Evaluation of the third solvent clusters fitting procedure for the prediction of protein-protein interactions based on the results at the CAPRI blind docking study

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · JULY 2003

Impact Factor: 2.63 · DOI: 10.1002/prot.10385 · Source: PubMed

5 AUTHORS, INCLUDING:

Hideaki Umeyama
Kitasato University
165 PUBLICATIONS **2,533** CITATIONS

SEE PROFILE

PROTEINS: Structure, Function, and Genetics 52:15–18 (2003)

Evaluation of the Third Solvent Clusters Fitting Procedure for the Prediction of Protein–Protein Interactions Based on the Results at the CAPRI Blind Docking Study

Katsuichiro Komatsu, Youji Kurihara, Mitsuo Iwadate, Mayuko Takeda-Shitaka, and Hideaki Umeyama*
School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan

ABSTRACT **To predict protein–protein interactions, rough or coarse handling for the induced fit problem is proposed. Our method involves the overlap of two hydrophobic interactions as "third solvent clusters fitting." Predictions for binding sites and geometric centers were acceptable, but those of the binding axes were poor. In this study, only the largest benzene cluster was used for the third solvent clusters fitting. For the next CAPRI targets, we must perform protein–protein interaction analyses, which include other smaller benzene clusters. Proteins 2003;52:15–18.** © **2003 Wiley-Liss, Inc.**

Key words: protein–protein docking; protein-binding site; hydrophobic interaction; benzene cluster; induced fit problem

INTRODUCTION

Because conformational changes of subunits are normally induced, the results of docking uncomplexed subunits have been unsatisfactory in the prediction of protein–protein interactions.[1,2] The basis of one strategy is that the conformational fit induced on binding is mimicked by allowing some overlap of the interacting surfaces according to the softening of the geometric criteria.[3] Then, a soft interaction energy function on a grid is used for rigid body docking followed by side-chain optimization.[4] However, the method of Fernández-Recio et al. requires the finding of an optimal evaluation function by adjusting several energy component terms such as van der Waals, electrostatic interactions, hydrogen bonds, hydrophobic interactions, and hydration energy.

Therefore, we have developed a different docking method that includes a kind of induced fit conformation. In a living cell, there are many other proteins around any particular protein. Between any two protein molecules in living cells are intermolecular forces. Electrostatic and ionic interactions play a dominant role over greater distances but are subject to interception due to disturbances by many other electrolytes such as proteins. However, hydrophobic interactions seem to play a more dominant role in cells crowded with proteins, although hydrogen bonds and van der Waals interactions are also very important when two proteins are in direct contact.

For entropy, rough or coarse handling is needed for the induced fit problem in protein–protein interactions. Consequently, a method involving the overlap of two hydrophobic interactions, our "third solvent clusters fitting," is proposed in this study. The docking procedure was determined by adding some new ideas and techniques to our protein docking methods and tools during CAPRI rounds 1 and 2. Our emphasis is on the construction of an automatic method of coarse protein–protein docking.

MATERIALS AND METHODS

Our scheme for constructing an automatic method for protein–protein docking is shown in Figure 1. The method is separated into three steps: (i) prediction of the binding site of the protein monomer for another one [(1)–(6) in Fig. 1], (ii) docking of the latter to the former [(7)–(8) in Fig. 1], and (iii), refinement of the protein–protein complex [(9) in Fig. 1]. Figure 2 illustrates the detailed procedures for (6) and (7).

Hydrogen atoms for bonding to heteroatoms were added geometrically. His residues were regarded as neutral, with the proton on ND1. We used no biochemical information for any target of CAPRI rounds 1 and 2 and used all the subunits without any special consideration. The protein–protein complexes were modeled by using only the procedures given below.

(1) TIP3P water molecules[5] such as in a sphere model are generated. (2) Water molecules within 3.5 Å on the surface of hydrophobic amino acid residues with maximum solvent accessibility of side-chain (MSAS)[6] >30% are substituted into benzenes except that any water molecules within 1.5 Å of a benzene molecule were not substituted. (3) Here all the remaining water molecules are removed from the region around the protein molecule, so that the three-dimensional coordinates of the protein monomer and benzene molecules are obtained.

