	0.000	0.87	**	3.35	1.45	0.509	0.33	0	46.9	11.9	0.000	1.40
summary	2*/2***					5*/4**/1***					2*/2**/2***				

Table S2: summary of structure prediction accuracy in unbound docking. Clusters are ranked by the median of its top 10 interface energies and represented by the lowest interface energy decoy. In column 'CQ' (CAPRI Quality), '0' indicates that none of the top 10 models was of accetable quality, '*', '**' and '***' indicates that at least one of the top 10 models is of acceptable, medium or high quality, respectively (Section 2.3.7). Columns 'L_rms', 'I_rms', 'f_{nat}' and '$f_{non\text{-}nat}$' record the respective information of the best model within these top 10 models. [1] CQ refers to CAPRI quality

Target	Exchange rate
1a2k	0.2404
1a2y	0.2394
1acb	0.2504
1bvn	0.2396
1cse	0.2423
1e96	0.2432
1f7z	0.2467
1fm9	0.2578
1fqj	0.2542
1jps	0.2448
1mlc	0.2471
1nby	0.2553
1oph	0.2564
1ppe	0.2687
1ppf	0.2737
1r0r	0.2586
1t6b	0.2566
1tmq	0.2521
1tx4	0.2474
1v7p	0.2582
1wq1	0.2427
1z5y	0.2513
2a42	0.2553
2a5t	0.2515
2b42	0.2494
2bnq	0.2678
2hle	0.2394
2mta	0.2385
2oob	0.3399
2pav	0.2498

Table S3: Average exchange rate between temperature levels across the tested benchmark.

Supplementary Methods

Method S1: protocol_capture/rosetta_dock/centroid

```
    # RosettaDock' shotgun approach sampling in low-resolution stage

       # COMMANDLINE:
$ROSETTA3_BIN/docking_protocol.mpi.linuxgccrelease \
-out:file:silent $DECOYS.out
-database $ROSETTA3_DB
-docking:randomize1 -docking:randomize2
-low_res_protocol_only
-nstruct $NSTRUCT
-in:file:native $PROTAB.pdb
-score:weights interchain_cen
-in:file:s $P.pdb
-partners $CHAIN1_$CHAIN2
```

About 120000 decoys for each target are generated in the low-resolution stage with RosettaDock's shotgun approach. For analysis and refinement we take the lowest 36% of decoys by energy and filter out *interchain_contact* > 10. An example decoys file is given in /protocol_capture/2012/replica_docking/example_runs/rosetta_dock/udock_1bvn/run/decoys.out

With this decoy selection would look like this:

```
   # get the tags of the selected decoys. 50 decoys in all in this example
file, so select 50*0.36=18 decoys for analysis and refine
for i in $(ls decoys_000?.out); do echo $i; cat $i | grep SCORE: >> decoys.fsc;
done

scripts/silent_data.py decoys.fsc score interchain_contact description | awk
'$2<=10{print}' | sort -n -k 1 | head -n 18 | awk '{print $3}' > tag_low


   # extract the selected decoys from decoys.out
for i in $(ls decoys_000?.out); do echo $i; scripts/extract_tagged_decoys.py $i
tag_low > low_$i; done
```

Method S2: protocol_capture/replica_dock/centroid

ReplicaDock is run with RosettaScripts. The temperature levels used are 2.0, 3.0 and 5.0. For each target, 4 trajectories with 3 replicas each are run. The trajectory is 5000,000 Monte-Carlo steps and snapshots are stored every 1000 steps. That is for each target ReplicaDock generates 3*4*5e+6/1000=60000 decoys. Decoys with interchain_contact <=10 are selected from the lowest two temperatures (2.0 and 3.0), which makes about 36000 decoys for each targets.

To run ReplicaDock, MPI-mode is required. Please note that specific numbers of processors have to be used: calculate number of processes using the formula: nstruct * n_replica + 2. The extra 2 processes are dedicated to the job distributor and File IO. n_replica is the number of temperature levels (here 3), and nstruct can be any positive integer. ReplicaDock outputs the trajectory in the form of two silent-files: one containing decoy+score information (name: decoys_<input_pdb>_nnnn_traj.out), and the second file is a copy of just the score information (scores_<input_pdb>_nnnn_traj.out). Decoytags are of the form P_tttt_rrr_ssssssss where tttt informs about trajctory number, rrr about the replica, and ssssssss about the snapshot number within the trajectory. The temperature levels are switched between different replicas. The current temp_level or temperature of a replica at the moment a decoy was recorded is found in the score-columns *temp_level* and *temperature*.

