Flexible macromolecular docking

CECAM Workhop
Lyon, 28-30 April, 2004

Abstracts

Conformation changes and the specificity of protein-protein interaction

Joël Janin, Laboratoire d'Enzymologie et de Biochimie Structurales, CNRS, Gif-sur-Yvette, France (janin@lebs.cnrs-gif.fr)

Molecular docking algorithms assemble a two-pieces puzzle, which would be a child game if the components were rigid like a lock and a key. In reality, molecules (small or large) change conformation as they associate, a feature that all protein-protein docking procedures must take into account. Whereas existing procedures generally succeed when the conformation changes are small, they fail to reproduce large changes. These are nevertheless common, and in many biological systems, they are essential to the function. Changes seen upon association can be local (loop movement) or global (dimerization), and they may include disorder-to-order transitions, making protein-protein interaction of similar complexity to protein folding.

Specific protein-protein complexes and homodimeric proteins form interfaces that are large and compact, with close-packed interface atoms. In contrast, the non-specific interaction observed in protein crystal packing generate small, loosely packed interfaces. These structural differences are easily interpreted in terms of geometric complementarity in cases where conformation changes are small and recognition takes place between preformed surfaces. In contrast, large changes at an interface imply that recognition first occurs between surfaces that are not complementary. A basic question in molecular assembly is how this process takes place, and whether we can reproduce it in docking procedure.

Analysis of the impact of deformations on the results of the CAPRI experiment

Miri Eisenstein, Weizmann Institute of Science, Rehovot, Israel (miriam.eisenstein@weizmann.ac.il)

The results of the CAPRI experiment indicate that rigid body docking procedures are able to tolerate considerable structural deformations (e.g. T01, T10 and T11). However, larger deformations that involve hinge movements, are a problem not only when they are real, as in T09, but also when they are anticipated, as in T11 and T13. It appears that the current scoring functions cannot be used to predict if a hinge movement is likely to occur or not. In cases when a hinge movement is known to occur, a multi-rigid-body approach can be used to predict the structure of the complex.

How Flexible your Rigid-Body Docking Should Be

Carlos J. Camacho, Boston University, USA (ccamacho@bu.edu)

I will discuss the feasibility of structural refinement in the context of protein-protein docking, emphasizing the benefits/disadvantages of backbone and side chain flexibility. I will mention the biophysical motivation for addressing the side chain refinement problem, and suggest how to solve it. Examples from blind predictions in the CAPRI experiment will also be mentioned.

ICM optimization of flexible interface side-chains in protein-protein docking: successes and limitations

Juan Fernandez-Recio, University of Cambridge, UK (juan@cryst.bioc.cam.ac.uk)

The ICM Docking and Interface Side-Chain Optimization (ICM-DISCO) was benchmarked in 24 unbound pairs for protein-protein docking [Fernandez-Recio et al. (2002) Protein Sci. 11, 280-291] and successfully evaluated in the blind CAPRI experiment (http://capri.ebi.ac.uk) [Fernandez-Recio et al. (2003) Proteins 52, 113-117]. The rigid-body docking step is able to provide thousands of candidate poses ranked by interacting energy, and gives important information about the location of the putative protein-protein interaction sites [Fernandez-Recio et al. (2004) J.Mol.Biol. 335, 843-865]. However, it is the global energy optimization of the flexible ligand interface side-chains that ultimately helps to identify the correct geometry of the complex. This flexible refinement step is especially efficient in protease-inhibitors, and generally, in cases where only a few steric clashes between the unbound side-chains need to be resolved in order to achieve the final complex structure. However, in some other cases the ligand side-chain optimization protocol is not enough to achieve the optimized fit of the interacting molecules. Both successful and not-so-successful stories will be analyzed here,and we will discuss new ways of improving flexible refinement of the interfaces in internal coordinates.

Flexibility as part of the geometric filtering problem.

Ludwig Krippahl, Universidade Nova de Lisboa, Portugal (ludik@netcabo.pt)

When modelling protein complexes, geometric complementarity is generally the most important criterion for filtering the large set of possible models and reducing it to a manageable sub-set. This necessary reliance on geometry makes protein flexibility a major problem. Two solutions for this problem are either to account for flexibility implicitly by relaxing the stringency of the filter, or to retain a stringent filter and model different conformations explicitly. Though the latter approach has become more popular with the increase in computation power, we propose that, in many cases, flexibility can be considered as part of the general problem of the reliability geometric complementarity as a predictor of protein interaction.

To this end, our docking algorithm BiGGER [1] uses an implicit representation of protein flexibility that can distinguish rigid and flexible regions, and can incorporate experimental data as an additional filter to compensate the lower stringency of the geometric complementarity filter [2].