(4) Again, the surfaces of the protein and benzene regions are filled with TIP3P water molecules[5] in a periodic boundary condition. (5) The whole molecules are minimized by molecular mechanics (MM), followed by molecular dynamics (MD) at 300K for 30 ps. The Apricot program,[7] which was developed in our laboratory, and the AMBER united-atom force-field parameter[8] were used for

Grant sponsor: Ministry of Education, Culture, Sports, Science and Technology of Japan.

*Correspondence to: Hideaki Umeyama, School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan. E-mail: umeyamah@pharm.kitasato-u.ac.jp

Received 31 October 2002; Accepted 13 December 2002

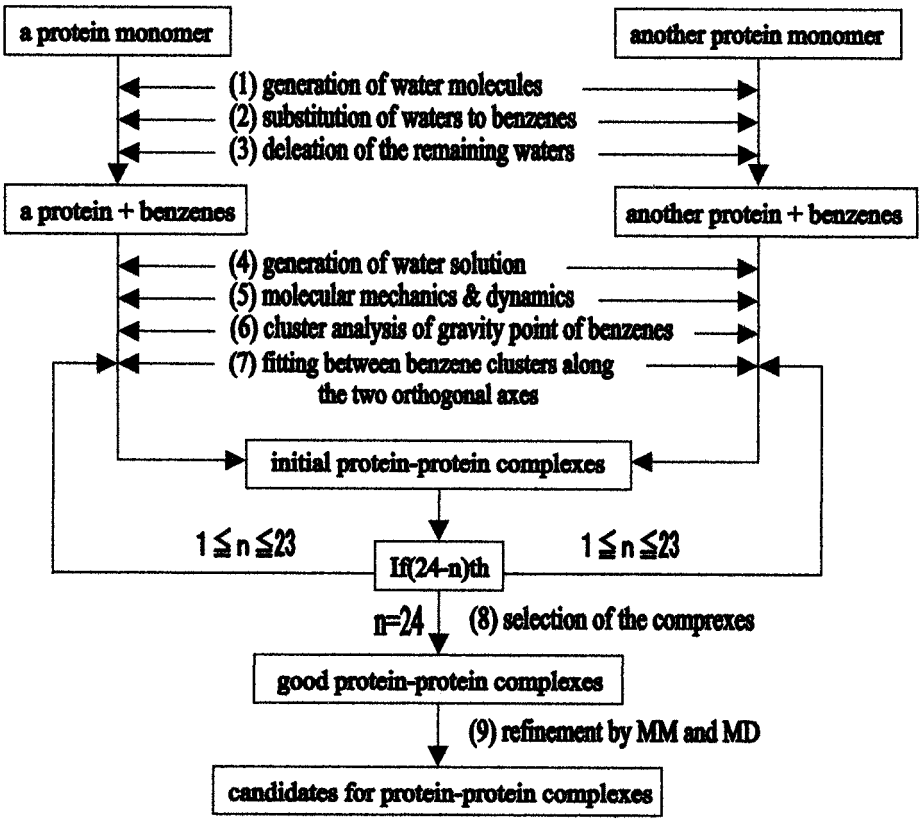

Fig. 1. Global strategies of third solvent clusters fitting.

