MIAX: A Novel System for Assessment of Macromolecular Interaction in Condensed Phases. 1) Description of the Interaction Model and Simulation Algorithm

Carlos Adriel Del Carpio Munoz carlos@translell.eco.tut.ac.jp

Atsushi Yoshimori yosimori@translell.eco.tut.ac.jp

Department of Ecological Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 441-8580, Japan

Abstract

We describe a novel computer system directed to evaluate protein complex formation in a liquid environment. The relevant feature of the system is a potential function expressing the main thermodynamic and kinetic factors leading to protein interaction in solution. The protein interaction model expresses the interaction energy as basically composed of three forces: electrostatic (hydrogen bond), van der Waals, and hydrophobic. The latter is defined in function of the forces that the solvent molecules exert on the surface of the complex, and the van der Waals forces between the monomers and the solvent.

The interaction model implemented in the system has proven a high discrimination ability between different protein dockings, scoring high those close to the observed crystal structures. These results have led to the establishment of the basic principles underlying protein interaction, which constitutes the main way of expression of the biological function of these macromolecules.

1 Introduction

With the advent of the post-genome era, and with enormous amounts of data product of the translation of the genome code to be analyzed, one of the topics that remains to be unveiled is that involving the rules and fundamentals governing protein interaction, and macromolecular interaction in general.

Protein interaction is the expression of biological function of these macromolecules and is intimately related to molecular activity and specificity. Meticulous understanding of this phenomenon would pave the path to the prediction of the geometry at interaction of the components forming the complex when only information of the isolated monomers and the conditions under which they interact is known a priori, yielding in this way insights into the underlying processes of molecular recognition, Moreover, this understanding would lead to establish the rules within the genetic code for encoding complementary proteins in the organism.

Many of the systems developed up to date focus the problem from the geometrical and electrophysical complementarity of the monomers constituting the complex, being a common characteristic the fact that they are based on rigid body docking of the complex components with the conformation adopted at interaction [2,7,10,16,18] or at their native state [8,9].

Results derived by these systems consist in a series of predictions scored on the base of shape and electrostatic complementarity. The method described here is oriented to the prediction of the optimal structure of a complex composed of two or more monomers at interaction taking into account the environment and conditions under which the complex formation occurs. In order to accomplish this objective, we propose a model that accounts for hydrophobic interactions by means of a combined term including the cavity formation, surface tension and van der Waals interaction of the atoms

constituting the molecules and also the surrounding solvent. Electrostatic interactions among intermolecular atoms and the surrounding solvent are calculated by a generalized Born Equation, while inter atomic interactions within the monomers are considered by means of a force field that expresses the intramolecular forces leading to conformational changes in the monomers as the complex formation proceeds. Furthermore, hydrogen bonding is also explicitly considered . Optimization using his model leads to the prediction of the lowest energy complex configuration. We have carried out all these studies and implemented a computer system that we have come to name MIAX(*M*acro-molecular *I*nteraction *A*ssessment system *X*) , and we report the results in two parts being this the first one, in which we describe the protein interaction model in a condensed phase.

2 Protein Interaction Model

Protein interaction is a phenomenon by which these macromolecules express their function, and they are the cornerstone of processes controlling almost all the regulatory functions in an organism. Processes such as signal transmission, antigen-antibody interaction, molecular recognition -to cite some of them- are controlled by this phenomenon and consequently its understanding is of the utmost relevance in many fields of life and the medical sciences. Although the association of several proteins can occur through covalent bonding like sulfide bonding among cysteinic amino acids composing the interacting bio molecules, leading in such cases to non reversible complexes, experimental observation of several processes controlled by protein interaction has led to the conclusion that protein association and binding is similar to an enzyme substrate interaction but with the important distinction that unlike the latter, the former leads to a reversible process. The noncovalent interactions that form the basis of protein-protein binding include hydrogen bonds, ionic bonds, hydrophobic interactions, and van der Waals interactions among the atoms constituting the interacting monomers and with the atoms of the surrounding environment. Because of the weakness of these types of intermolecular interactions (further weakened in a liquid environment) as compared to a covalent bond, a large number of them is required to form a relatively stable complex. Moreover, strong protein-protein interactions are observed at very small distances (less than 10 Åof separation). This fact reflects the high degree of specificity characteristic of the interactions between this type of bio-molecules.

