Incorporating flexibility into protein-ligand docking

Manuel Rueda, PhD
Joint BSC-IRB research programme in Computational Biology
Barcelona Supercomputing Center
C/ Jordi Girona 29, 08034 Barcelona, Spain

Summary
Structure-based drug design (SBDD) is routinely used to identify new active compounds in early stages of drug development. Among the computer-aided drug design approaches, docking methods are the tool of choice for the prediction of ligand-receptor interactions. Unfortunately, the success of molecular docking is hampered by the sparse representation of protein dynamics. The principal objective of the "IFLPD" project has been to develop methods and software to introduce protein flexibility in the context of small molecule docking and protein-protein docking. Many Universities and Pharmaceutical companies are using the public applications and websites created during the project.

Description of the work performed and main results achieved during the "outgoing phase" and the "returning phase".
Dealing efficiently with protein flexibility and understanding the effects of dynamics in ligand binding is one of the main challenges in the world of computer-aided drug design. One simple and efficient way of representing protein plasticity is by using multiple "static" receptor conformations, also known as ensemble docking. The ensembles can consist of multiple experimental structures, computationally created models, or both. Most of the results obtained during this project consist of variations of the *ensemble docking* procedure, applied in different medicinal chemistry scenarios. The most significant results will be summarized in the next paragraphs.

Exploiting experimental conformational diversity in protein-ligand docking. Experimental 3D structures are the main source of conformational variability for protein-ligand docking and virtual screening campaigns. According to our studies, ensemble docking with x-ray protein structures displayed better discrimination of actives from inactive ligands than single receptor docking. Moreover, the use multiple receptor conformations enhanced the chemical diversity of the compounds. On average, ensembles consisting of 3-5 conformers generated better discrimination values than individual conformations (1-3).

ALiBERO: A new computational method that improves the recognition of active compounds in distant homology models. For ~50% of the therapeutically relevant protein targets, no experimental 3D structures are available. For these cases, homology modeling can be used to generate protein models. Unfortunately, raw homology models usually display poor docking performance. During this project, we developed a method that optimizes homology models so that their docking performance is on par with x-ray structures. The method is called ALiBERO (3)(Automatic Ligand-guided Backbone Ensemble Receptor Optimization) and performs an iterative search based on two main steps: (i) generation of multiple receptor conformers (4), and, (ii) selection of the conformers according to docking/VLS performance. The method has being successfully

applied to many important nervous system targets, such as the A2A adenosine receptor, the Dopamine D3 receptor and the 5-HT_{1A} serotonin receptor. We also developed a method to check the precision of the models (SimiCon) that allows automated identification of equivalent protein–ligand atomic contacts (5). SimiCon was implemented in a free web server located at http://abagyan.ucsd.edu/SimiCon.

Worldwide modeling assessment on GPCR structure. GPCR Dock 2010 community-wide assessment was conducted by our group to evaluate the status of molecular modeling and ligand docking for three recent GPCR targets of varying difficulty: dopamine D3 and CXCR4 chemokine receptors bound to small molecule antagonists and CXCR4 with a synthetic cyclopeptide. According to the results, the fact that D3 had closer homolog for homology modeling, combined with the use of modern docking protocols and QSAR information, allowed for prediction of complexes with atomic details of accuracy approaching experimental. On the contrary, CXCR4 complexes, that only possess distant homology to the available GPCR structures, still remained very challenging (6). The results were published as a web site at http://abagyan.ucsd.edu/GPCRDock2010.

Protein-protein interactions. In a recent paper (7), we investigated the cost of backbone conformational changes upon association in a large dataset consisting of 2090 unique unbound-to-bound domain transitions. According to our estimates, 65% of the transitions did not show significant changes upon binding (i.e., the RMSD unbound to bound was $\leq$ 1.5 Å), 13% explored the bound conformation in the unbound state (conformational selection model), and only 2% clearly require external energy (induced fit model). An important fact that came up from the study was that domains with many partners tend to undergo smaller changes upon association and are less likely to freely explore larger adaptations. Some of the estimates were derived from molecular dynamics (MD) simulations performed to 70 protein domains, starting from the unbound form. The trajectories were part of a large repository of MD simulations (MoDEL: Molecular Dynamics Extended Library) which consists of > 1700 trajectories of proteins representative of monomeric soluble structures in the Protein Data Bank (PDB)(8).

References:

1. Bottegoni G, Rocchia W, Rueda M, Abagyan R, & Cavalli A (2011) Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLoS One* 6(5):e18845.
2. Rueda M, Bottegoni G, & Abagyan R (2010) Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *Journal of chemical information and modeling* 50(1):186-193.
3. Rueda M, Totrov M, & Abagyan R (2012) ALiBERO: evolving a team of complementary pocket conformations rather than a single leader. *Journal of chemical information and modeling* 52(10):2705-2714.
4. Rueda M, Bottegoni G, & Abagyan R (2009) Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *Journal of chemical information and modeling* 49(3):716-725.
5. Rueda M, Katritch V, Raush E, & Abagyan R (2010) SimiCon: a web tool for protein-ligand model comparison through calculation of equivalent atomic contacts. *Bioinformatics* 26(21):2784-2785.
6. Kufareva I, Rueda M, Katritch V, Stevens RC, & Abagyan R (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* 19(8):1108-1126.
7. Stein A, Rueda M, Panjkovich A, Orozco M, & Aloy P (2011) A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. *Structure* 19(6):881-889.
8. Meyer T, *et al.* (2010) MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories. *Structure* 18(11):1399-1409.