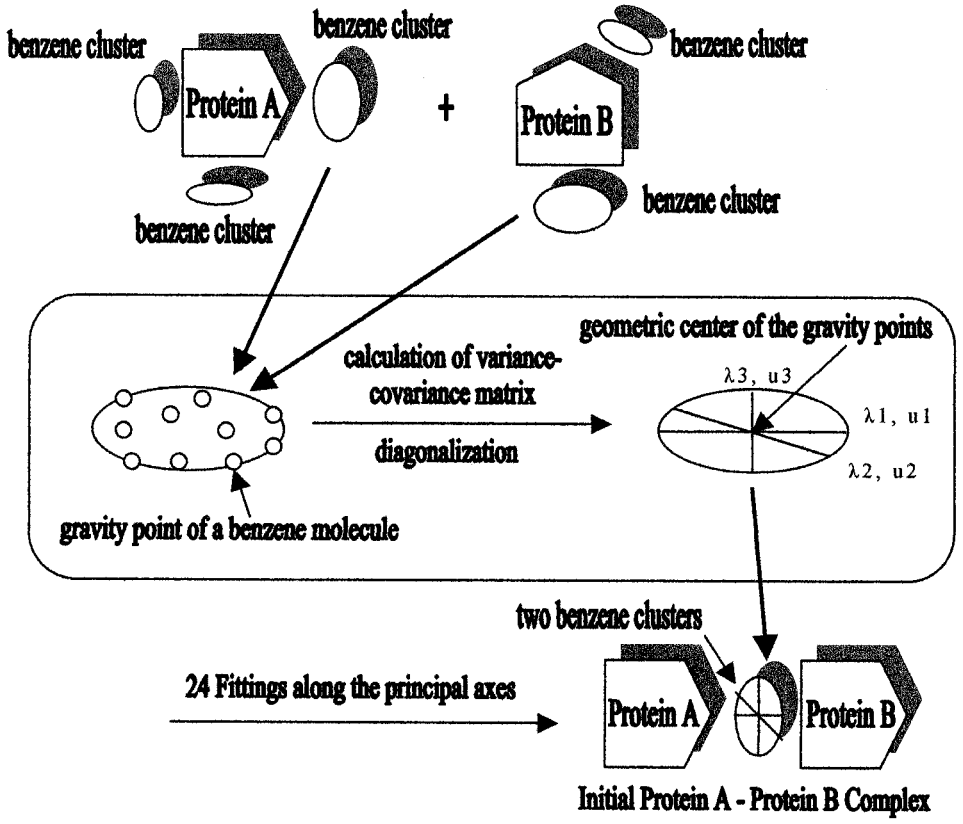

Fig. 2. Schematic of cluster analysis of gravity points of benzenes and 24 fittings along the principal axes between the largest benzene clusters for each protein A and B. By solving a problem by eigenvalues for a variance–covariance matrix on central coordinates of benzenes in the cluster, eigenvalues λ_1, λ_2, and λ_3 ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvectors u_1, u_2, and u_3 corresponding to each eigenvalue are obtained. Twenty-four pairs of the three principal axes of protein B were fitted on the first and second axes of protein A.

