PROTEIN-PROTEIN DOCKING REFINEMENT USING RESTRAINT MOLECULAR DYNAMICS SIMULATIONS

MARTIN ZACHARIAS

Physics Department T38, Technical University of Munich
James-Franck-Str. 1, 85748 Garching, Germany

(received: 29 August 2016; revised: 26 September 2016;
accepted: 30 September 2016; published online: 18 October 2016)

Abstract: A realistic prediction of the structure of protein-protein complexes is of major importance to obtain three-dimensional models for the interaction of proteins to form complexes and assemblies. In addition to the systematic search for putative binding sites on the surface of two binding partners, the second step of a docking effort, the refinement of docked complexes, is a major bottleneck to obtain realistic interaction geometries. Typically, the first initial systematic search employs rigid partner structures or few flexible degrees of freedom, whereas the refinement step involves fully flexible partner structures. The possibility to refine docked complexes using restraint MD simulations combined with an implicit solvent (Generalized Born) model was explored on three example test complexes starting from unbound partner structures. Significant improvement, both in scoring and agreement with the native complex structure after refinement was observed for two test cases. No improvement was found for a test case of a complex with lower binding affinity. The method can be easily applied to any docked protein-protein complex, however, more general applicability requires further improvements in the scoring function.

Keywords: protein-protein complex, docking prediction, force field modeling, implicit solvent modeling

DOI: https://doi.org/10.17466/tq2016/20.4/d

1. Introduction

The great majority of biological processes are based on protein-protein interactions. Hence, detailed understanding at the molecular or atomistic levels of these interactions is required to comprehend the function of these protein-protein interactions in a cell. Ultimately it also elucidates the possibilities of influencing these interactions by specifically designed drug molecules to interfere with protein-protein interactions. Functional understanding requires knowledge of the three-dimensional structure of protein-protein complexes [1–3]. The experimental determination of all complex structures is not possible due to the time and costs

involved and for many transient interactions it may not even be possible to obtain high resolution structural information. Therefore, a realistic prediction of protein-protein complexes (protein-protein docking) is of increasing importance [4]. Hence, the development and improvement of virtual docking methods to achieve realistic predictions is a rapidly evolving field. The prediction of protein complex structures is achieved by computer programs that evaluate all the information available on atomic interactions that are determined by the amino acid sequence of proteins, comparable to a complicated puzzle, where one tries to identify the correct mode of protein binding out of a myriad of alternative arrangements [1–4].

Typically, one can distinguish two stages of a protein-protein docking prediction simulation [1–3]. In the first phase the protein partners are often treated as rigid irregular bodies and the task is to identify arrangements that allow the best possible interface complementarity. A variety of methods to efficiently solve this task are available (reviewed in [1–3]). Evidently, the neglect of possible conformational changes upon binding during this early stage of docking may interfere with the chances to identify realistic solutions. Hence, in some approaches some degree of conformational change is included even at the first phase of a systematic search [3]. In the second step the docking solutions obtained from the first search step are subjected to a refinement procedure followed by a re-ranking step [4]. A number of refinement methods have been developed based on energy minimization [5], Monte Carlo or Molecular Dynamics (MD) simulations [6]. MD simulations are in principle very well suited for a fully flexible refinement because in such simulations every atom is mobile and surrounding water molecules and ions can be included explicitly. However, a major drawback is the large computational demand, if an explicit solvent is included and, even more severe, possible inaccuracies of the force field that may result in refined structures that deviate even more from the native structure than the rigidly docked initial model.

In order to avoid significant conformational changes during the MD-simulation refinement it is possible to include conformational restraints during the simulations. In the present contribution we test this possibility on the refinement of several docking starting models that deviate only modestly form the native complex structure. In addition, for computational efficiency, an implicit solvent model based on a Generalized Born continuum solvent model was employed. For the tested examples, the methodology shows promising results.

2. Materials and Methods

2.1. Protein-protein docking approach

For the generation of initial protein-protein docking solutions the ATTRACT docking program was employed [7]. It is based on a coarse-grained model such that the protein main chain is represented by two pseudo atoms per residue (located at the backbone nitrogen and backbone oxygen atoms). Small amino acid side chains (Ala, Asp, Asn, Cys, Ile, Leu, Pro, Ser, Thr, Val) are represented by one

pseudo atom (geometric mean of side chain heavy atoms) whereas larger and more flexible side chains are represented by two pseudo atoms accounting for the shape and dual chemical nature of some side chains [8]. The effective pseudo-atom interactions are described by soft distance(r_{ij})-dependent Lennard-Jones(LJ)-type potentials of the following form:

$$V = \varepsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 - \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \quad \text{in case of attractive pair} \quad (1)$$