At the end of a trajectory the final decoy is written to the file 'decoys.out'; this file is a relict of using the JD2-framework and can be ignored. Additionally, the file 'trial.stat' is produced which gives information about acceptance rates in each temperature level.

```
   #  ReplicaDock sampling in low-resolution docking

      # COMMANDLINE:
$ROSETTA3_BIN/rosetta_scripts.mpi.linuxgccrelease \
-database $ROSETTA3_DB
-parser:protocol dock_cen.xml
-n_replica 3
-run:intermediate_structures
-out:file:silent decoys.out              # store final decoys, which can be
                                         # ignored.
-out:file:scorefile scores.fsc
-nstruct 4                               # 4 trajectories running
-partners $CHAIN1_$CHAIN2
```

```
-in:file:native $PROTAB.pdb
-in:file:s $P.pdb
-score:weights interchain_cen
```

```
    # ReplicaDock SUPPORT_FILES:

       # RosettaScripts file dock_cen.xml
<dock_design>
  <SCOREFXNS>
    <score_dock_low weights="interchain_cen" />
  </SCOREFXNS>
  <FILTERS>
  </FILTERS>
  <MOVERS>

    # sampling in low-resolution stage, thus switch2centroid
    <SwitchResidueTypeSetMover name=switch2centroid set=centroid/>

    # unbiased rigid-body move. Constant parameter for step size, alternatively
use dock_cen_inter.xml, which interpolates the parameter based on temperature
level.
    <ThermodynamicRigidBodyPerturbNoCenter name=rb_mover rot_mag=1
trans_mag=0.5/>

    # setup jumps via fold_tree, store movable_jump into RigidBodyInfo
    <DockSetupMover name=setup_jump/>

    # very loose AtomPair constraint between closest-to-mass-center Ca-atoms
    <AddEncounterConstraintMover name=encounter_cst gap=8/>

    # randomly reorient the two docking partners
    <DockingInitialPerturbation name=init_pert randomize2=1 randomize1=1 />

    # sampling engine

    # acceptance rate recorder, write to file 'trial_stats' at the end
    <TrialCounterObserver name=count file="trial.stats"/>

    # temp_file contains the temperature configurations. exchange between
neighbor temperatures is attempted every 1000 steps. No specific reason for
using HamiltonianExchange mover instead of ParallelTempering, only because I
started with it and it works as well for temperature only exchange.
    <HamiltonianExchange name=h_exchange temp_file="hamiltonians_cen.txt"
temp_stride=1000/>

    # take snapshots every 1000 steps and write the snapshots into a trajectory
silent file
    <SilentTrajectoryRecorder name=traj score_stride=1 stride=1000
cumulate_replicas=1 />

    # normally use trials=5000,000. Empirically, it is enough for converge.
    <MetropolisHastings name=sampler trials=5000000 scorefxn=score_dock_low >
      <Add mover_name=h_exchange/>
      <Add mover_name=traj/>
      <Add mover_name=count/>
      <Add mover_name=rb_mover/>
    </MetropolisHastings>

  </MOVERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
    <Add mover_name=switch2centroid/>
    <Add mover_name=setup_jump/>
    <Add mover_name=encounter_cst/>
    <Add mover_name=init_pert/>
    <Add mover_name=sampler/>
  </PROTOCOLS>
</dock_design>
```

```
    # temperatures configuration file "hamiltonians_cen.txt"
GRID_DIM 1
GLOBAL_PATCH atom_pair_constraint = 5    # set atom_pair_constraint weight to 5
                                         # for all replicas
1 2.0 interchain_cen                     # define temp and score of replicas
2 3.0 interchain_cen
3 5.0 interchain_cen
```

For analysis and refinement, decoys are selected as follows:

```
    # get the tags of the selected snapshots
cat scores_P_000*fsc > scores_traj.fsc

scripts/silent_data.py scores_traj.fsc temperature interchain_contact
description | awk '$1<4&&$2<=10{print $3}' > tag_low

    # extract selected decoys from the trajectory silent file
for i in $(ls decoys_P_000?_traj.out); do echo $i;
scripts/extract_tagged_decoys.py $i tag_low > low_$i; done
```

Method S3: protocol_capture/rosetta_dock/refine

Refinement of low-resolution ensemble is carried out with standard parameters. If disulfide bonds are present in the unbound structures, extra flags are added:

```
    -detect_disulf  true -rebuild_disulf  true  -fix_disulf  $DISULF_FILE
```

```
    # COMMANDLINE:
$ROSETTA3_BIN/docking_protocol.mpi.linuxgccrelease \
-database $ROSETTA3_DB
-evaluation:rmsd IRMS _input FULL
-in:file:native $PROTAB.pdb
-docking_local_refine
-ex1
-ex2aro
-nstruct 1
-score:weights docking
-use_input_sc
-unboundrot $PROTAB.pdb

-detect_disulf true                      # these three disulf related flags used
-rebuild_disulf true                     # only if disulfide bonds are present
-fix_disulf $DISULF_FILE                 # in the unbound structures

-in:file:silent $SELECTED_LOW_RES.out
-out:file:silent $REFINED.out
```

Method S4: protocol_capture/replica_dock/refine

Refinement of low-resolution ensembles from ReplicaDock as well as ReplicaDock-LoT are done using exactly the same protocol as refinement of low-resolution ensemble from shotgun approach.

Method S5: protocol_capture/relax_native

Using the same refinement protocol as for refinement of low-resolution ensembles, we generated the RelaxedNative ensembles starting from the superimposed reference structure with 1000 decoys for each target.

```
    # COMMANDLINE:
```

```
$ROSETTA3_BIN/docking_protocol.mpi.linuxgccrelease \
-database $ROSETTA3_DB
-evaluation:rmsd IRMS _input FULL
-in:file:native $PROTAB.pdb
-docking_local_refine
-ex1
-ex2aro
-nstruct 1000                    # generate 1000 decoys as RelaxedNative for
                                 # each target
-score:weights docking
-use_input_sc
-unboundrot $PROTAB.pdb

-detect_disulf true
-rebuild_disulf true
-fix_disulf $DISULF_FILE

-in:file:s $PROTAB.pdb           # start from the superimposed structure
-out:file:silent $REFINED.out
```

Method S6: Flags in commandlines in Method S1-S5

Here we list out the explanations for the flags mentioned in Method S1-S5 as follows.

flags	interpretation
-out:file:silent $DECOYS.out	specify the filename of output silent file
-database $ROSETTA3_DB	specify the directory of your rosetta database
-docking:randomize1 -docking:randomize2	randomize1 and randomize2 using together to do global docking
-low_res_protocol_only	low-resolution stage only with RosettaDock --- shotgun sampling
-nstruct $NSTRUCT	specify how many decoys to generate
-score:weights interchain_cen	centroid docking energy function
-partners $PARTNER1_$PARTNER2	specify the docking partners, for example "AB_C" means docking chain C to chain AB, and keep chain AB rigid
-evaluation:rmsd IMRS_input FULL	evaluate the Ca-RMSD for the entire structure
-docking_local_refine	only do refinement with RosettaDock
-ex1 -ex2aro	adding extra sidechain rotamers
-use_input_sc	Use accepted rotamers from the input structure between Monte-Carlo with Minimization (MCM) cycle
-unboundrot $PROTAB.pdb	use unbound rotamers
-detect_disulf true	these three flags are added if disulfide bonds are