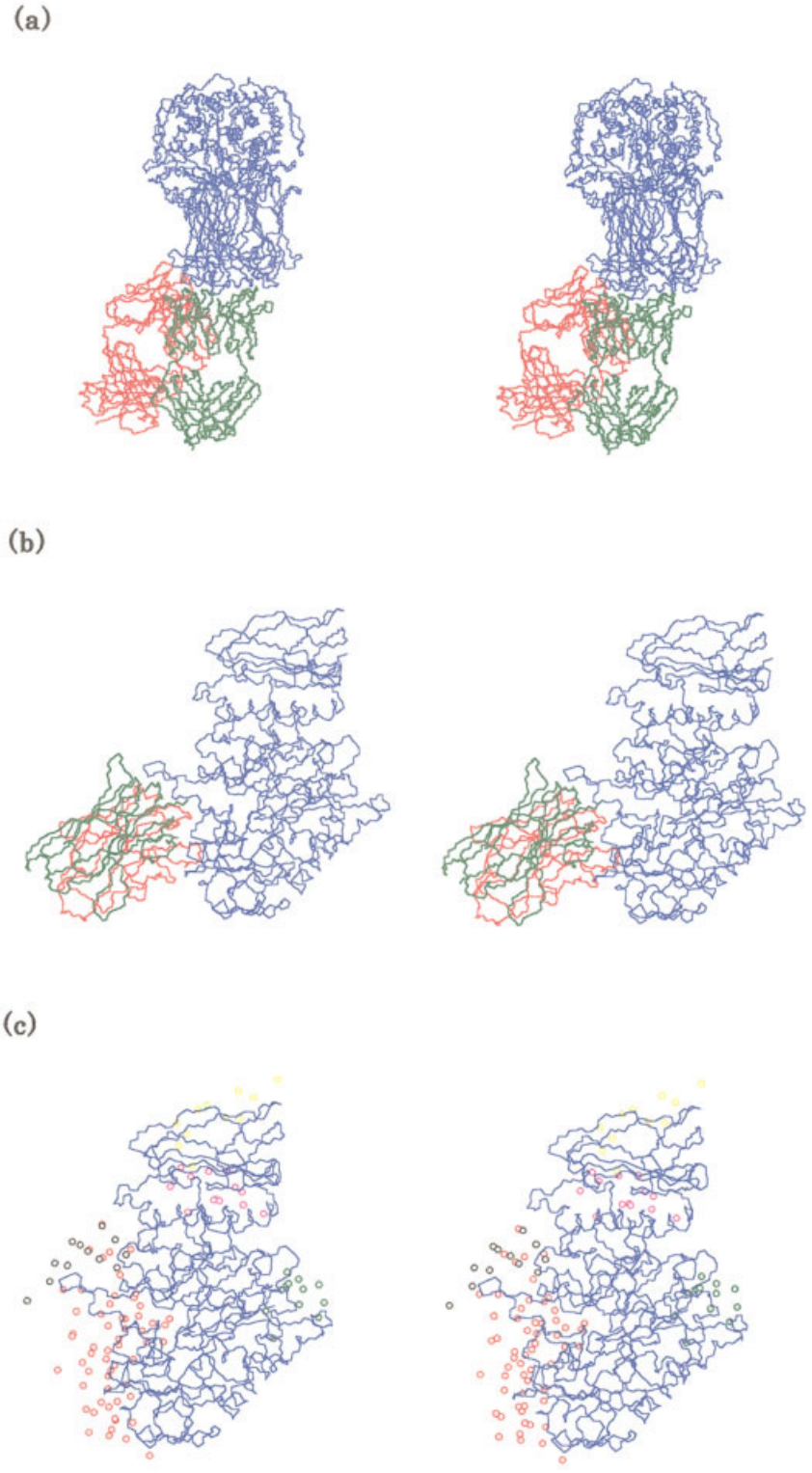

Fig. 3. Stereoview of the prediction result for CAPRI target 2 (a), target 6 (b), and the benzene clusters estimated by our method (c). **a:** Main chains of bovine rotavirus VP6 (ABC chains) and Fab HL chains of the antibody in crystal; that predicted for the latter are shown in blue, red, and green, respectively. **b:** Main chain of α-amylase and camelid VHH Cab D09 antibody in crystal; that predicted for the latter are shown in the same colors. **c:** Main chain of α-amylase is indicated by a blue line; the first, second, third, fourth, and fifth benzene clusters are indicated by red, magenta, green, black, and yellow circles, respectively.

these empirical potential calculations. Consequently, benzene molecules diffuse or integrate around or inside the protein. The benzenes that attained a solvated equilibration are divided into clusters with an 8 Å threshold as the minimum intermolecular distance between benzene molecules in our cluster analysis.

(6) Because a benzene cluster can be represented as an ellipsoid, the three principal axes of inertia of the benzene cluster can be estimated by solving a problem by eigenvalues for a variance-covariance matrix on central coordinates (*x, y,* and *z*) of the benzenes. As a result, the center of the gravity points, the three eigenvalues, and the corresponding eigenvectors can be determined. (7) Any two principal axes of protein B were fitted on the two axes of protein A. After 24 fittings along the two principal axes, the complexes of protein A and protein B that overlapped were omitted. Thus, many of the initial protein A–protein B complexes were obtained. Here it should be cautioned that the largest benzene cluster was used in our trial as a typical example for protein A or protein B. (8) We selected good models manually, so that about 10 model complexes were gone forward.

