Structure, Vol. 12, 2095–2102, December, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.11.001

Introducing a 4th Dimension to Protein-Protein Docking

A three-step model (diffusion, conformer selection, and induced fit) is proposed to describe molecular recognition processes, merging previous theories of protein association (Grünberg et al., 2004, this issue of *Structure*). The model was arrived at by docking ensembles of structures.

The formation and disassembly of specific molecular complexes are a part of almost every biological process. Complexes are stabilized by a myriad of weak interactions, some of which require precise relative positioning (e.g., hydrogen bonds), whereas others are geometrically less restrictive (e.g., van der Waals interactions). Hence, molecular recognition is an intricate process, particularly when in solution, governed by thermodynamics and kinetics (Janin, 1995) and directed to form biologically functional complexes. The latter are as stable as necessary since functionality calls for a balance between stability and flexibility. The article by Grünberg, Leckner, and Nilges, "Complementarity of Structure Ensembles in Protein-Protein Binding" (2004) is concerned with the process of recognition, studied via molecular docking. It therefore touches two related fields: understanding of the principles that underlie molecular recognition and predictive docking.

The molecules in a complex (bound) have different conformations than in the free states (unbound). The differences are small and local in some cases but in other cases large deformations are observed (Betts and Sternberg, 1999; Lo Conte et al., 1999). Hence, the molecular recognition process should be envisaged as a movie. The players in this motion picture are the vibrating ever-changing solvated molecules. Previously, two models of the recognition process were proposed: induced fit, namely changing of conformations as the molecules come close together and start affecting each other (Koshland, 1958), and conformer selection, which postulates that the ensemble of conformers of the free molecule includes a bound-like conformer (Kumar et al., 2000). The latter model is an extension of the allosteric regulation model (Monod et al., 1965).

Grünberg et al. (2004) combine the former recognition models by proposing a three-step recognition process: diffusion, conformer selection, and induced fit. They arrive at this notion by crossdocking ensembles of structures. Each molecule is described by an ensemble of conformers, a series of snapshots which when combined together portray a vibrating molecule. Interestingly, the ensembles, generated by molecular dynamics, do not include either the bound conformer or global transitions from the free to bound interface. This may result from the sparse sampling of conformers, but it may also describe a real situation—low abundance of the bound conformer due to its relatively high energy. The latter notion is supported by the fact that ensembles of NMR structures of a free molecule do not include its bound conformer.

The crossdocking of ensembles produces considerable enrichment of encounter models that possess some near native contacts, compared to docking of single conformers. Often more accurate models are obtained (Grünberg et al., 2004). Yet, many of the encounter models differ from the structure of the native complex and from one another. This led to the suggestion that recognition does not depend upon the bound structure and association can proceed from different coexisting recognition complexes, with partially correct contacts. Notably, the docking of ensembles corresponds to the second step in the proposed three-step recognition process—the selection of conformers. This step is characterized by entropy changes— the loss of conformational entropy and the gain of desolvation entropy. Progressive desolvation and short-range forces facilitate the next step, the refolding into the bound structure. In this context, it is worthwhile to mention the study by Kimura et al. (2001), in which smaller conformation changes, confined to side chain movements, were analyzed. It was found that key side chains, most important for association, frequently sample the proper conformation for binding.

The difference between the conformations of bound and unbound molecules is a severe obstacle in predictive docking. Docking starts from known native structures of molecules, usually experimental structures obtained by X-ray crystallography or NMR spectroscopy. These structures represent a single snapshot of an averaged prevalent conformer in the crystalline environment in the first case and a selection of snapshots in the second case that can provide information on possible deformations in the molecule, but not on the bound structure. In many docking algorithms, the flexibility of the interacting molecules is considered globally, by treating the molecules as soft rigid bodies, namely rigid objects with a "shock absorbing" surface able to tolerate small clashes (Eisenstein and Katchalski-Katzir, 2004). This approach was recently extended to allow for large domain movements, while treating each domain as a rigid body (Inbar, et al., 2003). Other docking methods consider conformation changes explicitly (Vajda et al., 1997; Fernandez-Recio et al., 2003; Gray et al., 2003; Zacharias, 2003). Can the new approach of ensemble-to-ensemble docking improve the predictions? The enrichment of models with near native interactions suggests that it can. However, in predictive docking, the detection of near native models is not enough. Such models must also be distinguished from false models and ranked high. The immense number of solutions produced in crossensemble docking includes many false solutions, some of which rank higher than the near native ones. The enrichment of such solutions has yet to be analyzed. Finding a needle in a haystack is hard; it is

even harder to single out an approximate needle in a larger haystack.