$$V = -\varepsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 - \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \quad \text{repulsive pair if } r_{ij} > r_{\min} \quad (2)$$

$$V = 2e_{\min} + \varepsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 - \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \quad \text{repulsive pair if } r_{ij} \le r_{\min} \quad (3)$$

where R_{AB} and ε_{AB} are effective pairwise radii and attractive or repulsive Lennard-Jones parameters. At the distance $r_{\min}$ between two pseudo atoms the standard LJ-potential has the energy $e_{\min}$. A Coulomb type term accounts for electrostatic interactions between real charges (Lys, Arg, Glu, Asp) damped by a distance dependent dielectric constant ($\varepsilon = 15r$). In order to account for some flexibility of partners the program also allows protein partners to relax (deform) along pre-calculated soft collective degrees of freedom during the docking search. It is also possible to include energy minimization along the directions of a subset of normal modes simultaneously with the rigid body degrees of freedom [9, 10]. The soft collective degrees of freedom corresponded to eigenvectors of the proteins calculated using an approximate normal-mode analysis method (harmonic potential model) based on an Anisotropic Elastic Network model [10]. Normal modes were calculated with respect to the protein backbone (C_α atoms) and the side chains followed the same global motion as the corresponding C_α atoms. For a systematic docking search $\sim$250 different orientations of one protein partner (ligand) with respect to a second partner (receptor) were generated at approximately equally spaced points (separated by $\sim$7 Å) on the receptor protein surface. The resulting $\sim$ 50000–100000 starting arrangements were all energy-minimized to convergence and a subset of 20 complexes with a root mean square deviation of the ligand ($RMSD_{lig}$) after superposition of the receptor onto the native complex in the range of 5–15 Å was selected for further refinement. Docking was performed on unbound protein partner structures.

2.2. Generation of atomic resolution start structures for refinement

After docking based on the coarse-grained model (see above) atomic resolution protein partner structures were superimposed onto the docking solutions using the reduced representation in order to obtain docked protein-protein complexes. Energy minimization (2500 steps) using the Sander program from the Amber package [11] was used to eliminate the sterical overlap between atomic resolu-

tion partner structures. During the energy minimization a Generalized Born (GB) model was employed to implicitly account for solvation effects as implemented in Amber (igb = 5 option).

2.3. Refinement using restraint MD simulations

During restraint MD-simulation refinement the protein backbone structure of the partner molecules was restraint to stay reasonably close to the unbound reference structure using harmonic distance restraints between backbone CA atoms within each protein partner structure. All distances in the interval between 5 Å and 11 Å were considered and using a force constant of 2.0 kcal $mol^{-1}Å^{-2}$ to penalize any deviation with respect to the distance in the unbound reference structure. This choice keeps the secondary structure of the proteins close to the reference and also keeps close contacts between protein elements near the reference structure but allows full side chain flexibility and limited global rearrangement of protein secondary structure elements (long-range distances are not controlled). The test simulations using the GB implicit solvent model indicated that MD simulations using these restraints resulted in RMS deviations of individual protein partners from the references structure of < 1.5–2 Å (heavy atom RMSD) which was in the same order as the RMSD between most bound and unbound protein structures in the data base. For each selected docked starting structure 12 different sets of initial atom velocities were assigned. The refinement simulations were performed for 200 ps at a temperature of 250 K and using the Langevin integration scheme implemented in the pmemd.cuda program of the Amber package [11]. The final structure was energy minimized for 5000 steps and the interaction energy (including Coulomb, Lennard-Jones and GB reaction field contributions) between partners in the complex was evaluated (calculated by taking the difference in the energy of the complex minus the energy of isolated partners).

3. Results and Discussion

The accurate refinement of docked protein-protein complexes is one of the most critical steps for generating and identifying realistic complex structures [2, 3]. The performance of an MD-simulation based refinement procedure based on an atomistic resolution representation of the protein partners and a GB implicit solvent model was tested on three different protein-protein complexes. The protein complexes differed in affinity between partners in the complex. Whereas the complex pdb1CGQ and pdb1PPE corresponded to high affinity complexes that third case represented a medium affinity complex (pdb2OOB).

