Identification of protein assemblies is at the heart of functional genomics and drug discovery.

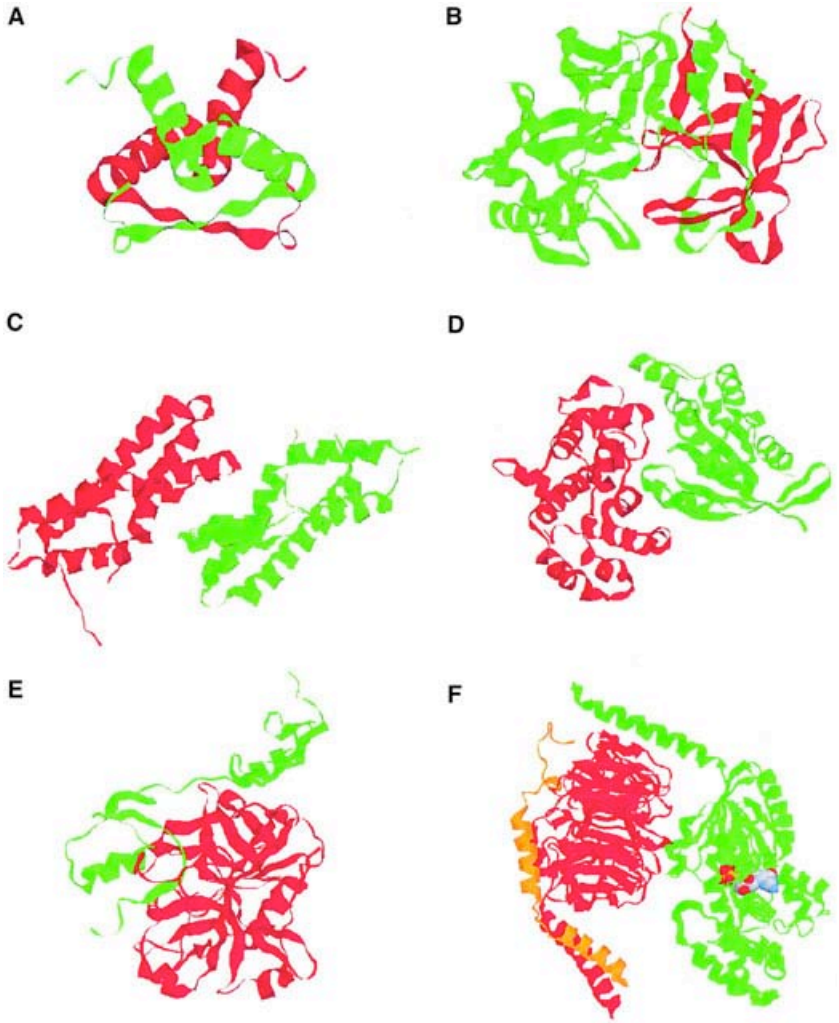

Figure 1. Examples of different types of protein–protein interactions : (A) obligate homodimer, P22 Arc repressor; (B) obligate heterodimer, human cathepsin D that consists of a non-homologous light (red) and heavy (green) chain; (C) non-obligate homodimer, sperm lysin; (D) non-obligate heterodimer, RhoA (green) and RhoGAP (red) signalling complex; (E) non-obligate permanent heterodimer, thrombin (red) and rodniin inhibitor (green); (F) non-obligate transient heterotrimer, bovine G protein, i.e., the interaction between Ga (green) and Gb (red, orange) is transient. The proteins in an obligate interaction are not found as stable structures on their own *in vivo*.

.

Published in: Irene M.A. Nooren; Janet M. Thornton. *EMBO J*. (2003) **22**, 3486-3492.

Prediction of protein assemblies is crucial for understanding cellular organization.