[1] Palma, P.N., Krippahl, L.,Wampler, J.W., Moura, J.G., A New (Soft) Docking Algorithm for Predicting Protein Interactions. Protein:Struc. Func. Gen. 2000 Jun 1;39(4):372-84.

[2] Krippahl, L., Moura, J.J., Palma, P.N., Modeling Protein Complexes with BiGGER. Prot: Struc.Funct. Gen, V. 52(1):19-23.

Modelling the structure of protein-protein complexes

Michael J E Sternberg, Patrick Aloy, Philip Carter, Henry Gaab, Suhail Islam, Richard Jackson, Victor Lesk, Gidon Moont, F. Pazos & Graham Smith.
Structural Bioinformatics Group, Dept of Biological Sciences, Imperial College London, UK (m.sternberg@imperial.ac.uk, www.sbg.bio.ic.ac.uk)

This talk will describe the current status of the package 3D-DOCK that aims to predict the 3D structure of a protein-protein complex starting from the coordinates of the unbound components. Protein flexibility is introduced in the final stage via a program MULTIDOCK that performs a rigid-body refinement coupled with optimisation of side-chain / side-chain packing is performed The talk will describe the status of the above approach on a test data set. Finally the results of the recent blind trial of protein-protein docking (CAPRI, www.capri.ebi.ac.uk) will be reported.

The talk will also report recent work to predict functional residues from sequence and structure using a new approach that predicts protein function and the responsible residues over a wide range of functional specificities.

Simulating induced fit in molecular docking.

Ruben Abagyan, Maxim Totrov, Juan-Fernandez Recio, Julio Kovacs, Claudio Cavasotto.
The Scripps Research Institute, La Jolla, USA (abagyan@scripps.edu)

The main complicating factor in molecular docking is receptor rearrangement upon ligand binding (induced fit). It is the induced fit that complicates cross-docking of ligands from different ligand receptor complexes. To improve on discriminating between binders and nonbinders in the virtual screening process we developed a protocol which performs receptor-flexible docking of known ligands in order to simulate possible pocket rearrangements. This protocol was applied to a benchmark of kinases and was demonstrated to improve both the cross-docking accuracy as well as the "enrichment" in virtual ligand screening. In protein-protein docking and peptide protein docking the side-chain sampling may be sufficient to account for induced fit. The induced changes of the backbone are more problematic. We show how the slow modes of soft harmonic Cα-model can be used to generate alternative conformations.

Modeling large-scale hinge-bent motions in docking.

Dina Schneidman & Ruth Nussinov, Tel Aviv University, Israel (duhovka@tau.ac.il)

Proteins are very flexible molecules. The flexibility may range from small-scale side-chain motions to large-scale intra and inter domain motions or even partial refolding. In my talk I will focus on approaches to handle hinge-bent protein flexibility in docking algorithms. Hinge detection strategies will also be mentioned. I will discuss the problems of the current methods and the challenges of the field. In addition, I will present various examples, including CAPRI targets.

Accounting for protein loop flexibility during macromolecular docking

Karine Bastard & Chantal Prévost, Laboratoire de Biochimie Théorique, Paris, France
(Karine.Bastard@ibpc.fr)

Upon macromolecular association, some proteins undergo large conformational changes that can result in surface loop movements. When the Met repressor binds to DNA, an eight residue loop of Met repressor changes its hairpin conformation into a conformation that wraps around the DNA phosphate backbone. Such an examples confirm the necessity to account for induced surface remodeling during the search for interacting surfaces, by allowing the receptor to adapt to its partner in an induced fit process. To address this problem, we have recently developped a new docking method, termed MC2, which takes into account the loop and side-chain movements at the protein surface during macromolecular association. The objectives of MC2 are to precisely position the ligand, predict the loop conformations that optimally interact with the ligand and adjust the side-chain conformations, in order to predict the atomic level interactions between the two partners. The loop flexibility is artificially introduced by using a multiple copy representation. Each loop copy results from ab intio construction and represents one possible main-chain conformation of the loop with rigid backbone and flexible side-chains. The ligand position, the conformation of the protein side-chains and of the loop copy side-chains are sampled by a Monte-Carlo Simulated Annealing process. The multiple copy representation and Monte Carlo simulation are coupled via the copy weights which are recalculated at the end of each Monte Carlo cycle, finally resulting in selecting a unique loop copy at the end of MC2 process. Final loop adjustments, via energy minimzation, is found to play an important role in establishing the correct energy ranking. In a test-case study, the method was able to predict the structure of the complex at the atomic level and to unambiguously predict the conformation of an interfacial loop.