Considering all these factors, we introduce a new potential function constituted by interaction terms among the protein monomers as well as terms accounting for interaction with the surrounding solvent.

The model can be illustrated simply as in Fig. 1, where the energies of formation of the solvated complex can be computed by the ideal process represented by the dot arrows.

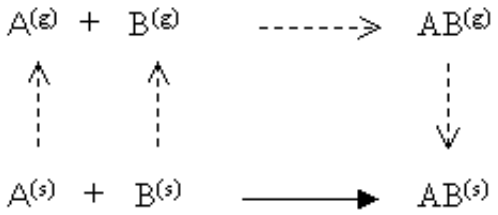

Figure 1: Calculation of the formation free energy for solvated complex AB.

Accordingly the complex in solution $AB^{(s)}$ from the solvated monomers, can be calculated as the

energy of interaction of the monomers in the gas state plus the energy of solvation of the complex $AB^{(g)}$ plus the energies of desolvation of the monomers A and B, that is:

$$\Delta G_{AB^{(s)}} = G^{AB^{(g)}}_{A^{(g)},B^{(g)}} + G^{AB^{(g)}}_{sol} + G^{A^{(s)}}_{desol} + G^{B^{(s)}}_{desol} \tag{1}$$

The first term of the interaction potential energy introduced in the present model is the hydrophobic interaction term(E_{hy}), which expresses the tendency of the monomers to aggregate in solution. This effect is accounted for by calculating the pressure that the solvent exerts on the complex interface with the solvent and the van der Waals interactions of the solvent and the atoms on the solvent accessible surface of the complex. The pressure term can also be regarded as a term related to the energy of cavity formation for the complex within the solvent. The second term takes into account hydrogen bond energies (E_{hb}) explicitly. Similarly, an electrostatic energy term expresses the electrostatic energies among any pair of atoms constituting the complex as well as each complex atom and the solvent molecules (E_{elec}).

The fourth term expresses the changes in torsional energy due to conformational changes within the monomers at interaction (E_{tor}). The fifth term is endowed to the potential function as a correction factor to take into account the desolvation energy of the contact surfaces of the monomers constituting the complex (E_{desol}).

Consequently the potential function introduced here has the form:

$$\Delta G = E_{hy} + E_{hb} + E_{elec} + E_{tor} + E_{desol} \tag{2}$$

The calculation of each term for the complex for a given conformation is described briefly in the following sections.

2.1 The Hydrophobic Interaction Energy

In our computational model of protein-protein interaction in solution, the forces acting on atom i of the complex can be divided in mainly two types : (i) the force field of the complex itself, and (ii) the interaction of the atoms of the monomers forming the complex with the surrounding solvent. The latter is computed as the sum of the potential gradient expressing the van der Waals energy (among intramolecular atoms and with the solvent molecules) and the pressure that the solvent exerts on the surface of atoms laying on the surface of the complex.

The energy resulting from the pressure of the solvent on the complex is computed by evaluating the product $p\Delta$V arising from the solvent pressure on the atoms composing the complex, being ΔV the change in solvent excluded volume due to complex formation.

