**Miriam Eisenstein**

**Ephraim Katchalski-Katzir**

Efrat Ben-Zeev
Alex Berchanski
Alex Heifetz
Boaz Shapira

# Development and application of theoretical tools for predicting the structures of molecular complexes

## Depts. of Chemical Services & Biological Chemistry

**Tel.** 972 8 934 3031 **Fax.** 972 8 934 4136
**E-mail:** miriam.eisenstein@weizmann.ac.il

Living organisms rely on the specific recognition of pairs of molecules in practically every biological process. Hence, the importance of understanding molecular recognition and determining the structures of molecular complexes cannot be overestimated. Our group's goal is to develop and implement reliable theoretical tools for predicting the structures of molecular complexes (docking procedures). Such tools must be able to deal with molecules whose activity is not fully understood and with modeled structures, where the accuracy may be limited. Therefore, incorporation of all the available knowledge regarding intermolecular interfaces is important. An adequate representation and quantification of this information must be formulated, such that reduces the sensitivity of the prediction procedure to structural errors.

Geometric recognition (shape complementarity), electrostatic complementarity and hydrophobic complementarity are considered by our algorithm as well as experimental data from biological, biochemical and bioinformatics studies. The algorithm, implemented in a computer program named MolFit, uses correlation techniques to assess the shape and chemical complementarity of molecular surfaces. In MolFit the common three-dimensional (3D) atomic representation of the molecules to be docked (denoted A and B) is replaced by a 3D grid representation. The grids are then correlated using discrete Fourier transformations (Katchalsky-Katzir et al., 1992; Eisenstein et al., 1997).

Each grid point in the representation is a complex number, such that the real part contains the geometric descriptor of the molecular surface and the imaginary part includes an additional descriptor, derived from the electrostatic or hydrophobic character of the surface or from experimental data. The real part of grid points outside the molecule are given the value 0, whereas points on the surface of the molecule are given the value g. The parameter g is the geometric descriptor of the molecular surface and it is usually 1. Grid points in the interior of the molecule are given either a negative value (-15), or a positive value (+1), for molecules A and B, respectively. The imaginary part can contain either an electrostatic descriptor, derived from the electrostatic potential of the molecule, or a hydrophobic descriptor based on the character of the residues nearby, or a weight parameter, which reflects the probability of a given residue to be involved in binding, according to experimental data.

We apply our algorithm to a large selection of known protein-protein complexes and oligomers starting from either the bound structures of the component molecules (i.e. as they are in the complex) or their unbound structures (i.e. separately determined). Geometric docking successfully reassembles know structures and often, unknown structures are also correctly predicted (Strynadka et al., 1996; Eisenstein et al., 1998). Yet, in general, geometric docking of unbound structures is less successful than docking of bound structures.

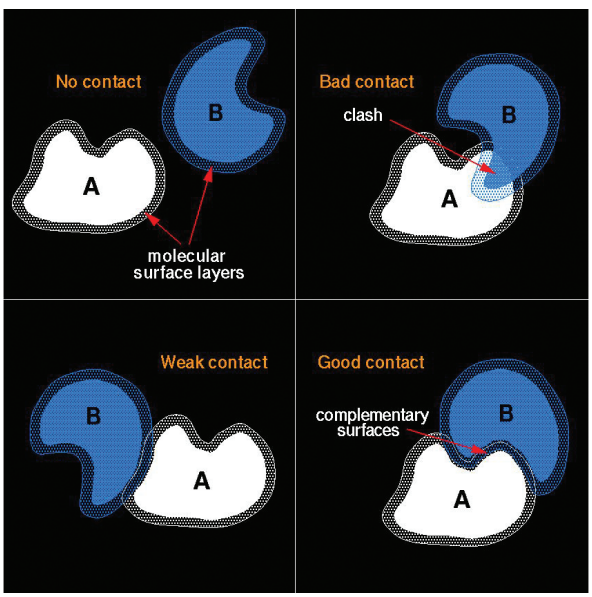


***Fig. 1*** *A section through the grid representation of two molecules, where the surface layer is distinguished from the interior of the molecule. The figure presents different possibilities in docking: no contact, weak contact, interpenetration and a good contact.*

Geometric-electrostatic docking is tested on 17 systems and it significantly improves the results of geometric docking.