-rebuild_disulf true -fix_disulf $DISULF_FILE	present in the unbound structures. in $DISULF_FILE residue pairs forming disulfide bonds are specified, like "28 35"
$TARGETLIB	protocol_capture/2012/replica_docking/dock_targetlib
$PROTAB.pdb	PDB-file with unbound partners superimposed onto the bound complex, used as the reference structure for RMSD related calculation
$P.pdb	PDB-file with the two binding partners in $PROTAB.pdb fully randomized, used as the start conformation to avoid initial bias
-n_replica	number of replicas, need to be consistent with the number of temperature levels in file hamiltonians_cen*.txt

Method S7: Automated Setup

All the production runs of the benchmark in this work are generated with the automated setup tools available with the CS-Rosetta toolbox(www.csrosetta.org). The workflow of this automated setup is shown in Figure S12. Methods for docking used in this work are included in the folder under *protocol_capture/replica_docking/csrosetta3/flag_library/methods/*, i.e. *_docking_base, rosetta_dock* and *replica_dock*. To run this toolbox, a minmal installation is required as detailed below.

To install:

```
    # go to the directory
cd /csrosetta3/

    # check all the options available for the installation
install.py -h

    # use option -nopicking since fragment picking is not required for docking,
and provide path to ROSETTA as follows
install.py -nopicking -rosetta ~/rosetta -rosetta_database
~/rosetta/rosetta_database

    # You will see some installation information. Finally you should see
'created symlink /home/zhezhang/csrosetta3/com/init.bashrc -->
/home/zhezhang/csrosetta3/com/init' with /home/zhezhang/ replaced by your own
user directory. To use the toolbox in your shell run
source /home/zhezhang/csrosetta3/com/init
```

setup targets library

This step assembles target related input files to build the target library (Figure S12). By default the library is stored in ~/cs_targetlib. You can also specify a directory using flag '-target_prefix' as follows. Absolute path is recommended for '-target_prefix'.

```
    # Setup a target for rosetta_dock
setup_target -method rosetta_dock -target udock_1bvn -target_prefix
$TARGETLIB_DIR -disulf disulf_file -native protAB.pdb -pdb P.pdb -partners
partners

    # Setup a target for replica_dock either explicitly using
setup_target -method replica_dock -target udock_1bvn -target_prefix
$TARGETLIB_DIR -native protAB.pdb -pdb P.pdb -partners partners

    # or copying the inputs from a previously prepared 'rosetta_dock' setup as
follows:
setup_target -method replica_dock -target udock_1bvn -target_prefix
$TARGETLIB_DIR -transfer_method rosetta_dock
```

setup runs

This step creates a run-ready directory as specified with flag '-dir', in which job-scripts, input files, RosettaScripts-xml as well as flag files are contained. For flag '-dir', absolute path is recommended. For job scripts, you can use different types (e.g. moab) according to your queuing system, as shown in Figure S12. I will keep slurm (-job slurm) here as example.

For single-machine/interactive use, you can simply specify with flag '*-job interactive'* when setup the run, then start the running under the corresponding run/ directory using '*source production.interactive.job -n $Np'* with *$Np* specifying the processor numbers.

```
    # setup a run of ReplicaDock in queuing system
setup_run -method replica_dock -target udock_1bvn -target_prefix $TARGETLIB_DIR
-dir $TEST_REPLICA_DIR -job slurm -extras mpi -score interchain_cen -nstruct 1
-protocol rep_cen -xml uniform -n_replica 3

    # rosetta_dock's shotgun sampling in low-resolution stage
setup_run -method rosetta_dock -target udock_1bvn -target_prefix $TARGETLIB_DIR
-dir $TEST_ROSETTA_DIR -job slurm -extras mpi -protocol centroid -batches 2
-score interchain_cen -nstruct 100

    # refine decoys sets
setup_run -method rosetta_dock -target udock_1bvn -target_prefix $TARGETLIB_DIR
-dir $TEST_REFINE_DIR -job slurm -extras mpi -protocol refine -pattern
"low_decoys_*out" -prefix refine -score docking -nstruct 1 -start
$ABSOLUTE_PATH_OF_FOLDER_FOR_DECOYS

    # refine protAB.pdb to generate relaxNative ensemble
setup_run -method rosetta_dock -target udock_1bvn -target_prefix $TARGETLIB_DIR
-dir $TEST_RELAX_NATIVE_DIR -job slurm -extras mpi -protocol refine -out
relax_native.out -score docking -nstruct 1000
```

change/add file in /jobtemplates

You can easily modify the jobtemplates for use with your own queuing system. Shown here is a job template for slurm. The $CM_ variables are replaced by the automatic setup

tool when run setup_run. #SBATCH and $SLURM_ are specific to SLURM queuing environment.

```
#!/bin/bash -x
#SBATCH -J csrosetta
### start of jobscript

NSLOTS=$SLURM_NTASKS

module load openmpi/gcc

LOGS=logs_`echo $SLURM_JOB_ID | awk -v FS="." '{print $1}'`
mkdir -p $LOGS

## have NSLOTS - 3 worker processes -- determines number of structures per
generation...
NSTRUCT=`echo $NSLOTS | awk '{print $1-3}'`

echo "running on $NSLOTS cpus ..."
EXE=$CM_EXECUTEABLE.$CM_EXEC_EXT
CMDLINE="-out:level 300 -mute all_high_mpi_rank_filebuf -out:mpi_tracer_to_file
$LOGS/log -database $CM_ROSETTA_DATABASE $CM_COMMANDLINE"
CYCLES=$CM_AUTO_NSTRUCT
$MPI_RUN -n $NSLOTS $EXE $CMDLINE $CYCLES
```

Bibliography

1. Stein A, Mosca R, Aloy P. Three-dimensional modeling of protein interactions and complexes is going “omics.” Current Opinion in Structural Biology. Elsevier Ltd; 2011;21: 200–208.

2. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440: 637–643. doi:10.1038/nature04670

3. Berman HM, Kleywegt GJ, Nakamura H, Markley JL. The Protein Data Bank archive as an open data resource. J Comput Aided Mol Des. 2014;28: 1009–1014. doi:10.1007/s10822-014-9770-y

4. Janin J. Assessing predictions of protein–protein interaction: the CAPRI experiment. Protein Sci. Wiley Online Library; 2005;14: 278–283.

5. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KMS, Morozov AV, et al. Protein-protein docking predictions for the CAPRI experiment. Proteins: Structure, Function, and Bioinformatics. 52nd ed. 2003;52: 118–122. doi:10.1002/prot.10384

6. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science. 2000;289: 905–920.

7. Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, et al. Structure of the 30S ribosomal subunit. Nature. 2000;407: 327–339. doi:10.1038/35030006

8. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, et al. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell. 2000;102: 615–623.

9. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature. 2004;430: 223–226. doi:10.1038/nature02679

10. Liu Q, Greimann JC, Lima CD. Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell. 2006;127: 1223–1237. doi:10.1016/j.cell.2006.10.037

11. Egea PF, Shan S-O, Napetschnig J, Savage DF, Walter P, Stroud RM. Substrate twinning activates the signal recognition particle and its receptor. Nature. 2004;427: 215–221. doi:10.1038/nature02250

12. Kay LE. New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. Journal of Molecular Biology. 2016;428: 323–331. doi:10.1016/j.jmb.2015.11.028

13. Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. Nature Publishing Group. 2013;10: 47–53. doi:10.1038/nmeth.2289

14. Wodak SJ, Janin J. Computer analysis of protein-protein interaction. Journal of Molecular Biology. 1978.

15. Liang S, Liu S, Zhang C, Zhou Y. A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. Proteins: Structure, Function, and Bioinformatics. 2007;69: 244–253. doi:10.1002/prot.21498

16. Ritchie DW. Recent progress and future directions in protein-protein docking. Current Protein and Peptide Science. Bentham Science Publishers; 2008;9: 1–15.

17. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. Principles of flexible protein-protein docking. Proteins: Structure, Function, and Bioinformatics. 2008;73: 271–289. doi:10.1002/prot.22170

18. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. II. European Journal of Inorganic Chemistry. WILEY-VCH Verlag; 1894;27: 3479–3483. doi:10.1002/cber.189402703169

19. Huang S-Y. Exploring the potential of global protein-protein docking: an overview and critical assessment of current programs for automatic ab initio docking. Drug Discov Today. 2015;20: 969–977. doi:10.1016/j.drudis.2015.03.007

20. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. Journal of Molecular Biology. 1997;272: 106–120. doi:10.1006/jmbi.1997.1203

21. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, E N, et al. Protein docking using continuum electrostatics and geometric fit. Protein Engineering. 2001;14: 105–113.

22. Roberts VA, Thompson EE, Pique ME, Perez MS, Eyck Ten LF. DOT2: Macromolecular docking with improved biophysical models. J Comput Chem. 2013;34: 1743–1758. doi:10.1002/jcc.23304

23. Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein-docking algorithm. Proteins: Structure, Function, and Bioinformatics. Wiley Online Library;

2003;52: 80–87.

24. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA. 1992;89: 2195–2199.

25. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein-protein docking. Protein Sci. 2002;11: 571–587.

26. Zhang C, Lai L. SDOCK: A global protein-protein docking program using stepwise force-field potentials - Zhang - 2011 - Journal of Computational Chemistry - Wiley Online Library. J Comput Chem. 2011.

27. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics. 2006;65: 392–406. doi:10.1002/prot.21117

28. Chowdhury R, Rasheed M, Keidel D, Moussalem M, Olson A, Sanner M, et al. Protein-Protein Docking with F(2)Dock 2.0 and GB-Rerank. PLoS ONE. 2013;8: e51307. doi:10.1371/journal.pone.0051307

29. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Research. 2006;34: W310–W314. doi:10.1093/nar/gkl206

30. Eisenstein M, Katchalski-Katzir E. On proteins, grids, correlations, and docking. Comptes Rendus Biologies. 2004.

31. Ritchie DW, Kemp GJL. Protein docking using spherical polar Fourier correlations. Proteins: Structure, Function, and Bioinformatics. Wiley Online Library; 2000;39: 178–194.

32. Ritchie DW, Kozakov D, Vajda S. Accelerating and focusing protein-protein docking correlations using multi-dimensional rotational FFT generating functions. Bioinformatics. 2008;24: 1865–1873. doi:10.1093/bioinformatics/btn334

33. Garzon JI, Lopez-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernández Recio J, et al. FRODOCK: a new approach for fast rotational protein-protein docking. Bioinformatics. 2009;25: 2544–2551. doi:10.1093/bioinformatics/btp447

34. Padhorny D, Kazennov A, Zerbe BS, Porter KA, Xia B, Mottarella SE, et al. Protein–protein docking by fast generalized Fourier transforms on 5D rotational manifolds. Proceedings of the …. 2016.

35. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and

SymmDock: servers for rigid and symmetric docking. Nucleic Acids Research. 2005;33: W363–W367. doi:10.1093/nar/gki481

36. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, et al. Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. Journal of Molecular Biology. 2003;331: 281–299.

37. Zacharias M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci. 2003;12: 1271–1282.

38. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003;125: 1731–1737.

39. Fernández Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins: Structure, Function, and Bioinformatics. Wiley Online Library; 2003;52: 113–117.