Bastard K, Thureau A, Lavery R, Prevost C. Docking macromolecules with flexible segments.
J.Comput.Chem. 2003 Nov 30;24(15):1910-20.

http://www.ibpc.fr/~bastard/MC2/mc2.html

How to efficiently account for side chain flexibility and global motions during docking

Martin Zacharias, International University Bremen, Germany (zacharias@iu-bremen.de)

Most current docking approaches to predict the binding geometry of protein-protein complexes use rigid protein partner structures. However, protein complex formation can involve both local conformational changes of side chains and loops at the protein-protein interface and global conformational relaxation of the protein partners. We have developed a docking approach that is based on energy minimization of translational and rotational degrees of freedom of protein partners and on a reduced protein representation allowing efficient search for docking minima. A multicopy approach is used to select the most favourable side-chain conformation at the protein-protein interface during the docking process [1]. To approximately account for possible global conformational adaptation a method has been developed that allows to relax the protein structure in pre-calculated flexible degrees of freedom (soft modes) during docking [2]. Such flexible modes can for example be obtained from molecular dynamics simulations or on the level of a reduced protein representation by employing an energy function that depends on the local protein density. Application of the approaches to test systems will be presented.

[1] Zacharias, M. 2003. Protein-protein docking with a reduced protein model accounting for side chain flexibility. Protein Sci. 12, 1271.
[2] Zacharias, M. 2004. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability:binding of FK506 to FKBP. Proteins 54, 759.

HADDOCK: an information-driven flexible docking approach

Alexandre M. J. J. Bonvin, Utrecht University, The Netherlands (a.m.j.j.bonvin@chem.uu.nl)

In my talk, I will describe our recently developed information-driven flexible docking approach HADDOCK (High Ambiguity Driven protein protein DOCKing) (http://www.nmr.chem.uu.nl/haddock), that makes use of biochemical and/or biophysical information. The experimental information is introduced as highly ambiguous interaction restraints (AIRs) to drive the docking process.
HADDOCK uses an all-atom representation of the system. Flexibility is accounted for in different ways during the docking protocol:
i) in the initial rigid body energy minimization stage by starting the docking from ensembles of conformations (e.g. a NMR ensemble of structures, snapshots from a MD simulation)
ii) during the semi-flexible simulated annealing refinement stage by allowing flexibility at the interface first, only for side-chain atoms, and then, for both side-chain and backbone atoms
iii) in the final SA refinement in explicit water by progressively allowing flexibility in the remaining of the system in addition to the defined, flexible interface.

Reference:
Dominguez, C. Boelens, R. and Bonvin, A.M.J.J. (2003). J. Am. Chem. Soc. 113. 1731

How may the use of MD and rigid-body docking algorithms overcome the protein flexibility problem associated with complex formation?

Graham R. Smith, Cancer Research UK London Research Institute (graham.smith@cancer.org.uk)
Michael J. E. Sternberg, Imperial College London, UK. (m.sternberg@imperial.ac.uk)
Paul A. Bates, Cancer Research UK London Research Institute (paul.bates@cancer.org.uk)

The formation of a protein-protein complex is a key event in an enormous number of cellular biochemical processes. However, to predict a wild-type complex computationally given the structures of the components (the "protein docking problem") is still difficult in cases where there is any more than a very small change in the conformation of the components upon the formation of the complex. As a first step to addressing this flexible docking problem, we have used Molecular Dynamics (MD) simulations to investigate the extent to which the conformational fluctuations undergone by proteins in solution reflect the conformational changes that they undergo when they form protein-protein complexes ("induced fit"). To do this, we study a set of over thirty proteins that form such complexes and whose 3-dimensional structures are known, both bound in the complex and unbound. We carry out MD simulations of 5 ns duration with Gromacs, starting from the unbound structures, and analyse

the resulting conformational fluctuations in comparison with the structures in the complex.

We find that in some cases the conformational fluctuations observed in MD correlate well with the regions of the proteins that move on complex formation, and in some cases take the protein towards its bound conformation.

Preliminary results are presented on how this information may be used to improve protein-protein docking, both for the test set described above and some targets from recent rounds of CAPRI.

Combinatorial docking for multi-molecular assembly and protein structure prediction

Yuval Inbar & Haim J. Wolfson, Tel Aviv University, Israel (inbaryuv@tau.ac.il)

The majority of proteins function when associated in multimolecular assemblies. Yet, prediction of the structures of multimolecular complexes has largely not been addressed, probably due to the magnitude of the combinatorial complexity of the problem. Docking applications have traditionally been used to predict pairwise interactions between molecules. We have developed an algorithm that extends the application of docking to multi-molecular assemblies.

