# Structural bioinformatics and molecular recognition

Department of
Plant Sciences

- Prof. Marvin Edelman
- Dr. Vladimir Sobolev

Ronen Levy

☎ 972 8 934 3301
FAX 972 8 946 9124
@ marvin.edelman@weizmann.ac.il
@ vladimir.sobolev@weizmann.ac.il
ligin.weizmann.ac.il

**Predicting metal binding sites from apo protein structures and translated gene sequences**

Metal ions are crucial for protein function. They participate in enzyme catalysis, play regulatory roles, and help maintain protein structure. Current tools for predicting metal-protein interactions are based on proteins crystallized with their metal ions present. However, a majority of resolved structures are free of metal ions. Moreover, metal binding is a dynamic process, often involving conformational rearrangement of the binding pocket. Thus, effective predictions need to be based on the structure of the apo state. Our group has developed an approach that identifies transition metal-binding sites in apo forms with a resulting selectivity >95%. Applying the approach to apo forms in the Protein Data Bank and structural genomics initiative identifies a large number (>1200) of previously unknown, putative metal-binding sites, and their amino acid residues, in some cases providing a first clue to the function of the protein. Fig. 1 presents an example from our CHED server for predicting soft metal binding sites in proteins. A major feature of our algorithm is that, in the first step (geometric search), structural rearrangements upon metal binding are taken into account. Filtration by machine learning is then applied to increase selectivity. The server produces a graphical presentation of the predicted binding site(s). PDB or user-generated structures can be submitted.

We also find that structures obtained by modeling translated gene sequences are sufficient for effective prediction of metal binding-sites. The basis for this is the major overlap already achieved between structural and linear database space (> 40%) and the minor extent by which side chain modeling reduces predictive accuracy (~ 5%). The procedure involves: inputting a translated gene sequence ("target"); seeking a homologous PDB sequence ("template"); structurally modeling target side-chains using the template backbone; outputting the predicted metal binding site using the CHED algorithm. Analysis of the approach finds selectivity to be uniformly high (~85-90%) irrespective of the level of sequence homology between template and target (Levy, in progress).