40. Gardiner EJ, Willett P, Artymiuk PJ. GAPDOCK: a Genetic Algorithm Approach to Protein Docking in CAPRI round 1. Proteins: Structure, Function, and Bioinformatics. 2003;52: 10–14. doi:10.1002/prot.10386

41. Zhang Z, Lange OF. Replica exchange improves sampling in low-resolution docking stage of RosettaDock. PLoS ONE. 2013;8: e72096. doi:10.1371/journal.pone.0072096

42. Solernou A, Fernandez-Recio J. pyDockCG: new coarse-grained potential for protein-protein docking. J Phys Chem B. 2011;115: 6032–6039. doi:10.1021/jp112292b

43. Pons C, Talavera D, la Cruz de X, Orozco M, Fernandez-Recio J. Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking. J Chem Inf Model. 2011;51: 370–377. doi:10.1021/ci100353e

44. Ravikumar KM, Huang W, Yang S. Coarse-grained simulations of protein-protein association: an energy landscape perspective. Biophys J. 2012;103: 837–845. doi:10.1016/j.bpj.2012.07.013

45. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. Chem Rev. 2016;116: 7898–7936. doi:10.1021/acs.chemrev.6b00163

46. Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci USA. 1958;44: 98–104.

47. Gao Y, Douguet D, Tovchigrechko A, Vakser IA. DOCKGROUND system of databases for protein recognition studies: unbound structures for docking. Proteins: Structure, Function, and Bioinformatics. 2007;69: 845–851. doi:10.1002/prot.21714

48. Schindler CEM, de Vries SJ, Zacharias M. iATTRACT: Simultaneous global and local interface optimization for protein-protein docking refinement. Proteins: Structure, Function, and Bioinformatics. 2014;83: 248–258. doi:10.1002/prot.24728

49. Fernández-Recio, Press PBD. Refinement of rigid-body protein–protein docking using backbone and side-chain minimization with a coarse-grained model. 2010;: 1–9.

50. Elcock AH, Sept D, McCammon JA. Computer Simulation of Protein−Protein Interactions. J Phys Chem B. American Chemical Society; 2001;105: 1504–1518. doi:10.1021/jp003602d

51. Zhang Z, Schindler CEM, Lange OF, Zacharias M. Application of Enhanced Sampling Monte Carlo Methods for High-Resolution Protein-Protein Docking in Rosetta. PLoS ONE. 2015;10: e0125941. doi:10.1371/journal.pone.0125941

52. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Engineering. 1999;12: 713–720.

53. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R. Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci. 2000;9: 10–19. doi:10.1110/ps.9.1.10

54. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends in Biochemical Sciences. 2010;35: 539–546. doi:10.1016/j.tibs.2010.04.009

55. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nature Publishing Group. 2009;5: 789–796. doi:10.1038/nchembio.232

56. Bonvin AMJJ. Flexible protein-protein docking. Current Opinion in Structural Biology. 2006;16: 194–200. doi:10.1016/j.sbi.2006.02.002

57. Chaudhury S, Gray JJ. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. Journal of Molecular Biology. 2008;381: 1068–1087. doi:10.1016/j.jmb.2008.05.042

58. Ellingson SR, Miao Y, Baudry J, Smith JC. Multi-conformer ensemble docking

to difficult protein targets. J Phys Chem B. 2015;119: 1026–1034. doi:10.1021/jp506511p

59. Korb O, Olsson TSG, Bowden SJ, Hall RJ, Verdonk ML, Liebeschuetz JW, et al. Potential and Limitations of Ensemble Docking. J Chem Inf Model. 2012;52: 1262–1274. doi:10.1021/ci2005934

60. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Current Opinion in Structural Biology. 2008;18: 178–184. doi:10.1016/j.sbi.2008.01.004

61. Craig IR, Essex JW, Spiegel K. Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. J Chem Inf Model. 2010;50: 511–524. doi:10.1021/ci900407c

62. Novoa EM, Pouplana LR de, Barril X, Orozco M. Ensemble Docking from Homology Models. J Chem Theory Comput. 2010;6: 2547–2557. doi:10.1021/ct100246y

63. Huang S-Y, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins: Structure, Function, and Bioinformatics. 2007;66: 399–421. doi:10.1002/prot.21214

64. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins: Structure, Function, and Bioinformatics. 2002;47: 409–443.

65. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins: Structure, Function, and Bioinformatics. 2010;78: 3073–3084. doi:10.1002/prot.22818

66. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins: Structure, Function, and Bioinformatics. 2013;81: 2082–2095. doi:10.1002/prot.24428

67. Vajda S, Hall DR, Kozakov D. Sampling and scoring: A marriage made in heaven. Proteins: Structure, Function, and Bioinformatics. 2013;81: 1874–1884. doi:10.1002/prot.24343

68. Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. Proteins: Structure, Function, and Bioinformatics. 1997;Suppl 1: 226–230.

69. Norel R, Lin SL, Wolfson HJ, Nussinov R. Shape complementarity at protein-protein interfaces. Biopolymers. 1994;34: 933–940.

doi:10.1002/bip.360340711

70. Mainz A, Religa TL, Sprangers R, Linser R, Kay LE, Reif B. NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. Angew Chem Int Ed Engl. 2013;52: 8746–8751. doi:10.1002/anie.201301215

71. Fernández Recio J, Totrov M, Abagyan R. Identification of protein–protein interaction sites from docking energy landscapes. Journal of Molecular Biology. 2004.

72. Li L, Guo D, Huang Y, Liu S, Xiao Y. ASPDock: protein-protein docking algorithm using atomic solvation parameters model. BMC Bioinformatics. BioMed Central Ltd; 2011;12: 36. doi:10.1186/1471-2105-12-36

73. Cheng TM-K, Blundell TL, Fernandez-Recio J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. Proteins: Structure, Function, and Bioinformatics. 2007;68: 503–515. doi:10.1002/prot.21419

74. Moreira IS, Fernandes PA, Ramos MJ. Protein-protein docking dealing with the unknown. J Comput Chem. 2009;31: 317–342.

75. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins: Structure, Function, and Bioinformatics. 2007;69: 139–159. doi:10.1002/prot.21495

76. Lorenzen S, Zhang Y. Monte Carlo refinement of rigid-body protein docking structures with backbone displacement and side-chain optimization. Protein Sci. 2007;16: 2716–2725.

77. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. Uversky VN, editor. PLoS ONE. 2011;6: e22477.

78. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. Journal of Molecular Biology. 1994;235: 983–1002. doi:10.1006/jmbi.1994.1052

79. Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 1997.

80. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. J Chem Phys. 2002;116: 9058. doi:10.1063/1.1472510

81. Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble.

Physical Review Letters. 2010;104: 190601.

82. Garma L, Mukherjee S, Mitra P, Zhang Y. How Many Protein-Protein Interactions Types Exist in Nature? PLoS ONE. 2012;7: e38913.

83. Melquiond ASJ, Karaca E, Kastritis PL, Bonvin AMJJ. Next challenges in protein-protein docking: from proteome to interactome and beyond. WIREs Comput Mol Sci. 2011;2: 642–651.

84. Nooren IMA, Thornton JM. Diversity of protein–protein interactions. EMBO J. Nature Publishing Group; 2003;22: 3486–3492.

85. Stratmann D, Boelens R, Bonvin AMJJ. Quantitative use of chemical shifts for the modeling of protein complexes. Proteins: Structure, Function, and Bioinformatics. 2011;79: 2662–2670.

86. Cowieson NP, Kobe B, Martin JL. United we stand: combining structural methods. Current Opinion in Structural Biology. 2008;18: 617–622.

87. Karaca E, Bonvin AMJJ. Advances in integrative modeling of biomolecular complexes. Methods. 2013;59: 372–381. doi:10.1016/j.ymeth.2012.12.004

88. Koehler J, Meiler J. Expanding the utility of NMR restraints with paramagnetic compounds: background and practical aspects. Progress in Nuclear Magnetic Resonance Spectroscopy. 2011;59: 360–389.

89. Schneider S, Zacharias M. Scoring optimisation of unbound protein-protein docking including protein binding site predictions. J Mol Recognit. 2011;25: 15–23.

90. Smith GR, Sternberg MJE. Prediction of protein–protein interactions by docking methods. Current Opinion in Structural Biology. Elsevier; 2002;12: 28–35.

91. Vajda S, Kozakov D. Convergence and combination of methods in protein–protein docking. Current Opinion in Structural Biology. Elsevier; 2009;19: 164–170.

92. Sternberg MJ, Gabb HA, Jackson RM. Predictive docking of protein-protein and protein-DNA complexes. Current Opinion in Structural Biology. 1998;8: 250–256.

93. Fleishman SJ, Baker D. Role of the Biomolecular Energy Gap in Protein Design, Structure, and Evolution. Cell. Elsevier Inc; 2012;149: 262–273.

94. Aloy P, Querol E, Aviles FX, Sternberg MJE. Automated structure-based prediction of functional sites in proteins: applications to assessing the

validity of inheriting protein function from homology in genome annotation and to protein docking. Journal of Molecular Biology. 2001;311: 395–408.

95. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Geometry-based flexible and symmetric protein docking. Janin J, editor. Proteins: Structure, Function, and Bioinformatics. 2005;60: 224–231.

96. Swendsen RH, Wang JS. Replica Monte Carlo simulation of spin-glasses. Physical Review Letters. APS; 1986;57: 2607–2609.

97. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. Elsevier; 1999;314: 141–151.

98. Kim YC, Tang C, Clore GM, Hummer G. Replica exchange simulations of transient encounter complexes in protein–protein association. Proc Natl Acad Sci USA. National Acad Sciences; 2008;105: 12855–12860.

99. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Phys. 1953;21: 1087–1092.

100. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. Biometrika Trust; 1970;57: 97–109.

101. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. Uversky VN, editor. PLoS ONE. 2011;6: e20161.

102. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins: Structure, Function, and Bioinformatics. Wiley Online Library; 2003;52: 51–67.

103. Patriksson A, van der Spoel D. A temperature predictor for parallel tempering simulations. Physical Chemistry Chemical Physics. Royal Society of Chemistry; 2008;10: 2073–2077.

104. Rathore N, Chopra M, de Pablo JJ. Optimal allocation of replicas in parallel tempering simulations. J Chem Phys. 2005;122: 024111. doi:10.1063/1.1831273

105. Sanbonmatsu KY, Garc a AE. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. Proteins: Structure, Function, and Bioinformatics. 2002;46: 225–234.

106. Hwang H, Vreven T, Pierce BG, Hung J-H, Weng Z. Performance of ZDOCK

and ZRANK in CAPRI rounds 13-19. Proteins: Structure, Function, and Bioinformatics. 2010;78: 3104–3110. doi:10.1002/prot.22764

107. Wiehe K, Pierce B, Tong WW, Hwang H, Mintseris J, Weng Z. The performance of ZDOCK and ZRANK in rounds 6-11 of CAPRI. Proteins: Structure, Function, and Bioinformatics. 2007;69: 719–725.

108. Wiehe K, Pierce B, Mintseris J, Tong WW, Anderson R, Chen R, et al. ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. Janin J, editor. Proteins: Structure, Function, and Bioinformatics. 2005;60: 207–213. doi:10.1002/prot.20559

109. Lorenzen S, Zhang Y. Identification of near-native structures by clustering protein docking conformations. Proteins: Structure, Function, and Bioinformatics. 2007;68: 187–194. doi:10.1002/prot.21442

110. Fraternali F, Cavallo L. Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. Nucleic Acids Research. Oxford Univ Press; 2002;30: 2950–2960.

111. Kleinjung J, Fraternali F. POPSCOMP: an automated interaction analysis of biomolecular complexes. Nucleic Acids Research. 2005;33: 342–346. doi:10.1093/nar/gki369

112. Tyka MD, Keedy DA, André I, DiMaio F, Song Y, Richardson DC, et al. Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. Journal of Molecular Biology. Elsevier Ltd; 2011;405: 607–618.