It is likely that the study of Grünberg et al. (2004) will inspire similar studies in which different scoring functions and ranking procedures will be employed. Several questions should be addressed: for example, how large should the ensemble be in terms of structural diversity and energy differences in order to simulate the association process? The answer probably depends on the system studied and on the anticipated conformation changes, leading to the next question: can molecular dynamics simulations provide the necessary diversity of conformers for systems in which domain movements occur? A very large ensemble, which can but does not necessarily include conformers close to the bound structure, is prohibitive in terms of computation time. These and other questions will have to be considered, perhaps keeping in mind the effects of macromolecular crowding in the cell (Minton, 2000).

Miriam Eisenstein
Department of Chemical Research Support
Weizmann Institute of Science
Rehovot 76100
Israel

Selected Reading

Betts, M.J., and Sternberg, M.J. (1999). Protein Eng. *12*, 271–283.

Eisenstein, M., and Katchalski-Katzir, E. (2004). C. R. Biol. *327*, 409–420.

Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2003). Proteins *52*, 113–117.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., and Baker, D. (2003). J. Mol. Biol. *331*, 281–299.

Grünberg, R., Leckner, J., and Nilges, M. (2004). Structure *12*, this issue, 2125–2136.

Inbar, Y., Benyamini, H., Nussinov, R., and Wolfson, H.J. (2003). Bioinformatics *19* (*Suppl. 1*), i158–i168.

Janin, J. (1995). Prog. Biophys. Mol. Biol. *64*, 145–166.

Kimura, S.R., Brower, R.C., Vajda, S., and Camacho, C.J. (2001). Biophys. J. *80*, 635–642.

Koshland, D. (1958). Proc. Natl. Acad. Sci. USA *44*, 98–104.

Kumar, S., Ma, B., Tsai, C.J., Sinha, N., and Nussinov, R. (2000). Protein Sci. *9*, 10–19.

Lo Conte, L., Chothia, C., and Janin, J. (1999). J. Mol. Biol. *285*, 2177–2198.

Minton, A.P. (2000). Curr. Opin. Struct. Biol. *10*, 34–39.

Monod, J., Wyman, J., and Changeux, J.P. (1965). J. Mol. Biol. *12*, 88–118.

Vajda, S., Sippl, M., and Novotny, J. (1997). Curr. Opin. Struct. Biol. *7*, 222–228.

Zacharias, M. (2003). Protein Sci. *12*, 1271–1282.

Structure, Vol. 12, December, 2004, DOI 10.1016/j.str.2004.11.005

Integrin Activation In Vivo and In Silico

A new computational study by Jin et al. (2004, this issue of *Structure*) tests the hypothesis that mechanical force induces the conformational changes leading to the activation of integrins.

Cells are glued to their surroundings through a family of transmembrane receptor proteins known as integrins. The growth, movement, and survival of cells are all dependent on bidirectional signals relayed by integrins across the membrane. Each integrin consists of two noncovalently associated heterogeneous subunits: α and β. In mammalian cells, eighteen α and eight β subunits form 24 different types of integrins, which selectively bind to extracellular matrix proteins such as collagen, fibronectin, and adhesion proteins on the surfaces of other cells. To be capable of binding to their various ligands, integrins must be activated in response to both extracellular and intracellular signals. One such signal is the mechanical force exerted by the cytoskeleton and transmitted through a mechanical linkage that couples the cytoskeleton to integrins.

It is known that integrin activation involves conformational changes (Hynes, 2002). Compelling structural evidence has become available since the ligand-mimetic bound (active) and unbound (inactive) forms of the integrin α_M I domain were crystallized almost a decade ago by Liddington, Arnaout, and coworkers (Lee et al., 1995). Half of the integrin α subunits have a homologous I domain inserted at the top of the integrin headpiece. The first I domain structures revealed that the 200 amino acids of the protein form a Rossman fold, which consists of a single mostly parallel β-sheet surrounded by seven α-helices. The major ligand binding site is located at the top of the domain and termed the metal-ion-dependent adhesion site (MIDAS), for it acquires a divalent metal ion recognized by the ligands. Compared to the inactive form, the active α_M I domain exhibits a few conformational changes that were attributed to the ligand binding; most notably, the C-terminal α-helix shifts 10 Å towards the tailpiece. The movement is linked to the rearrangement of loops bearing the MIDAS residues. Similar conformational changes have been observed for integrin α2 I domain structures subsequently (Emsley et al., 2000), confirming the notion that the C-terminal helix shift is a key feature of the activated integrin I domains. By introducing disulfide bridges that prevent the movement of the C-terminal helix, Springer and colleagues have successfully locked integrin αL in states with high affinity (active), low affinity (inactive), or intermediate affinity (Lu et al., 2001; Shimaoka et al., 2003).

The observed conformational changes of α I domains led to the hypothesis that the activation of integrin can be regulated by stretching the C-terminal helix. Steered