In scaled particle theory the pressure exerted by the solvent on the wall of a spherical cavity with radius r is given by:

$$p = k_B T \rho G(r) \tag{3}$$

where k_B is the Boltzman constant T the temperature ρ the number density of the solvent and $G(r)$ is the reversible work of cavity formation. Reiss *et al.* [14] suggest that regarding real molecules in the fluid as rigid cores of diameter α, $G(r_c)$ can be approximated by the following expression:

$$G(r_c) = k_0 + k_1 r_c + k_2 r_c^2 + k_3 r_c^3 \tag{4}$$

Thus:

$$E_{px} = \rho k_B T \sum_{i=1}^{N} G(r_c^i) \Delta V_c^i \tag{5}$$

where $G(r_c^i)$is the free energy for the formation of a cavity corresponding to the atom i, where the cavity radius is r_c^i, and the change of the cavity volume due to this atom is ΔV_c^i.

The hydrophobic interaction term proposed in this study is composed of the pressure on the complex surface, and the van der Waals interaction energies between any pair of atoms of the complex as well as the interaction energies of any atom of the complex and the molecules of the solvent surrounding it. The van der Waals interaction between pairs of atoms composing the complex is simply calculated by a Lennard-Jones (12-6) potential function of the form:

$$E_{vdw}^{(p)} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{A_{ij}}{d_{ij}}^{12} - \frac{B_{ij}}{d_{ij}}^{6}\right) \tag{6}$$

To compute the van der Waals interaction energy between the solvent and the complex, we use an empirical expression obtained by correlation of solvation energies for small organic compounds and the solvent accessible surface area SASA. The expression is:

$$E_{vdw}^{(s)} = \sum_{i=1}^{N} a_i SASA_i + b_i \tag{7}$$

Consequently,

$$E_{vdw} = E_{vdw}^{p} + E_{vdw}^{s} \tag{8}$$

And the hydrophobic interaction is:

$$E_{hy} = E_{px} + E_{vdw} \tag{9}$$

2.2 Electrostatic Interaction Energy

To express the electrostatic interaction energy among any pair of atoms constituting the complex as well as the electrostatic energy arising from interaction among solvent molecules and any complex atom we adopted the dielectric continuum-medium approach and reaction field theory [1].

The main postulates of the theory are based on the total electrostatic free energy (G_{el}) of a system of widely separated particles in a medium of dielectric constantthat can be expressed by the sum of Coulomb's law in a dielectric medium and the Born equation.:

$$G_{el} = 332 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{\epsilon} \frac{q_i q_j}{d_{ij}} - 166(1 - \frac{1}{\epsilon}) \sum_{i}^{N} \frac{q_i^2}{\alpha_i} \tag{10}$$

where α_i is the effective Born radius of atom i which we calculate by a new empirical relation obtained by PLS analysis of the main variants thought to influence it and the values obtained by the model of Still [17].

2.3 Hydrogen Bond Energy

Binding specificity and stabilization of bio-molecular complexes are well explained in terms of the free energy attributed to the hydrogen bonding. Analysis of reported complexes reveal a significant number of intermolecular hydrogen bonds within the adequate angular and distance geometrical constraints; being only a few of the donor/acceptor atoms involved in intermolecular binding also capable of forming intramolecular hydrogen bonds [12]. In the present study, hydrogen bonding interaction energy is computed as described by Del Carpio [5].

2.4 The torsional Potential Term

Bio-molecular binding brings about conformational changes in the interacting monomers. The model we have developed here has been endowed with a term to take account of the variation of the internal energy of the interacting monomers. The term considers, however, only torsional changes associated with the side chains of proteins participating in bio-macromolecular binding. The backbone of the

monomers is kept fix. The function to account for torsional energy variation is a well known harmonic function of the torsional angles [5].

