

Shide Liang
Zhijie Liu
Weizhong Li
Lisheng Ni
Luhua Lai

*Institute of Physical
Chemistry,
College of Chemistry and
Molecular Engineering,
Peking University,
Beijing 100871,
the People's Republic of
China*

*Received 27 October 1999;
accepted 20 April 2000*

Construction of Protein Binding Sites in Scaffold Structures

Abstract: We have developed a strategy for grafting a protein–protein interface based on the known crystal structure of a native ligand and receptor proteins in a complex. The key interaction residues at the ligand protein binding interface are transferred onto a scaffold protein so that the mutated scaffold protein will bind the receptor protein in the same manner as the ligand protein. First, our method identifies key residues and atoms in the ligand protein, which strongly interact with the receptor protein. Second, this method searches the scaffold protein for combinations of candidate residues, among which the distance between any two candidate residues is similar to that between relevant key interaction residues in the ligand protein. These candidate residues are mutated to key interaction residues in the ligand protein respectively. The scaffold protein is superposed onto the ligand protein based upon the coordinates of corresponding atoms, which are assumed to strongly interact with the receptor protein. Complementarity between scaffold and receptor proteins is evaluated. Scaffold proteins with a low superposing rms difference and high complementary score are accepted for further analysis. Then, the relative position of the scaffold protein is adjusted so that the interfaces between the scaffold and receptor proteins have a reasonable packing density. Other mutations are also considered to reduce the desolvation energy or bad steric contacts. Finally, the scaffold protein is cominimized with the receptor protein and evaluated. To test the method, the binding interface of barstar, the inhibitor of barnase, was grafted onto small proteins. Four scaffold proteins with high complementary scores are accepted. © 2000 John Wiley & Sons, Inc. *Biopoly* 54: 515–523, 2000

Keywords: protein–protein interaction; grafting; barnase–barstar; protein design

INTRODUCTION

To date, a variety of experiments have been performed to rearrange structural contexts in proteins or

aim at the construction of new functionality within a given scaffold. The specificity of DNA-binding proteins has been successfully altered by exchanging the solvent-exposed functional residues between homol-

Correspondence to: Luhua Lai
Contract grant sponsor: Department of Science and Technology of China and Natural Science Foundation of China
Biopolymers, Vol. 54, 515–523 (2000)
© 2000 John Wiley & Sons, Inc.

ogous proteins.¹ Jones et al. transferred antigen recognition specificity by replacing the complementarity-determining regions in a human antibody with those from a mouse.² Substrate specificity of trypsin is converted to that of chymotrypsin by exchanging related functional loops.³ In these studies, structural information was frequently gained by sequence alignment and structure superposition of the ligand protein to a homologous target structure. However, it is more exciting to graft protein functional epitope between nonhomologous proteins, which requires the help of computational tools. Because of the straightforward relationship between structure and functionality, ligand binding sites are often considered for transferring.

We are especially interested in protein-protein interaction. Usually, protein-protein complexes have buried solvent accessible surface areas at the interfaces in the range of 1200–2000 Å² with 10–30 amino acid residues on each of the partners.⁴ In most cases, only 3–4 residues dominate binding energy.^{5–9} We have developed a strategy for grafting protein-protein interaction epitopes,¹⁰ which search scaffold proteins with complementary interfaces based on molecular docking and suitable positions for transferring ligand protein key interaction residues. However, the docking procedure involves six dimensional (6D) search and is extremely CPU intensive. This paper describes an alternative strategy, which runs much more quickly.

METHODS AND MATERIALS

The prerequisite of our grafting strategy is the availability of a complex crystal structure of the ligand and receptor proteins. The key interaction residues and atoms at the ligand protein interface are identified according to mutation information and theoretic analysis of the complex structure. Then candidate residues in a scaffold protein for grafting are explored.

Searching for Candidate Residues

In the first stage of our program, we select qualified residues in a scaffold protein that are fit for transferring ligand protein key interaction residues. The candidate residues in the scaffold protein cannot be Pro, Cys, or Gly in a glycine turn, which are important for maintaining the structure of the scaffold protein.¹⁰ For Gly not in a glycine turn, it can be mutated and a pseudo C_β atom is added in convenience of computation. The candidate residues can not be the first two residues at the N-terminal or the last two residues at the C-terminal, as the backbone of protein usually is flexible at terminals. Since important functional residues are usually

solvent exposed, residues in the scaffold protein that are located on the surface and with an outward C_α → C_β vector are selected at first to reduce computational time. This is done by an atomic solvent accessible surface calculation using the NACCESS program.¹¹ All side-chain atoms except C_β are deleted and the probe radius is set to 3.0 Å, which is much larger than the frequently used probe radius of 1.4 Å. In such case, atoms located in the protein core usually are not solvent accessible. The solvent accessible surface of those atoms located on the protein surface may increase along with the probe radius. Only residues with a solvent accessible C_β atom are selected.

Then set reduction algorithm is applied to search the scaffold protein for combinations of residues forming a potential binding site.¹² For any two key interaction residues, residue1 and residue2, in the native ligand protein, four distances of C_{α1}—C_{α2}, C_{α1}—C_{β2}, C_{β1}—C_{α2}, and C_{β1}—C_{β2} are calculated. For any two residues, residue1' and residue2', in the scaffold protein, the four corresponding distances are also calculated. If the differences between relevant distances of residue1'—residue2' and residue1—residue2 are within 3 Å, we define that residue1'—residue2' and residue1—residue2 have the same geometric relationship. Combinations of residues in the scaffold protein, among which the geometric relationship between any two residues is the same as that of the two relevant key interaction residues in the ligand protein, are sought and treated as potential sites for grafting.