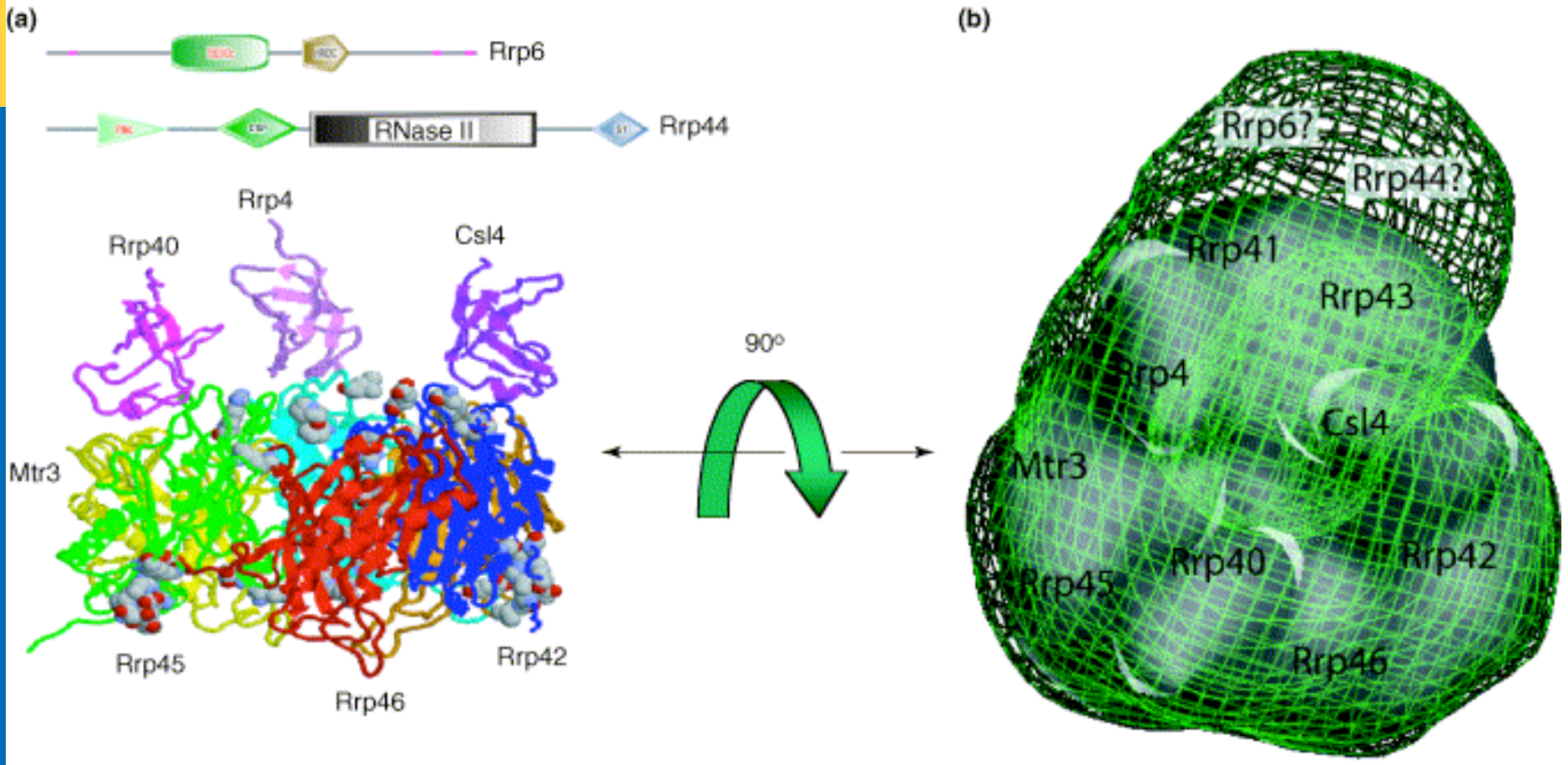

Figure 3. Putative structure through modeling and low-resolution EM. **(a)** Exosome subunits. The top of the panel shows the domain organization of two subunits present in the complex, but lacking any detectable similarity to known three-dimensional structures. The model for the nine other subunits (bottom) was constructed by predicting binary interactions using InterPReTS and building models based on a homologous complex structure using comparative modeling. **(b)** EM density map (green mesh) with the best fit of the model shown as a gray surface and the predicted locations of the subunits labeled. The question marks indicate those subunits for which no structures could be modeled

.

Published in: Robert B. Russell *et al*. *Curr Opin Struct Biol.* **2004**, 14, 313-324..

Proteins are flexible.

- A large fraction of cellular proteins are thought to be natively disordered, or unstable in solution.
- The structures of disordered proteins are not necessarily random. Rather, the disordered state has a significant residual structure. In the disordered state, a protein exists in an ensemble of conformers.
- In many cases, these regions constitute only certain parts or domains of the whole protein.
- While disordered on their own, their native conformation is stabilized upon binding.
- The global fold of disordered proteins does not necessarily change upon binding to different partners; however, local conformational variability can be observed, thereby complicating predictions of protein interactions.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Proteins are flexible.