113. Tyka MD, Jung K, Baker D. Efficient sampling of protein conformational space using fast loop building and batch minimization on highly parallel computers. J Comput Chem. 2012;33: 2483–2491. doi:10.1002/jcc.23069

114. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, et al. Estimating the size of the human interactome. Proc Natl Acad Sci USA. National Acad Sciences; 2008;105: 6959–6964.

115. Baker D. Protein folding, structure prediction and design. Biochem Soc Trans. 2014;42: 225–229. doi:10.1038/nbt.2109

116. Parmeggiani F, Huang P-S, Vorobiev S, Xiao R, Park K, Caprari S, et al. A general computational approach for repeat protein design. Journal of Molecular Biology. 2015;427: 563–575. doi:10.1016/j.jmb.2014.11.005

117. London N, Ambroggio X. An accurate binding interaction model in de novo computational protein design of interactions: if you build it, they will bind. Journal of Structural Biology. 2014;185: 136–146. doi:10.1016/j.jsb.2013.03.012

118. Baker D. Prediction and design of macromolecular structures and interactions. Philos Trans R Soc Lond B Biol Sci. 2006;361: 459–463. doi:10.1098/rstb.2005.1803

119. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol. 2004;11: 371–379. doi:10.1038/nsmb749

120. Huang S-Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. Drug Discov Today. 2014;19: 1081–1096. doi:10.1016/j.drudis.2014.02.005

121. Gray JJ. High-resolution protein–protein docking. Current Opinion in Structural Biology. 2006;16: 183–193. doi:10.1016/j.sbi.2006.03.003

122. London N, Schueler-Furman O. Funnel hunting in a rough terrain: learning and discriminating native energy funnels. Structure/Folding and Design. 2008;16: 269–279. doi:10.1016/j.str.2007.11.013

123. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol. 1997;4: 10–19.

124. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein-protein association. Biophys J. 1999;76: 1166–1178. doi:10.1016/S0006-3495(99)77281-4

125. Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chemical Physics Letters. 1996;281: 140–150. doi:10.1016/S0009-2614(97)01198-6

126. Falcioni M, Deem MW. A biased Monte Carlo scheme for zeolite structure solution. J Chem Phys. 1999;cond-mat.stat-mech: 1754. doi:10.1063/1.477812

127. Hritz J, Oostenbrink C. Hamiltonian replica exchange molecular dynamics using soft-core interactions. J Chem Phys. 2008;128: 144121. doi:10.1063/1.2888998

128. Luitz MP, Zacharias M. Protein-ligand docking using hamiltonian replica exchange simulations with soft core potentials. J Chem Inf Model. 2014;54: 1669–1675. doi:10.1021/ci500296f

129. Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, Isern NG, et al. High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. Journal of Molecular Biology. 2007;366: 1209–1221. doi:10.1016/j.jmb.2006.11.080

130. Yanover C, Schueler-Furman O, Weiss Y. Minimizing and learning energy functions for side-chain prediction. J Comput Biol. 2008;15: 899–911. doi:10.1089/cmb.2007.0158

131. Yan Q, de Pablo JJ. Hyper-parallel tempering Monte Carlo: Application to the Lennard-Jones fluid and the restricted primitive model. J Chem Phys. 1999;111: 9509. doi:10.1063/1.480282

132. Yan Q, de Pablo JJ. Hyperparallel tempering Monte Carlo simulation of polymeric systems. J Chem Phys. AIP Publishing; 2000;113: 1276–1282.

133. Ostermeir K, Zacharias M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. Biochim Biophys Acta. 2013;1834: 847–853. doi:10.1016/j.bbapap.2012.12.016

134. Earl DJ, Deem MW. Parallel tempering: Theory, applications, and new perspectives. Physical Chemistry Chemical Physics. 2005;7: 3910. doi:10.1039/b509983h

135. Barducci A, Bonomi M, Parrinello M. Metadynamics. WIREs Comput Mol Sci. 2011;1: 826–843. doi:10.1002/wcms.31

136. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys. 2008;71: 126601. doi:10.1088/0034-4885/71/12/126601

137. Deighan M, Bonomi M, Pfaendtner J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. J Chem Theory Comput. American Chemical Society; 2012;8: 2189–2192. doi:10.1021/ct300297t

138. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Meth Enzymol. Elsevier; 2004;383: 66–93.

139. Shapovalov MV, Dunbrack RL Jr. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. Structure/Folding and Design. Elsevier Ltd; 2011;19: 844–858. doi:10.1016/j.str.2011.03.019

140. Barducci A, Bussi G, Parrinello M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. Physical Review Letters. 2008;100: 020603. doi:10.1103/PhysRevLett.100.020603

141. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. Proteins: Structure, Function, and Bioinformatics. 2010;78: 3111–3114. doi:10.1002/prot.22830

142. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein-protein docking

benchmark version 3.0. Proteins: Structure, Function, and Bioinformatics. 2008;73: 705–709.

143. Smith GR, Sternberg MJE, Bates PA. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. Journal of Molecular Biology. 2005;347: 1077–1101. doi:10.1016/j.jmb.2005.01.058

144. Zacharias M. Accounting for conformational changes during protein-protein docking. Current Opinion in Structural Biology. Elsevier Ltd; 2010;20: 180–186. doi:10.1016/j.sbi.2010.02.001

145. Vakser IA. Protein-protein docking: from interaction to interactome. Biophys J. 2014;107: 1785–1793. doi:10.1016/j.bpj.2014.08.033

146. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Sci. 1999;8: 1181–1190. doi:10.1110/ps.8.6.1181

147. Vogt AD, Di Cera E. Conformational selection is a dominant mechanism of ligand binding. Biochemistry. 2013;52: 5723–5729. doi:10.1021/bi400929b

148. Grünberg R, Leckner J, Nilges M. Complementarity of structure ensembles in protein-protein binding. Structure/Folding and Design. 2004;12: 2125–2136. doi:10.1016/j.str.2004.09.014

149. Bastard K, Thureau A, Lavery R, Prévost C. Docking macromolecules with flexible segments. J Comput Chem. 2003;24: 1910–1920. doi:10.1002/jcc.10329

150. Jiang F, Kim SH. "Soft docking": matching of molecular surface cubes. Journal of Molecular Biology. 1991;219: 79–102.