2.5 Solvation and desolvation Energies

When two or more monomers in solution interact, solvent molecules ought to be pushed out from the interface of the monomers to form a cavity to accommodate the complex molecule. Moreover several monomer-solvent hydrogen bonds are broken to form new hydrogen bonds among the monomers hydrogen bond donors and acceptors. These factors added to the van der Waals contact energies which are lost with the solvent but compensated to certain extent by contacts between the monomers at complex formation [11,13,19] constitute together with the entropic changes brought up in the solvent the main terms to compute the work to desolvate the complex interface. Solvation energies for proteins have been approximated by a scheme proposed in the work of Eisenberg and McLachlan [6]. Cummings [4] adopt a similar scheme to compute the desolvation energy modifying the atomic solvation parameters introduced by Eisenberg and McLachlan, and express the desolvation energy as:

$$E_{desolv} = \sum_{i=1}^{N} \sigma_i \Delta SASA_i \tag{11}$$

where σ_i are the modified atomic solvation parameters (ASP) for the various atom types considered [4] and $\Delta SASA_i$ is the change in solvent accessible surface area for the N atoms involved in complex formation.

2.6 Computation of Solvent Accessible Surface Area and Excluded Volume

In the present work we have implemented the Richmond algorithm Richmond [15] , in the C language for the solvent accessible surface area, and used the algorithm of Connolly [3] for the computation of the excluded volumes.

3 Interaction Space Mapping

Mapping the configuration space for bimolecular complexes of the nature pursued here signifies dealing with a space of at least six degrees of freedom (5 rotations and one translation) as illustrated in Fig. 2, when the interacting monomers are assumed to be rigid bodies. The algorithm chosen to perform the configurational mapping in the present system is based on a simulated annealing machine, which scans the space by rotating and translating one of the monomers around the other fix at the origin of coordinates as illustrated in Fig. 6. The goal of the optimization is to locate the global minimum of the landscape depicted by the potential function described above. Although a system with many degrees of freedom and with a highly convoluted energy landscape such as the bio-macromolecular complexes treated in this study makes difficult to ascertain the finding of a global minimum, in a first stage, we have performed minimization of rigid bodies to validate the model we propose in the present study.

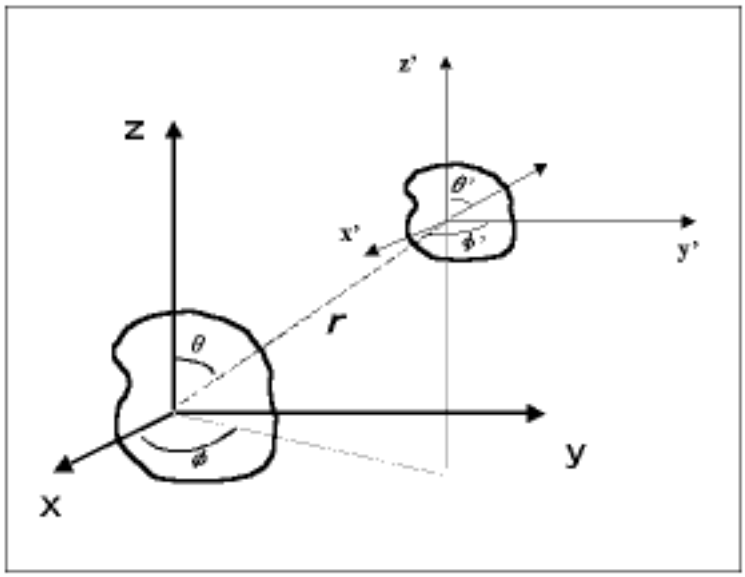

Figure 2: Degrees of Freedom for the docking of two rigid bodies.

4 Results and Discussion

Validation of the model for complex formation we propose here requires the computation of the optimal configuration of the monomers at interaction, and its comparison with experimental data registered in a data base such as the Protein Data Bank (PDB).

We have carried out detailed binding studies of one of such systems, and we report the results obtained so far. The first molecule treated with the system described here was the dimer Uteroglobin, (PDB number code:2UTG), composed monomers of 70 amino acids each. The first test carried out in this study was the docking of monomers treated as rigid bodies. In order to perform this, one of the monomers was placed at the origin of coordinates, while the second monomer was rotated and translated around the first monomer according to the probability function of the simulated annealing algorithm or the improvement in the value of the potential function. The scanning of the entire space possess limits on affordable computer processing times. Thus, before the optimization process, the complementarity of both monomers is analyzed by another algorithm that we report elsewhere (see Del Carpio and Yoshimori, submitted for publication). Thus the algorithm starts with the positioning of the monomers facing each other complementarity section and at a distance close enough for interaction to occur but where no atomic clashes are observed. Fig. 3 shows the monomers at the initial position. The separation among the centers of gravity of each monomer being 15 Å.