We apply it to predict both quaternary structures of oligomers and multi-proteins complexes. Moreover, adapting the algorithm to consider backbone connectivity, we also show that it may be useful in the prediction of protein tertiary structures when the structures of the protein parts are available. This application was tested both on domain assembly in order to predict the spatial arrangement of domains in multi-domain proteins, and on protein building blocks (substructures of domains with relatively high population times) assembly to predict their arrangement within a domain in the native protein.

Complementarity of structure ensembles in protein-protein binding

Raik Grünberg*, Johan Leckner* & Michael Nilges. Institut Pasteur, Paris, France (raik@pasteur.fr)

Our understanding of protein-protein interaction is caught in a contradiction: on the one hand,experimental rates of association suggest that, in many cases, practically every collision between two partner proteins leads to the formation of the complex. On the other hand, we often fail to predict the correct orientation of a protein complex because the two free partners simply don't sufficiently fit. This discrepancy is commonly explained by a fuzzy notion of induced fit, or by the assumption that the bound conformations is present in the structure ensembles of the two unbound proteins. However, both models appear to be inconsistent with our current knowledge about the forces and time scales of recognition.

In this study, we try to incorporate the additional dimensions of receptor and ligand variability into our picture of the protein-protein binding process. We performed two sets of molecular dynamics simulations for the unbound (free) structures of 17 receptor and 16 ligand proteins and applied shape-driven rigid body docking to all combinations of representative receptor and ligand snapshots as well as the free structure. In total, we analysed and compared 2,106,368 solutions from 4114 exhaustive rigid body dockings between 693 conformations of 33 different proteins. The cross-docking of ensemble snapshots increases the chances to find near native orientations. Our results suggest that there are complementary conformations within the free receptor and ligand ensembles, which, however are in general not necessarily related to the bound structure. In addition, we also performed molecular dynamics simulations on all 17 complexes and analysed the flexibility of free and bound proteins. Our results indicate that binding may not necessarily occur at the cost of entropy. We propose a refined model of the protein-protein recognition process that is combining the ideas of conformer selection and induced fit and is in better aggreement with our current understanding of interaction forces, time scales and kinetic data.

* these authors contributed equally to the work

Prediction of interacting surfaces by the Evolutionary Trace method

Olivier Lichtarge, Baylor College of Medicine, Houston, USA (lichtarge@bcm.tmc.edu)

Protein-protein interactions are the elementary units from which molecular pathways and cellular networks are built. But the description of the functional surfaces that determine protein binding still elude us. The Evolutionary Trace (ET) approach to this problem is to combine sequences, evolutionary trees, and structures to reveal the canonical determinants of a protein¹s function. Large-scale studies show that these determinants cluster spatially in the structure and that they match functional sites on proteins surfaces. Their discovery allows experimentalists to rationally design activity through targeted mutagenesis, for example along the G protein-signaling pathway. The scalability and generality of ET further suggest that proteome-wide annotation of functional sites is within reach. The activity of many protein structures may then be traced to narrow sets of relevant amino acids that form ³elementary units of function and of interaction². From a practical viewpoint, these units can be engineered to analyze and manipulate the molecular basis of protein function.The majority of proteins function when associated in multimolecular assemblies. Yet, prediction of the structures of multimolecular complexes has largely not been addressed, probably due to the magnitude of the combinatorial complexity of the problem. Docking applications have traditionally been used to predict pairwise interactions between molecules. We have developed an algorithm that extends the application of docking to multi-molecular assemblies.

We apply it to predict both quaternary structures of oligomers and multi-proteins complexes. Moreover, adapting the algorithm to consider backbone connectivity, we also show that it may be useful in the prediction of protein tertiary structures when the structures of the protein parts are available. This application was tested both on domain assembly in order to predict the spatial arrangement of domains in multi-domain proteins, and on protein building blocks (substructures of domains with relatively high population times) assembly to predict their arrangement within a domain in the native protein.