151. Palma PN, Krippahl L, Wampler JE, Moura JJ. BiGGER: a new (soft) docking algorithm for predicting protein interactions. Proteins: Structure, Function, and Bioinformatics. 2000;39: 372–384. doi:10.1002/(SICI)1097-0134(20000601)39:43.0.CO;2-Q

152. May A, Zacharias M. Accounting for global protein deformability during protein–protein and protein–ligand docking. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics. 2005;1754: 225–231. doi:10.1016/j.bbapap.2005.07.045

153. May A, Zacharias M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. Proteins: Structure, Function, and Bioinformatics. 2008;70: 794–809. doi:10.1002/prot.21579

154. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein-protein docking. Protein Sci. 2005;14: 1328–1339.

155. Bastard K, Prévost C, Zacharias M. Accounting for loop flexibility during protein-protein docking. Proteins: Structure, Function, and Bioinformatics. 2005;62: 956–969. doi:10.1002/prot.20770

156. Pierce B, Weng Z. A combination of rescoring and refinement significantly improves protein docking performance. Proteins: Structure, Function, and Bioinformatics. 2008;72: 270–279. doi:10.1002/prot.21920

157. Pallara C, Rueda M, Abagyan R, Fernandez-Recio J. Conformational Heterogeneity of Unbound Proteins Enhances Recognition in Protein–Protein Encounters. J Chem Theory Comput. American Chemical Society; 2016;12: 3236–3249. doi:10.1021/acs.jctc.6b00204

158. Fiorucci S, Zacharias M. Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. Proteins: Structure, Function, and Bioinformatics. 2010;78: 3131–3139. doi:10.1002/prot.22808

159. de Vries S, Zacharias M. Flexible docking and refinement with a coarse-grained protein model using ATTRACT. Proteins: Structure, Function, and Bioinformatics. 2013;81: 2167–2174. doi:10.1002/prot.24400

160. de Vries SJ, Schindler CEM, Chauvot de Beauchêne I, Zacharias M. A Web Interface for Easy Flexible Protein-Protein Docking with ATTRACT. Biophys J. 2015;108: 462–465. doi:10.1016/j.bpj.2014.12.015

161. Frauenfelder H, Sligar S, Wolynes P. The energy landscapes and motions of proteins. Science. 1991;254: 1598–1603. doi:10.1126/science.1749933

162. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins Structure Function and …. 1998.

163. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. Journal of Molecular Biology. 2008;380: 742–756. doi:10.1016/j.jmb.2008.05.023

164. Davis IW, Arendall WB, Richardson DC, Richardson JS. The backrub motion: how protein backbone shrugs when a sidechain dances. Structure/Folding and Design. 2006;14: 265–274. doi:10.1016/j.str.2005.10.007

165. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. Amber 14. University of California, San Francisco 2014. 2014.

166. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C.

ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

167. Zacharias M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. Proteins: Structure, Function, and Bioinformatics. 2004;54: 759–767. doi:10.1002/prot.10637

168. Lensink MF, Wodak SJ. Blind predictions of protein interfaces by docking calculations in CAPRI. Proteins: Structure, Function, and Bioinformatics. 2010;78: 3085–3095. doi:10.1002/prot.22850

169. Habeck M, Nilges M, Rieping W. Replica-exchange Monte Carlo scheme for bayesian data analysis. Physical Review Letters. 2005;94: 018105. doi:10.1103/PhysRevLett.94.018105

170. Habeck M, Nilges M, Rieping W. Bayesian inference applied to macromolecular structure determination. Phys Rev E. 2005;72: 031912.

171. MacCallum JL, Perez A, Dill KA. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. Proc Natl Acad Sci USA. 2015;112: 6985–6990. doi:10.1073/pnas.1506788112

Acknowledgement

I would like to first express my special appreciation and thanks to my advisor Professor Dr. Martin Zacharias. He have been a tremendous mentor for me. I would particularly like to thank him for taking me as a member of group T38 to continue my Ph.D after Lange group dismissed. This means so much for me back then and now. His constructive guidance has fitted so very well with my former project and knowledge, and at the same time has also been prospective and novel. I also want to thank him for the discussions about my career as well as his kind regards to my kid and my family.

I owe many thanks to Dr. Oliver F. Lange for his kind yet strict training on me. I am very grateful that, as a highly efficient person, he would also slow down his pace to teach me with so great patience. He have literally taught me from zero about Rosetta and how to develop protocols in Rosetta. His office was always open. I could then always turn to him when there was a problem. He has also spent time to help to improve my presentation skills, which I will be benefited for sure for my whole life. Apart from work, I would also thank him for generously sharing and imparting his wisdom from life experience.

I would then like to thank Prof. Dr. Michael Sattler. After Oliver left, Prof. Sattler agreed to take me as his Ph.D student administratively. He has been very nice and supportive. I am still grateful and feel moved even now, that when I was facing a frustrating situation once in a time, he has asked with concern.

Hereby I wish to give some special thank to Mrs. Sonja Ortner, Mrs. Waltraud Wolfson and Mrs. Anna Hilber for their administrative support.

Now my thanks go to Dr. Sjord de Vries and Christina EM Schindler and all the colleagues in group T38, Dr. Zaiyong Zhang, Tatjana Braun, Justin Porter and all the other colleagues in former Lange group.

I also want to say thanks to my family, my parents, my parents in law and of course especially my husband and my son. They have always been there to support me.

Last but not least, I would like to thank Technische Universität München, Center for Integrated Protein Science Munich (CIPSM) and German Research Foundation (SFB1035) for financing my project.