Fig. 4 illustrates the variation of the energy potential function and the RMS of the complex with respect to the crystal structure through the annealing process. The characteristics of the variation of the total energy function can be described as steadily at the beginning of the process to then change into a steep variation around the minimum. The discrimination power among docking conformations is apparent from this figure, although at the end of the minimization process a slight increase in the RMS value can be observed. This change can be attributed to the threshold value for the van der Waals energy, as clashes between atoms produce infinite values in the Lennard-Jones potential function employed in the present study.

The simulated annealing then proceeds in the direction of lower potential energy for the complex.

For illustration of the annealing process, as well as the discrimination power of the model introduced in the present work, the variation of the different energetic terms of the potential function are shown for the last part of the minimization process, i.e., when the RMS reaches close to 5 Åin difference with

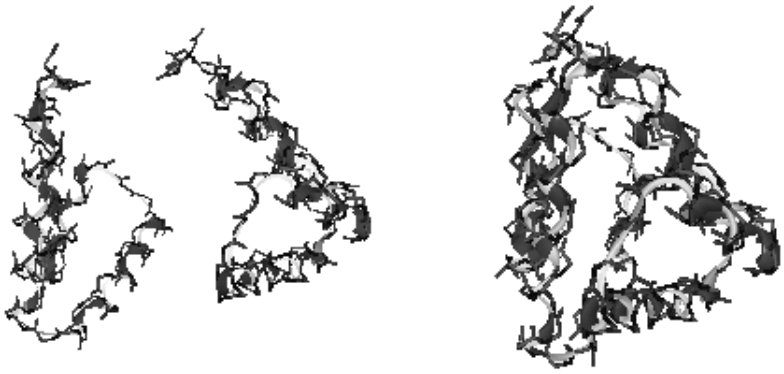

Figure 3: Starting Position (left) for the interaction simulation experiment compared with the crystal complex (right).