Modeling Correlated Protein Main-chain Motions in Proteins and their Ligands

Leslie A. Kuhn(1), Maria I. Zavodszky(1), Sameer Arora(2), Ming Lei(3), and Michael F. Thorpe(4)

(1) Department of Biochemistry & Molecular Biology and Center for Biological Modeling, Michigan State University, 502C Biochemistry Building, East Lansing, MI 48824-1319; http://www.bch.msu.edu/labs/kuhn (2) Departments of Biochemistry & Molecular Biology and Computer Science & Engineering, Michigan State University, (3)Department of Biochemistry, Brandeis University, and (4)Physics & Astronomy Department, Arizona State University (KuhnL@msu.edu)

We describe a new method for modeling protein and ligand main-chain flexibility in docking. The goal is to sample the full conformational space, including conformations not yet observed by crystallography, MD, or NMR. Flexibility analysis is performed using the graph-theoretic algorithm FIRST, which identifies coupled networks of covalent and non-covalent bonds within the protein. ROCK then explores available conformations by only sampling dihedral angles that preserve the coupled bond network in the protein. A representative set of protein conformations can then be used as targets for docking with SLIDE, which models protein and ligand side-chain flexibility. This combined approach for incorporating main-chain flexibility in docking is illustrated for cyclophilin A-cyclosporin and estrogen receptor-zearalenol complexes. Very recent results show that the maintenance of correlated motions between hydrogen-bonded and hydrophobic side chains is also a key aspect of ligand recognition across diverse protein-ligand complexes.

Protein flexibility and drug design: How to hit a moving target

Heather A. Carlson, University of Michigan, USA (carlsonh@umich.edu)

The use of multiple protein structures (MPS) is a growing trend in structure-based drug design. Different techniques will be discussed, and our MPS method for developing receptor-based pharmacophore models will be highlighted. By using MPS, we are able to identify flexible and rigid regions within the binding site and use that information to our advantage. An additional advantage of the method is that an unbound protein structure can be used sucessfully for structure-based inhibitor design !

Posters

1. **Structural Study of the Tenase Complex**

Ludovic Autin
Equipe Bioinformatique Structurale, Paris, France (ludovic.autin@univ-paris5.fr)

2. **Understanding Molecular Recognition: A Dissection of Specific and Non-specific Protein-Protein Interfaces**

Ranjit P. Bahadur[1], Pinak Chakrabarti[1], Francis Rodier[2] and Joel Janin[2]
[1]Department of Biochemistry, Calcutta, India (b_ranjit@bic.boseinst.ernet.in)
[2]Laboratoire d'Enzymologie et de Biochimie Structurales, Gif-sur-Yvette, France

3. **A new docking scoring function based on interface geometry and physico-chemical residue properties**

Julie Bernauer
Laboratoire d'Enzymologie et Biochimie Structurales, Gif-sur-Yvette Cedex, France
(bernauer@lebs.cnrs-gif.fr)

4. **Molecular Shape Analysis based upon the Morse-Smale Complex and the Connolly Function**

Frederic Cazals
INRIA Sophia-Antipolis, France (Frederic.Cazals@sophia.inria.fr)

5. **Haddock's adventures in CAPRI**

A.D.J. van Dijk, C. Dominguez, S.J. de Vries and A.M.J.J. Bonvin
Department of NMR Spectroscopy, Utrecht University, Netherlands (a.j.vandijk@chem.uu.nl)

6. **Automatic structural modelling for peptide/MHC complexes**

Quentin Kaas
Laboratoire d'ImmunoGenetique Moleculaire, Institut de Genetique Humaine, Montpellier
(kaas@ligm.igh.cnrs.fr)

7. **Investigating Different Starting Structures Using the MPS Pharmacophore Method.**

Kristin L Meagher and Heather A Carlson,
College of Pharmacy, University of Michigan, US (kmeagher@umich.edu)

8. **In silico studies of type-I interferons:using a protein docking method to highlight differences of IFN-alpha and IFN-beta binding to the IFN receptor chain 2 Towards a 3-D model of the IFN/receptor complex**

Florence Nosal
Department of Bioinformatics & IT, GenOdyssee S.A, Courtaboeuf , France
(nosal@genodyssee.com)

9. **Application of a new potential scaling approach to refine protein-ligand interfaces and protein cores**

Ralph Nico Riemann

School of Engineering and Science, Campus Ring 1, 28759 Bremen, Germany
(r.riemann@iu-bremen.de)

10. **Using robotic algorithms as new tools to model loop flexibility for protein interactions**

J. Cortes[1], V. Tran[1], T. Simeon[2]

[1]Unite de recherche sur la Biocatalyse, Faculte des Sciences et Techniques, Nantes, France
(Vinh.Tran@chimbio.univ-nantes.fr)

[2]LAAS/CNRS, Toulouse, France (nic@laas.fr)

11. **Evaluation of some docking procedure basics: from geometric to atomistic picture**

P. Puech[1] H. Hoyet[1], M. Djafari Rouhani[2], M. Erard[3], A. Esteve[2], G. Landa[1]

[1] Laboratoire de Physique des Solides, Toulouse

[2] LAAS-CNRS, Toulouse (djafari@laas.fr)

[3] IPBS, Toulouse