(9) We performed MM and MD with minor restraints on Cα atoms of the residues in vacuum for the selected models with a few short contacts between proteins A and B. Then, the 6 (target 1), 10 (targets 2 and 3), or 5 (targets 4–7) models, which were energetically and visually selected, were submitted as good complexes in a CAPRI contest for each target.

RESULTS AND DISCUSSION

Figure 3(a) shows the results of the prediction for CAPRI (round 1) target 2 complex [bovine rotavirus VP6 (ABC chains) and Fab HL chains of the antibody]. The relative arrangement of the two protein groups is somewhat acceptable, even though the theta angle and the distance published by the CAPRI organizing committee were 117° and 20 Å, respectively, which may lead to the poor root-mean-square values shown by Wodak (SCMB) at the first CAPRI meeting.

Figure 3(b) shows the results of the prediction for CAPRI (round 2) target 6 complex (α-amylase and camelid VHH Cab DO9 antibody). The relative arrangement of the two protein groups is somewhat acceptable, and the distance, 5 Å, published by the CAPRI organizing committee means a better placed model protein close to the experimental structure. However, the theta angle, 142°, indicates an incorrectly rotated model for the experimental ligand arrangement and, as a result, the poor root-mean-square values shown by Wodak (SCMB) at the first CAPRI meeting.

Our other results for the CAPRI targets are shown in Table I. Figure 3(c) shows the predicted binding sites of α-amylase for CAPRI targets 4, 5, and 6. The red circles in Figure 3(c) indicate the positions of the benzene clusters estimated by our method. In this study, only the largest benzene cluster was used for the third solvent clusters fitting. The second, third, •, •, and nth benzene clusters were not used. On reflection, it might be significant that the other benzene clusters shown in Figure 3(c) contributed to the complex formation between the two protein groups.

TABLE I. Our Best Results for CAPRI Targets 1–7

Target-Model	Init-Fit[a]	Theta Angle[b]	Distance[c]
T01-Model 2	8/24	135.71	32.790
T02-Model 3	10/24	117.30	19.936
T03-Model 5	9/24	37.58	32.588
T04-Model 4	11/24	121.51	29.074
T05-Model 1	9/24	131.23	28.451
T06-Model 5	9/24	142.20	4.928
T07-Model 1	8/24	153.83	31.877

[a]Number of initial protein–protein complexes not having too great a degree of overlap.
[b]Rigid body rotation angle (degree) of smaller size group model for the corresponding experimental structure.
[c]Distance (Å) between the geometric centers of the smaller size group model and the corresponding experimental structure.

We tested the effectiveness of our procedure by using four systems after CAPRI round 2, for which ternary structures of protein–protein complexes are registered in PDB.[9] When we used the largest benzene clusters for each protein A and B (three of four systems), and when did the first and the third largest clusters (one of four systems), the results were that predictions for the binding sites and the geometric centers were acceptable but that those for the binding axes were poor.

For the next CAPRI targets, we must perform protein–protein interaction analyses that include other smaller benzene clusters. Even so, for this CAPRI meeting, our new rough handling of the third solvent clusters fitting has been shown to be useful for predicting protein–protein interactions.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

REFERENCES

1. Vakser IA, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
2. Maurer MC, Trosset J-Y, Lester CC, DiBella EE, Scheraga HA. New general approach for determining the solution structure of a ligand bound weakly to a receptor: structure of fibrinogen Aα-like peptide bound to thrombin(S195A) obtained using NOE distance constraints and an ECEPP/3 flexible docking program. Proteins 1999;34:29–48.
3. Vakser IA. Protein docking for low-resolution structures. Protein Eng 1995;8:371–377.
4. Fernández-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. Protein Sci 2002;11:280–291.
5. Jorgensen WL, Chandrasekhar J, Madura JD. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
6. Akahane K, Nagano Y, Umeyama H. Hydrophobic effect on the protein-ligand interaction: hydrophobic field-effect index and hydrophobic correlation index. Chem Pharm Bull 1989;37:86–92.
7. Yoneda S, Umeyama H. Free energy perturbation calculations on multiple mutation bases. J Chem Phys 1992;97:6730–6736.
8. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.
9. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.