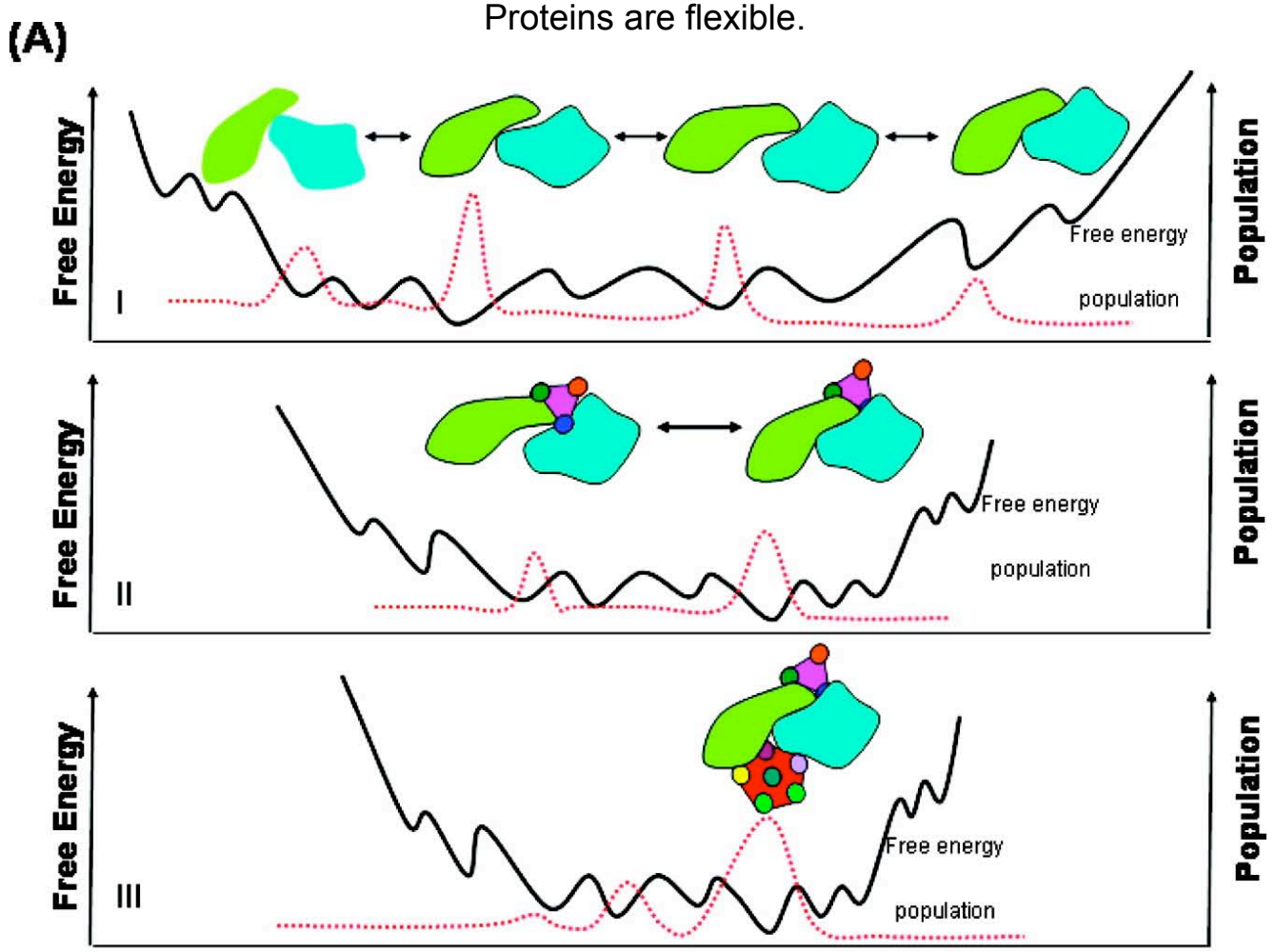

Figure 2 (A) The free energy landscape of a protein may change upon binding to another protein. Binding may induce a shift in the distribution of the populations of the conformational states of the protein; consequently, the relative population of the conformer with an altered binding site shape at another location on the protein surface may increase. The solid black line refers to the free energy landscape, and the dashed red line refers to the relative populations. (I) Distribution of the substates of the protein conformations, presenting several binding possibilities. (II) When a ligand binds at the first binding site, it shifts the conformational energy landscape and the distribution of the populations to favor selective binding at a second, allosteric site. (III) The final dominant conformer recognizes both ligands.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Proteins are flexible.