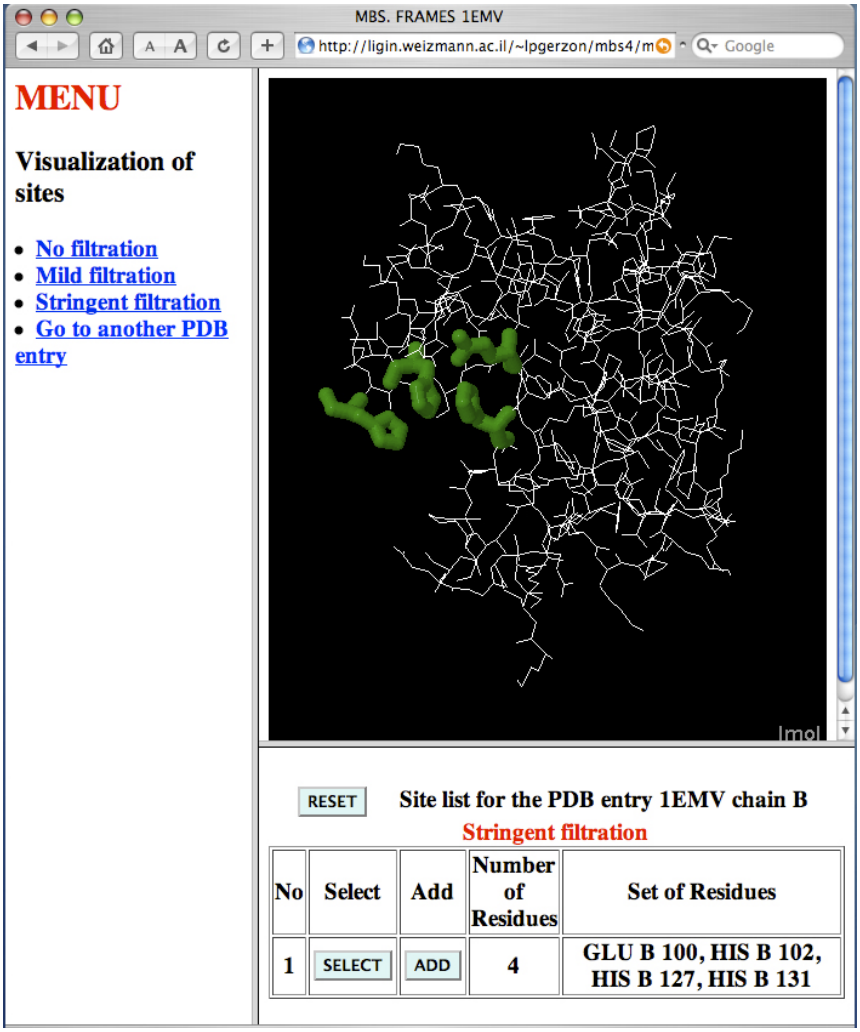


***Fig. 1*** *Output of CHED server (http://ligin.weizmann.ac.il/ched) for the apo form of the Colicin E9 Dnase domain. A metal binding site is predicted containing four residues. Analysis of the holo form of the protein (PDB entry 1fr2) confirms correctness of the binding site prediction.*

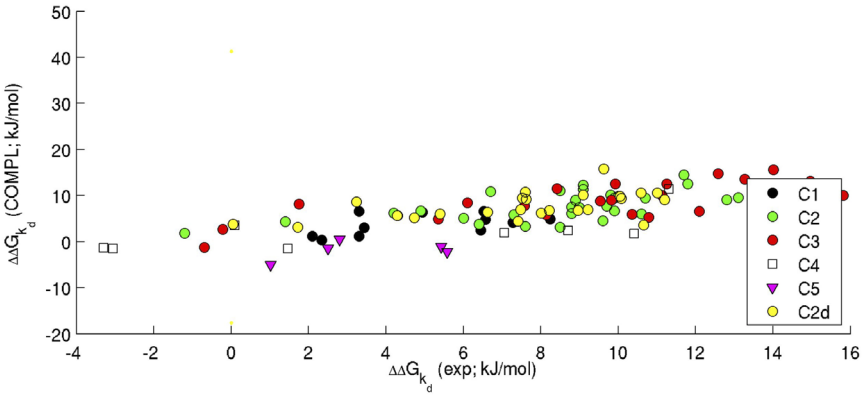


***Fig. 2*** *Calculated ΔΔGs compared to experimental ΔΔGs (based on surface complementarity) for 92 mutants in the TEM1-BLIP complex. Computational and biochemical studies of Dana Reichmann and others in Gideon Schreiber's group showed that the protein-protein interface can be considered as a set of mostly independent modules. Points on the graph are color-coded according to the module from which a given mutation originated. Note that good correlation is achieved for all modules.*

**Predicting stability of protein-protein complexes**

The ability to modulate protein-protein interactions and to engineer proteins with new functions and specificities is one of the goals of protein science, with implications for practical application. This requires quantitative understanding of the energetics of protein-protein binding and mechanism of specificity. We developed an approach for computational redesign of protein-protein interfaces in collaboration with Gideon Schreiber's group, combining natural template fragments from resolved structures with a new scoring function. An algorithm for predicting the change in stability of protein-protein complexes upon interface mutation was developed (Vladimir Potapov, PhD thesis, Department of Plant Sciences).