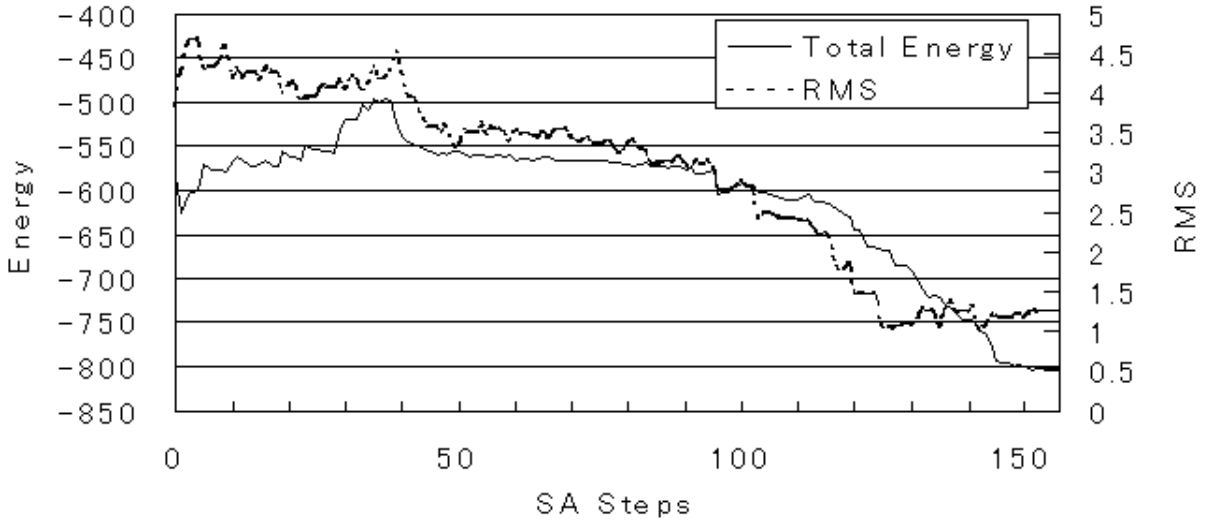

Figure 4: Variation of the RMSD and total interaction energy through the annealing process.

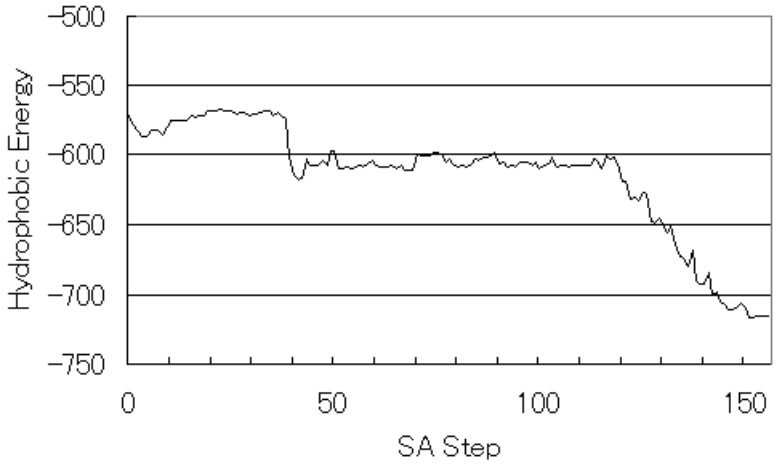

Figure 5: Variation of the Hydrophobic energy through the annealing process.

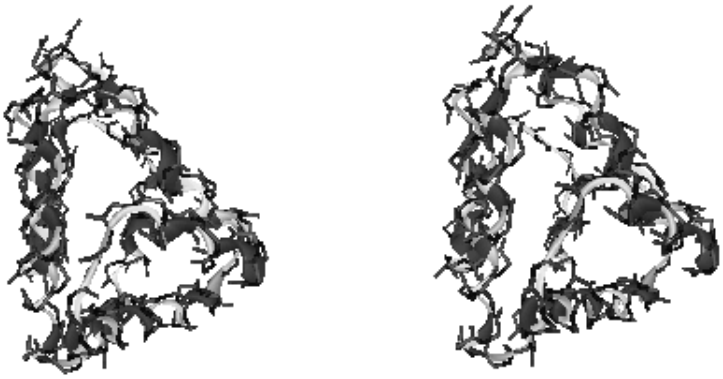

Figure 6: Complex obtained by automatic docking (left) and observed crystal structure (right).

the crystal reference. This analysis shows that all terms of the potential function decrease as the RMS value decreases, this behavior is particularly apparent with the hydrophobic interaction term (Fig. 5).

Consequently the discrimination power of the potential energy model introduced in the present work is evidently high as illustrated by the final structure output by the system and shown in Fig. 6, for which a 1.2 Åin RMS respect to the crystal structure was obtained and shown in the superimposed model in Fig. 7 . Although the minimization carried in the present work has been of rigid bodies, which means that the torsional energy Etor =0, the results are qualitatively and, to a good extent, quantitatively in agreement with theory and experiment.

Figure 7: Superposition of the automatic docking and the crystal structure.