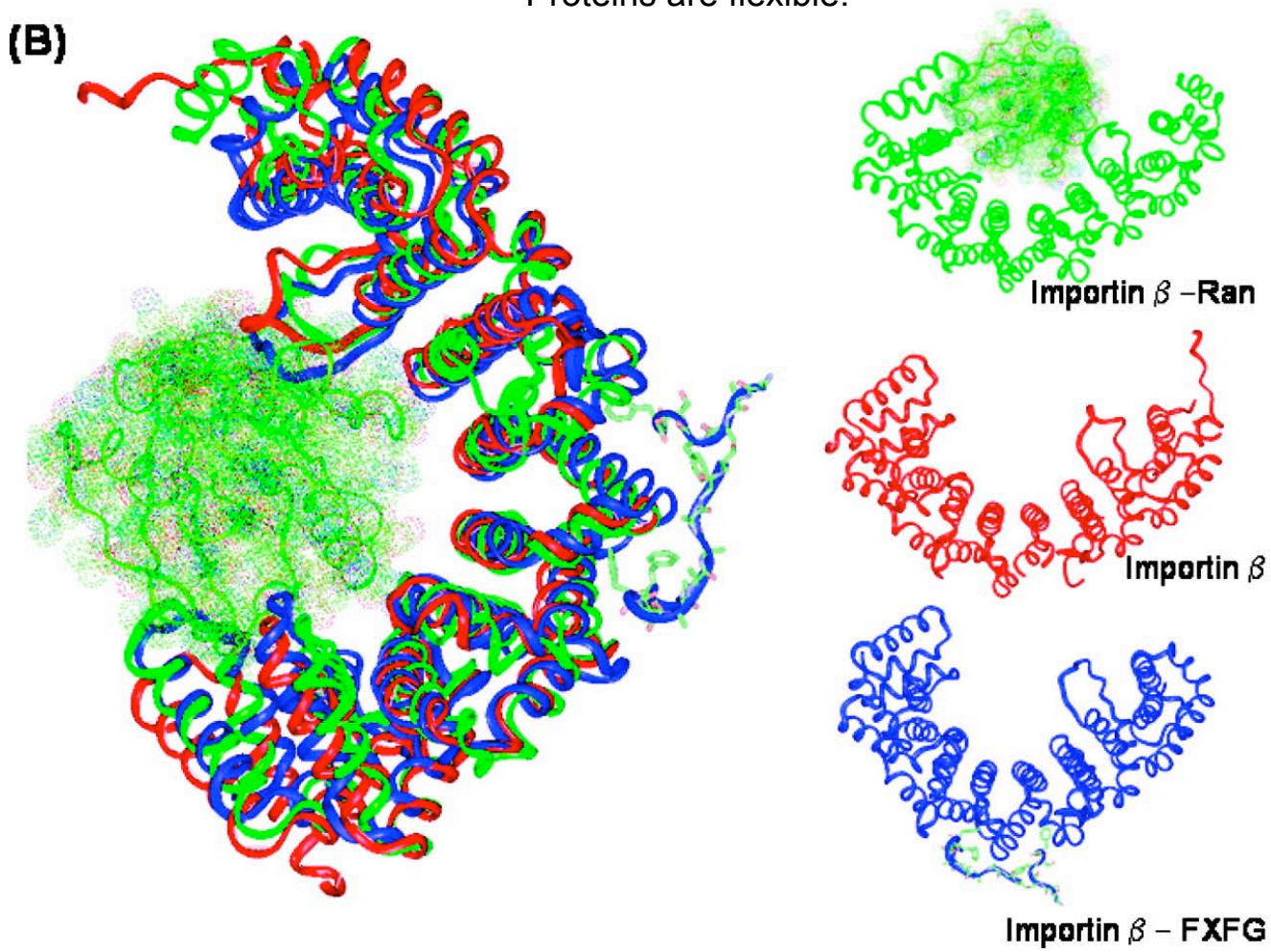

Figure 2 (B) Conformational variability is very important for importin to mediate nucleo-cytoplasmic transportation. Shown here are the superimposition (left panel) of three crystal structures of importin in the free state (red ribbon, left panel, PDB ID: 1gcj), bound to RanGTP (green ribbon, left panel, PDB ID: 1ibr; RanGTP is represented by ribbon and surface dots), and bound to nucleoporin (blue ribbon, left panel, PDB ID: 1f59). The bound/unbound conformational states are coupled with the importin functions of cargo binding and release by RanGTP binding. The importin conformations in the three crystal structures differ significantly in their binding sites with an overall rmsd around 3.5 . In solution, SAXS revealed much larger conformational variations.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Proteins are flexible.