The refinement was performed starting from 20 different docked complexes obtained after a systematic search using the ATTRACT docking program [5] (see Methods section). The 20 complexes were selected based on similarity with respect to the bound complex representing the frequent scenario that the binding site of the partners was approximately known. The initial $RMSD_{lig}$ of the 20 start complexes was in the range of 5–15 Å and examples are illustrated for one of the complexes in Figure 1. Distance restraints within each partner were included

during refinement to avoid large scale conformational changes and dissociation of complexes during the refinement simulations. The restraints allowed full flexibility of the side chains but limited mobility of the backbone relative to the unbound structure of each partner protein. For all three cases, the MD-based refinement resulted in structures of significantly improved interaction energy between partners compared to slight energy minimization of the start structures (light blue circles in Figure 2). The drop in the interaction energy reaches 10–30 kcal mol^{-1} which indicates that energy minimization leads to local energy minima and MD simulations can overcome barriers which results in significant further optimization of the complex geometry and a drop in the interaction energy. Simulations performed at 280 K resulted overall in slightly lower interaction energies than simulations at 200 K (after energy minimization of final complexes). Note, that even higher simulation temperatures resulted in more structures with larger $RMSD_{lig}$ (data not shown). Most remarkably, the rather short MD simulations starting from the initial energy minimized docked complexes can result in quite significant changes of the $RMSD_{lig}$ meaning that the partners undergo significant rotational and translational motions on the surface of the protein partner even within 200 ps simulation time per refinement case. For the 1GCQ and 2PPE cases a large fraction of the simulations resulted in structures in a much closer agreement with the native structure of the complex (lower $RMSD_{lig}$ than $RMSD_{lig}$ of the start structure, light blue circles in Figure 2). Especially for the 1PPE case one can observe an interaction energy funnel for the energy landscape vs. $RMSD_{lig}$ in the range of $RMSD_{lig} < 10$ Å. The interaction energy funnel resulted in many solutions

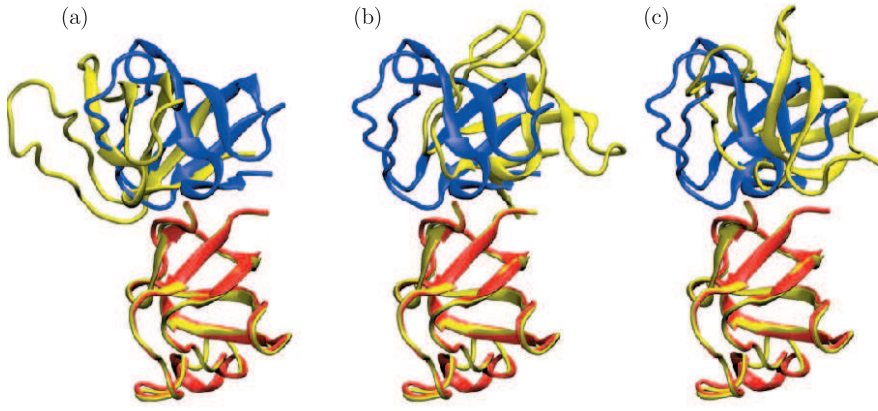

Figure 1. Example start structures for docking refinement using restraint MD simulations of target pdb1GCQ; The native complex structure is indicated as cartoon for the two partner proteins (indicated in red for the ligand and blue for the receptor, respectively); In the three example start structures (shown as yellow cartoon) obtained from systematic docking searches the ligand protein placement deviates from the native arrangement with $RMSD_{lig}$ in the range of 5–10 Å; Solutions of this quality are frequently reached in systematic searches employing rigid protein partner structures (in the unbound conformation)