5 Conclusions

Here, we present a new potential function to evaluate protein-protein interaction in solution. The function contains terms expressing the inter-molecular interaction as well as the interaction between solvent and the complex. It has been shown that the discrimination power of the function is high, especially for conformations close to the observed system. The representation of the hydrophobic effect as the sum of van der Waals interactions between the proteins and the surrounding solvent plus the pressure due to the solvent molecules on the surface of the complex, leads to a quantification of this effect so often mentioned and only considered qualitatively in this type of discussions. These preliminary results are encouraging, and applying the function to other systems constituted by proteins, RNA or DNA molecules would be the next step. One aspect which requires to be improved is, however, that related to the complementarity of the monomers to reduce the search space of the minimization algorithm. We have implemented a method (Del Carpio and Yoshimori) that computes geometrical complementarity between monomers composing the complex. The algorithm outputs a limited number of starting positions which are then treated with the SA described here.

A further interesting factor in the potential function introduced here, is the fact that a new solvent effect model can be obtained straightforwardly, and can be used in the analysis of hydration of small molecules, as well as proteins, and other compounds.

References

[1] Bottcher, C.J.F., *Theory of Electric Polarization*, Elsevier Scientific Publishing Company, 1973.

[2] Cherfils, J., Duquerroy, S., and Janin, J., Protein-protein recognition analysed by docking simulation, *Proteins: Struct. Funct. Genet.*, 11:271–280,1991.

[3] Connolly, M.L., Solvent-accessible surfaces of proteins and nucleic acids, *Science*, 221:709-713, 1983.

[4] Cummings, M.D., Hart, T.N., and Read, R.J., Atomic solvation parameters in the analysis of protein-protein docking results, *Protein Science*, 4:2087–2099, 1995.

[5] Del Carpio, C.A., A parallel genetic algorithm for polypeptide three dimensional structure prediction. a transputer implementation, *J. Chem. Inf. Comput. Sci.*, 36:258–269,1996.

[6] Eisenberg, D. and McLachlan, A.D., Solvation energy in protein folding and binding, *Nature*, 319:199–203, 1986.

[7] Fischer, D., Lin, S.L., Wolfson, L., and Nussinov, R., A geometry-based suite of molecular docking processes, *J. Mol. Biol.*, 248:459–477, 1995.

[8] Gabb, H.A., Jackson, R.M., and Sternberg, M.J.E., Modelling protein docking using shape complementarity, electrostatics and biochemical information, *J. Mol. Biol.*, 272:106–120, 1997.

[9] Jackson, R.M., Gabb, H.A., and Sternberg, M.J.E., Rapid refinement of protein interfaces incorporating solvation: application to the docking problem, *J. Mol. Biol.*, 276:265–285,1998.

[10] Jiang, F. and Kim, S., Soft docking: matching of molecular surface cubes, *J. Mol. Biol.*, 219:79–102, 1991.

[11] Krystek, S., Stouch, T., and Novotny, J., Affinity and specificity of serine endopeptidase-protein inhibitor interactions, *J. Mol. Biol.*, 250:258–275, 1993.

[12] Meyer, M., Wilson, P., and Schomburg, D., Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking, *J. Mol. Biol.*, 264:199–210, 1996.

[13] Nicholls, A., Sharp, K., and Honig, B., Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins*, 11:281–296, 1991.

[14] Reiss, H., Frisch, H.L., and Helfand, E., Aspects of statistical thermodynamics of real fluids, *J. Chemical Physics.*, 32:119–124, 1960.

[15] Richmond, T.J., Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect, *J. Mol. Biol.*, 178:63–89, 1984.

[16] Shoichet, B.K. and Kuntz, I.D., Protein docking and complementarity, *J. Mol. Biol.*, 221:327–346, 1991.

[17] Still, W.C., Tempczyk, A., Hawley, R.C., and Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics, *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[18] Totrov, M.M. and Abagyan, R.A., Detailed ab initio prediction of lysozyme-antibody complex with 1.6Åaccuracy, *Nature Struct. Biol.*, 1:259–263, 1994.

[19] Weng, Z., Vajda, S., and Delisi, C., Prediction of protein complexes using empirical free energy functions, *Protein Science*, 5:614–626, 1996.