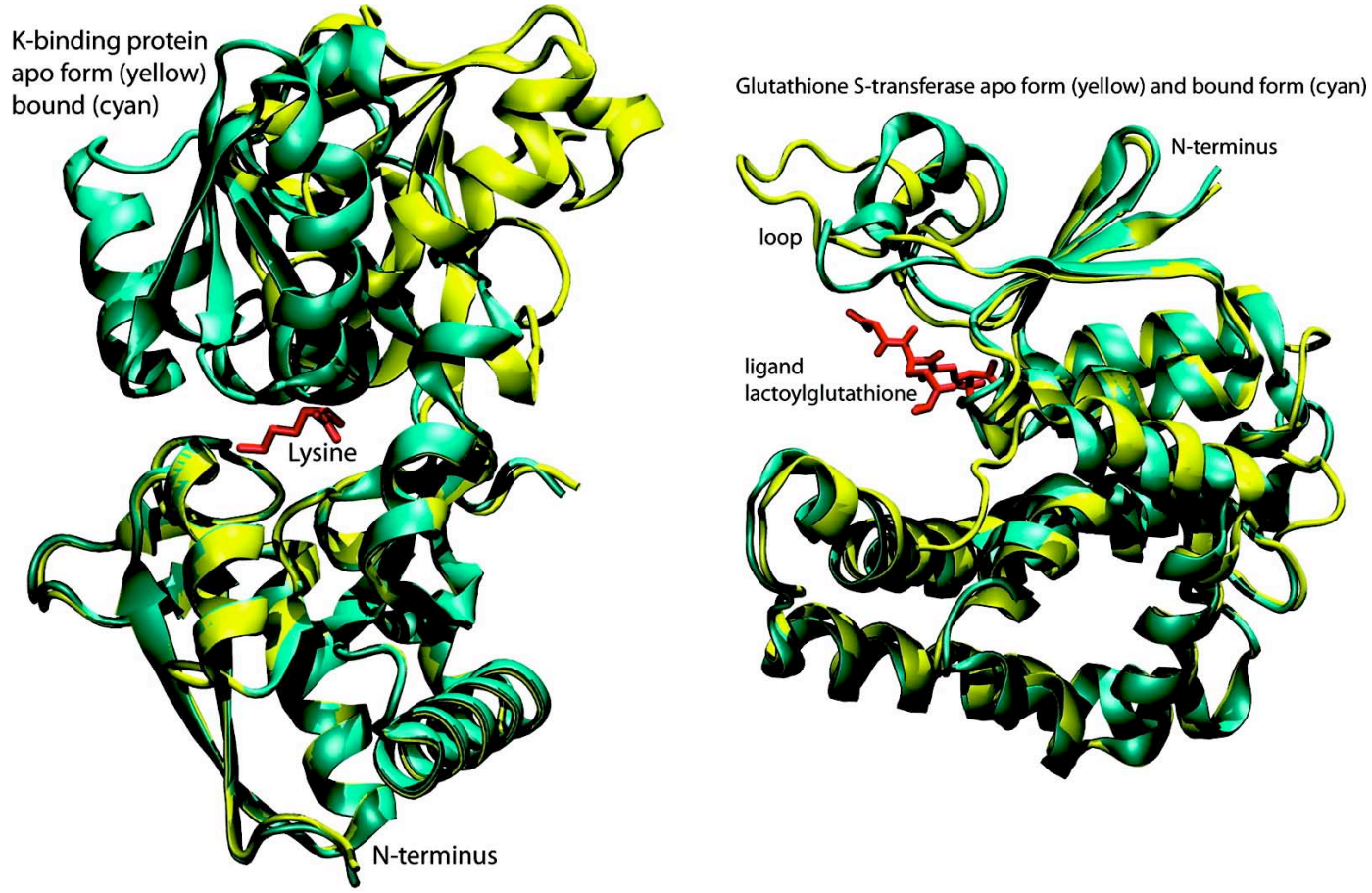

Figure 3 Comparisons of proteins in bound, complexed states versus in the free (apo) states. (A) The conformational changes undertaken by the K-binding protein (PDB IDs: 2lao (yellow) and 1lst (cyan)). The free structure (yellow) closes up and becomes stabilized when it is bound (cyan structure) to its ligand. The ligand, shown in red, belongs to the cyan structure. This is a domain motion example. (B) Glutathione S-transferase-I in free and bound forms (PDB IDs: 1aw9 (shown in cyan) and 1axd (yellow), respectively). The ligand introduces a conformational change in the loop.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Proteins interact through their surfaces.

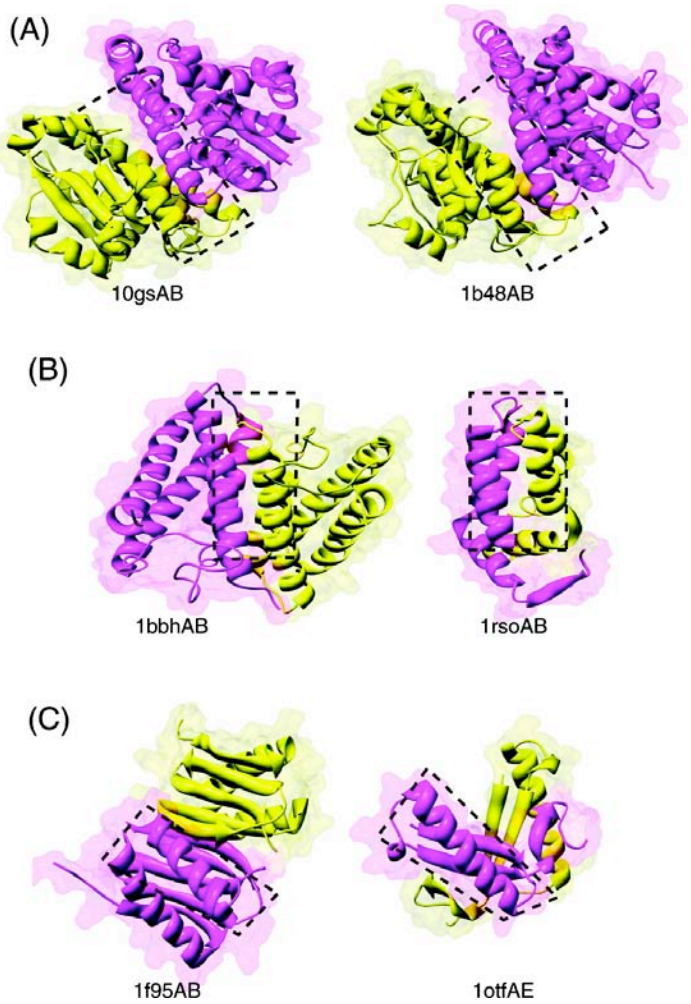

Figure 4 Several examples of crystal structures of binary protein complexes. The interfaces are highlighted with boxes. In part A, the two glutathione S-transferase complexes (PDB IDs: 10gs and 1b48) are homologous; they use similar interfaces to bind each other. In part B, the two complexes, cytochrome C and neuropeptide/membrane protein (PDB IDs: 1bbh and 1rso) are not related evolutionarily, yet the interface architecture is similar. Part C represents two complexes (dynein light chain 8 (PDB ID: 1f95AB) and 4-oxalocrotonate tautomerase (PDB ID: 1otfAE)) where only one side of the interface has similar architectures, the accompanying sides are unrelated. The similar side belongs to the magenta chains.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Protein binding sites are clusters of “hot spots”.