Based on a detailed analysis of the results we derive several 'good electrostatic docking rules', which specify when is geometric-electrostatic docking better than geometric docking (Fig. 2). Geometric-hydrophobic docking is currently being applied to the same selection of structures.

The structures of oligomers are predicted by combining MolFit with different oligomer forming algorithms, in which symmetry constraints are incorporated. For example: Homo-tetrameric oligomers are formed by combining homo-dimers (identified by MolFit) that comply with the symmetry requirements of the tetramer, either a dimer of dimers (D2 symmetry) or a planar tetramer (C4 symmetry). Geometric docking in combination with a dimer-of-dimers forming algorithm successfully predicts the structures of most oligomers, starting from either bound or unbound subunits (22 tetramers are tested). Currently the effect of hydrophobic- or electrostatic-complementarity on oligomer formation is studied.

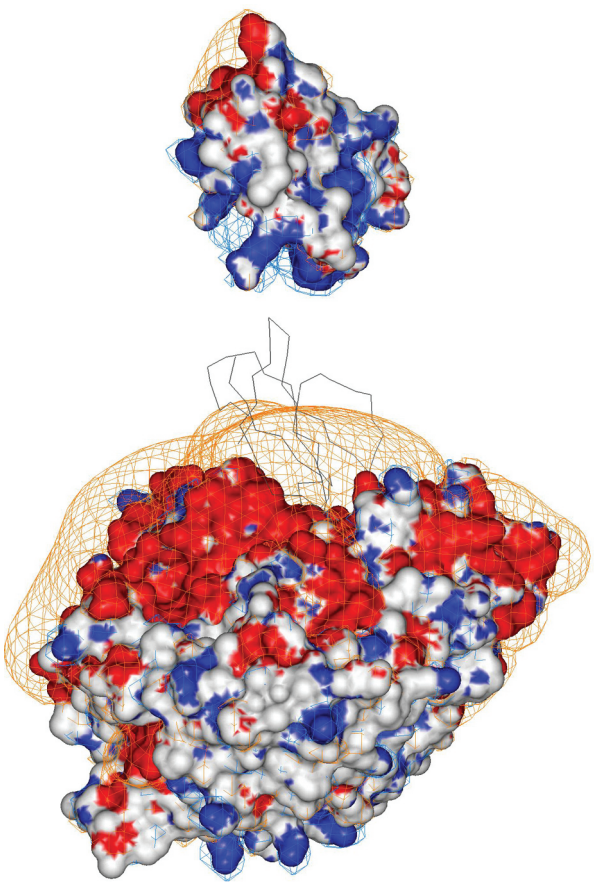

***Fig. 2*** *The electrostatic patches on the surfaces of acetylcholinesterase and fasciculine2, and two electrostatic potential contours at 3kT/e (blue) and -3kT/e (orange), illustrating the dipolar character of both molecules and the protrusion of the potential into the solvent.*

Sequence data contain valuable structural information. Functionally important residues can be extracted from the conservation patterns in homologous proteins and mapped onto the protein surface to generate clusters identifying functional interfaces (Lichtarge et al., 1996). Some of the clusters are plausible recognition sites, which are more conserved than the rest of the molecular surface. Up weighing a given cluster directs the geometric docking procedure toward formation of complexes where that cluster is at the interface. The predictions can be tested against available experimental data (e.g. the effect of point mutations) and suggest new experiments. Currently the method is tested on several proteins within the apoptosis cascade.

***Selected Publications***


Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, and Vakser IA (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. USA 89, 2195-2199.

Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MN, (1996). Molecular docking programs successfully predict the binding of a lactamase inhibitory protein to TEM-1. Nat. Struct. Biol. 3, 233-239.

Lichtarge O, Bourne HR et al, (1996).An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. 257, 342-358.

Eisenstein M, Shariv I, Koren G, Friesem AA, Katchalski-Katzir E (1997). Modeling supra-molecular helices: extension of the molecular surface recognition algorithm and application to the protein coat of the tobacco mosaic virus. J. Mol. Biol. 266, 135-143.

Eisenstein M and Katchalski-Katzir E, (1998). Geometric recognition as a tool for predicting structures of molecular complexes. LIPS 5, 365-369.

Heifetz A, Katchalsi-Katzir E and Eisenstein M, (2001). Electrostatics in protein-protein docking. Protein Science, (in press).