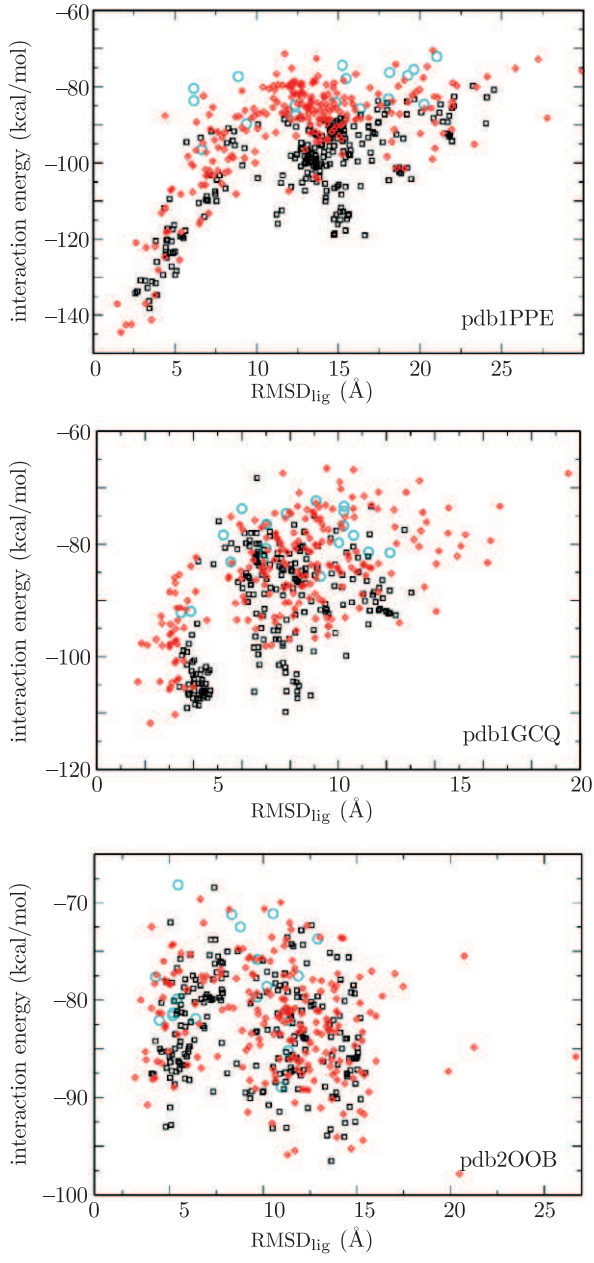

Figure 2. Interaction energy versus RMS deviation of the ligand protein (backbone atoms) from the native protein-protein complex for three test cases; The interaction energy was calculated after final energy minimization as the difference between complex energy and energy of isolated partner proteins (in the same conformation as in the complex); Light blue circles indicate start structures after energy minimization of the rigidly docked partners. For each of the 20 start structures 12 MD simulations with random velocity assignment were performed and after 200 ps at 200 K (black points) or 280 K (red points) structures were energy minimized and evaluated by calculating interaction energy and $RMSD_{lig}$

upon MD refinement with much improved $RMSD_{lig}$ and improved scoring (by the interaction energy). This result is also obtained, though to a lesser degree, for the second high-affinity protein-protein complex case (pdb1GCQ, Figure 2).

For the third case, a medium affinity complex, a different result was obtained. No interaction energy funnel could be observed and refined complexes were docked with large deviation from experiment giving a more favorable interaction energy score than the structures in closer agreement with the native complex structure. Also, in this case considerable displacement of the binding partners relative to the start geometry was observed during refinement but only very few structures with improved $RMSD_{lig}$ were sampled (Figure 2).

4. Conclusions

Refinement of docked protein-protein complexes to yield realistic complex geometries is one of the bottlenecks to predict the complex structure of interacting proteins. Refinement serves to achieve two main goals. A refinement procedure should improve the agreement between structural models and the native complex structure and, secondly, it should rank the structures close to the native complex as most favorable model structures. In the present study a simple and straightforwardly applicable MD-based refinement method was suggested that showed promising results at least for two high-affinity complexes in that it generated complex structures in closer agreement with experiment and much improved ranking. However, in the case of a medium affinity complex it failed to improve the scoring of near native complexes but still resulted in a broad ensemble of complexes (compared to the start structures). Hence, the study indicates that even relatively short MD simulations using partner structures restraint to the unbound reference conformations, in favorable cases, allow significant motions of the partners. However, improvements of the interaction energy scoring are necessary in the future to use the approach as a reliable method that could be generally applicable. Also, improvements of the scoring function need to be tested on a larger benchmark set of a complex structure.

Acknowledgements

I thank Dr. Isaure de Beauchene for helpful discussions and the Leibnitz Supercomputer center for providing computational resources (grantnumber pr48ko).

References

[1] Bonvin A M 2006 *Curr. Opin. Struct. Biol.* **16** 194
[2] Andrusier N, Mashiach E, Nussinov R, Wolfson H J 2008 *Proteins* **73** 271
[3] Zacharias M 2010 *Curr. Opin. Struct. Biol.* **20** 180
[4] Lensink M F, Wodak S J 2013 *Proteins* **81** 2082
[5] Fernandez-Recio J, Totrov M, Abagyan R 2003 *Proteins* **52** 113
[6] Mashiach E, Nussinov R, Wolfson H J 2010 *Nucleic Acids Res.* **38**, W457
[7] Zacharias M 2003 *Protein Sci.* **12** 1271
[8] Fiorucci S, Zacharias M 2010 *Prot. Struct. Funct. Bioinf.* **78** 3131

[9] Vries S de, Zacharias M 2013 *Prot. Struct. Funct. Bioinf.* **81** 2167
[10] May A, Zacharias M 2005 *Biochim Biophys Acta, Proteins Proteomics* **1754** 225
[11] Case D, Darden T, Cheatham III T, Simmerling C, Wang J, Duke R, Luo R, Walker R, Zhang W, Merz K, Paesani F, Roe D, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling C, Smith W, Swails J, Walker R, Wang J, Wolf R, Wu X, Kollman P 2014 *AMBER* **14**