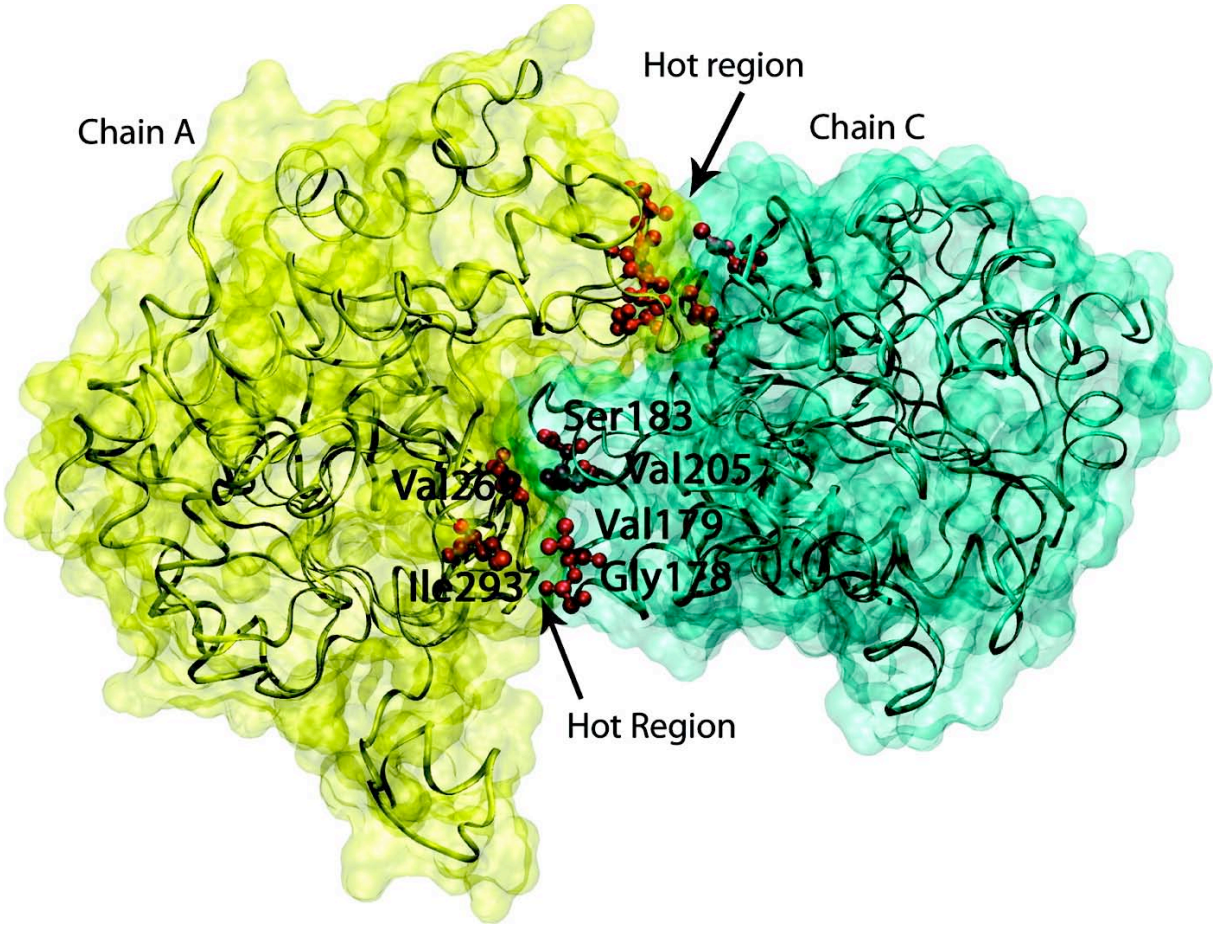

Figure 5 Crystal structure of a complex displaying the hot regions between two M chains of the human muscle L-lactate dehydrogenase (PDB ID: 1i10). Two interacting chains are shown in yellow and cyan. The hot spots (red), shown in ball and stick representation, are residues whose substitution by Ala leads to a significant ($\Delta\Delta G \geq 2$ kcal/mol) drop in the binding free energy (Clackson & Wells; *Science* **1995** 267, 383). There are two hot regions in this interface of the homodimer. The figure illustrates that hot spots are in contact with each other and form a network of interactions forming *hot regions*.

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.

Different protein partners may have similar binding motifs.