The scoring function was based on surface complementarity. The assumptions are that change in protein complex stability is proportional to change in contact surface areas, and that proper weights can be found based on experimentally determined stability changes. Protein atoms were divided into eight classes (hydrophilic, donor, acceptor, aromatic, hydrophobic and three neutral classes). To avoid overfitting, atom-atom contact areas were grouped according their physico-chemical properties. Contact solvent areas, and electrostatic and residue volume terms were also included while an entropic term was not, as it had negligible contribution. The areas were weighted based on experimentally determined stability changes for a set of over 90 interface mutations in the TEM1-BLIP complex (PDB entry 1JTG, 1.7 Å resolution). The overall correlation between theoretical and experimental values is 0.72 (Fig. 2), with the leave-one-out cross validation, 0.64.

**Selected publications**


Eyal E., Najmanovich R., Edelman M., Sobolev V. (2003) Protein side chain rearrangement in regions of point mutations. Proteins, 50, 272-282.

McConkey B.J., Sobolev V., Edelman M. (2003) Discrimination of native protein structures using atom-atom contact scoring. Proc. Natl. Acad. Sci. USA, 100, 3215-3220.

Yakobson E., Eisenberg S., Isakson R., Halle D., Levy-Lahad E., Catane R., Safro M., Sobolev V., Huot T., Peters G., Ruiz A., Malvehy J., Puig S., Chompret A., Avril M-F., Shafir R., Peretz H., Paillerets B.B. (2003) A single Mediterranean, possibly jewish, origin for the Val59Gly CDKN2A mutation in four melanoma-prone families. Eur. J. Hum. Gen., 11, 288-296.

Rosenberg N., Yatuv R., Sobolev V., Peretz H., Zivelin A., Seligsohn U. (2003) Major mutations in CALF-1 and CALF-2 domains of glycoprotein IIB in patients with Glanzmann thrombasthenia enable GPIIB/IIIA complex formation but abolish its transport from the endoplasmic reticulum to the Golgi apparatus. Blood, 101, 4808-4815.

Kuttner J., Sobolev V., Raskind A., Edelman M. (2003) A consensus binding structure for adenine at the atomic level permits searching for the ligand site in a wide spectrum of adenine-containing complexes. Proteins, 52, 400-411.

Eyal E., Najmanovich R., McConkey B.J., Edelman M., Sobolev V. (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. J. Comp. Chem, 25, 712-724.

Potapov V., Sobolev V., Edelman M., Kister A., Gelfand I. (2004) Protein-protein recognition: Juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. J. Mol. Biol., 342, 665-679.

Babor M., Greenblatt H.M., Edelman M., Sobolev V. (2005) Flexibility of metal binding sites in proteins on a database scale. Proteins, 59, 221-230.

Sobolev V., Eyal E., Gerzon S., Potapov V., Babor M., Prilusky J., Edelman M. (2005) SPACE: A suite of tools for protein Structure Prediction and Analysis based on Complementarity and Environment. Nucl. Acids Res., 33, W39-W43.

Eyal E., Gerzon S., Potapov V., Edelman M., Sobolev V. (2005) The limit of accuracy of protein modeling: influence of crystal packing on protein structure. J. Mol. Biol., 351, 431-442.

Eyal E., Frenkel-Morgenstern M., Sobolev V., Pietrokovski S. (2007) A pair-to-pair amino acids substitution matrix and its applications for protein structure prediction. Proteins, 67, 142-153.

Knox A.J.S., Meegan M.J., Sobolev V., Frost D., Zisterer D.M., Williams D.C., Lloyd D.G. (2007) Target specific virtual screening: Optimization of an estrogen receptor screening platform. J. Med. Chem., 50, 5301-5310.

Babor M., Gerzon S., Raveh B., Sobolev V., Edelman M. (2008) Prediction of transition metal-binding sites from apo protein structures Proteins, 70, 208-217.

***Acknowledgements***

M.E. is incumbent of the Sir Siegmund Warburg Chair in Agricultural Molecular Biology

V.S. is a Camea Fellow of the Ministry of Absorption, Division of the Absorption of Scientists

M.E. and V.S. acknowledge the support of the Ministry of Science

***INTERNAL support***

The Avron-Willstatter Center for Research in Photosynthesis