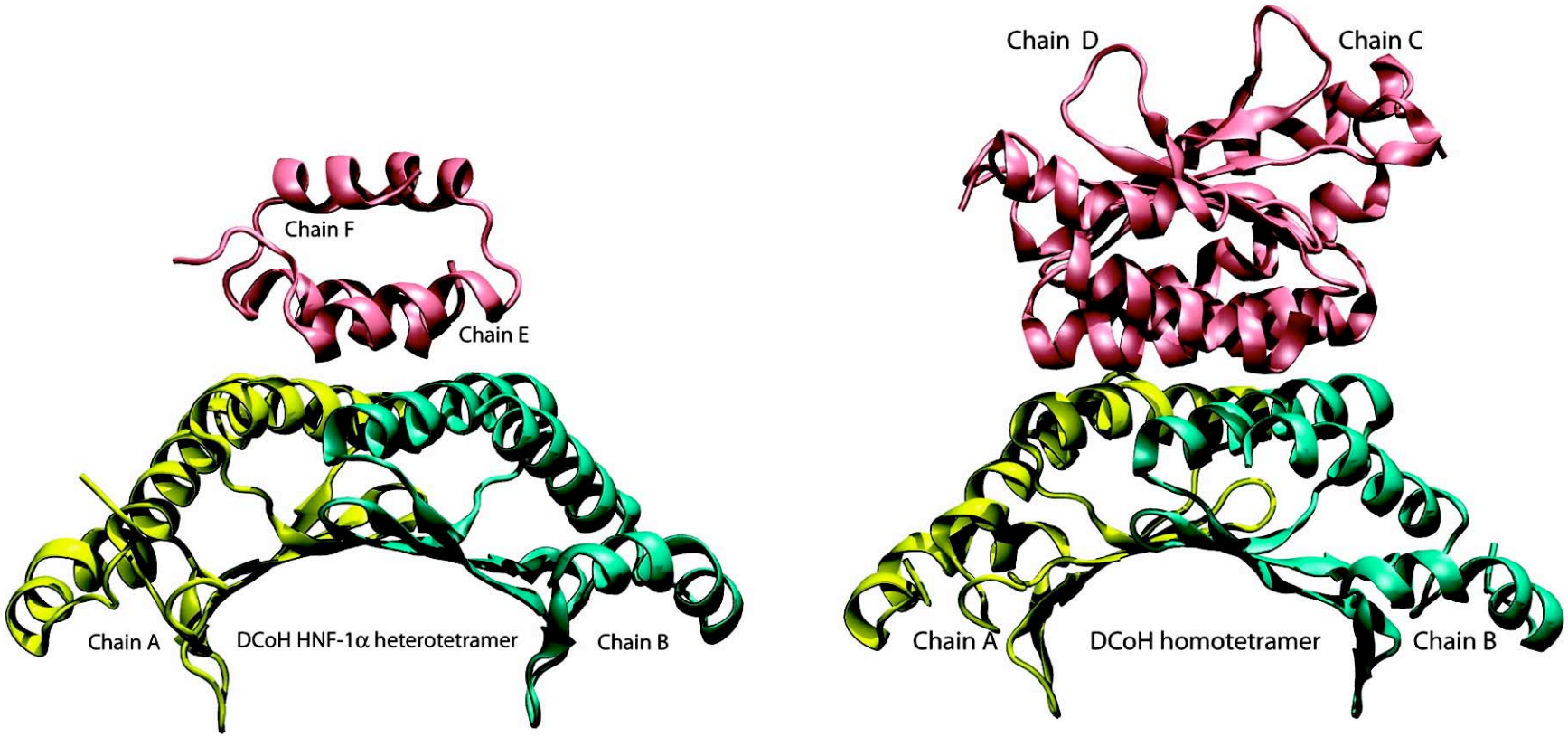

Figure 9 Example of multiple proteins binding at the same site on the protein surface, dimerization cofactor of hepatocyte nuclear factor (DCoH). DCoH serves as an enzyme and a transcription coactivator. The left figure is the crystal structure of hepatocyte nuclear factor dimerization domain, HNF-1, bound to a DCoH dimer (PDB ID: 1F93, Chains A, B of DCoH, and Chains E, F of HNF-1). In order to act as a coactivator, DCoH binds to HNF 1. The figure on the right displays the enzymatic form of the protein DCoH forming dimers of dimers (shown Chains A, B, C, and D, PDB ID: 1DCH).

Published in: Ozlem Keskin; Attila Gursoy; Buyong Ma; Ruth Nussinov; *Chem. Rev.* **2008,** 108, 1225-1244.