Superposing and Complementarity Evaluation

Candidate residues are mutated to key interaction residues of the ligand protein correspondingly. The side chain of the mutant residue adopts different favorable dihedral angles that are referred as rotamers at each candidate position. A set of rotamer combinations is formed at each potential binding site, which usually has multiple candidate residues. The basic rotamer library of Maeyer is used in this study, which contains 330 elements.¹³ Then the scaffold protein is superposed onto the ligand protein, which is in complex with the receptor protein, based on the corresponding coordinates of key interaction atoms.¹⁴ The scaffold protein is checked with each rotamer combination. If the key residues such as Asp, Glu, Val, Leu, Tyr, and Phe contain symmetric atoms, the names of symmetric atoms are exchanged for each rotamer and a new rotamer is generated. Since the superposition is done based on corresponding atoms with the same names, the new rotamers should also be checked, which will enlarge the set of rotamer combination significantly. Mirror images of the ligand geometry, which have the same geometry relationship between residues but a different handedness, are discarded due to a high rms difference at this stage. If the superposing rms difference is below the tolerance value, geometric complementarity between the scaffold and receptor proteins is evaluated using grid-based molecular representations.¹⁵ Van der Waals radius of nonhydrogen atoms is chosen as 1.8 Å. All hydrogen

Table I Packing Densities at the Interfaces of Protein Complexes^a

PDB	Names	No. of Atoms in Average	Packing Density at the Interfaces
2ptc	Trypsin/BPTI	143	0.744
1tec	Thermitase/Eglin C	147	0.746
4sgb	Proteinase B/PCI-1	122	0.760
1tpa	Anhydrotrypsin/BPTI	147	0.743
2sni	Subtilisin novo/CI-2	161	0.731
4tpi	Trypsinogen/inhibitor analog	156	0.743
3sgb	Proteinase B/OMTK3	126	0.758
1cho	α -Chymotrypsin/OMTKY3	141	0.714
1vfb	Lysozyme/D1.3 Fab	137	0.724
1mlc	Lysozyme/Fab D44.1	140	0.710
1fdl	Lysozyme/D1.3 F _v	127	0.676
3hfl	Lysozyme/HyHEL 5 Fab	162	0.698
1mel	Lysozyme/V _H	159	0.737
1dvf	D1.3 F _v /E5.2 F _v	153	0.681
1nca	Neuraminidase/Fab Nc 41	182	0.725
1nmb	Neuraminidase/Fab Nc 10	122	0.668
1hbs	Deoxyhemoglobin mol 1–2 contact	52	0.653
1axi	HGH/HGHBP	238	0.670
1brs	Barnase/barstar mutant	165	0.726

^a Packing densities were calculated according to Richards using method B.¹⁶ A 2.8 Å solvent shell was added onto the protein surface. The atoms at the interface were chosen on the basis of having buried solvent accessible surface upon association and with less than 15 Å² solvent accessible surface. Packing density for each of these atoms was calculated individually and averaged. Atom solvent accessible surface was calculated using the program NACCESS with a 1.4 Å probe radius.¹¹

atoms are neglected. The receptor protein is digitized onto a 3D grid with the grid step of 1 Å. Grid nodes inside of the receptor protein are assigned a value of 1. Those grid nodes outside of the receptor protein are set to zero. The scaffold protein is projected onto a similar grid. The grid nodes occupied by the scaffold protein backbone atoms are assigned a value of -10 and those occupied by side-chain atoms of scaffold protein are assigned a value of -2. Two layers of grid nodes on the scaffold protein surface are assigned a value of 1 and others are set to zero. The geometric complementary score is calculated by accumulating the products of the numbers assigned to the two grids at each coinciding grid point. These projections allow small gaps and penetrations, but bad steric contacts between the scaffold protein backbone atoms and the receptor protein are heavily punished. Among the set of rotamer combinations at each potential binding site, only the one with the highest complementary score is saved. If the score is above the tolerance level, we accept the scaffold protein, which is transferred with the saved rotamer combination, for further analysis.

Refinement

It is very common for atoms at interface to make small shifts during complex formation. Due to the induced fit, a free scaffold protein does not require a good complementary interface against the receptor protein before the complex formation. Thus, very low punishment on atom penetrations

is used in evaluating the complementarity between scaffold and receptor proteins. This may result in problems such as an unreasonable packing density at the interfaces, which can be overcome by regulating the relative position of the scaffold protein. Atom packing density is defined as the ratio of the volume enclosed by the van der Waals envelope to the volume that it occupies.¹⁶ Interface packing density is defined as the mean packing densities of atoms at the interfaces of scaffold and receptor proteins. Since the positions of water molecules on the surface of the complex are unknown, it is impossible to determine the actual volume occupied by solvent accessible atoms accurately. So atoms with more than 15 Å² solvent accessible surface area are not included in average. The packing densities of some atoms may be more than 1 due to overlapping. For these atoms, packing density is set to 1 to alleviate the effect on the mean value. If the interface packing density of the scaffold and receptor proteins is above the upper limit, the scaffold protein is moved 0.1 Å along the direction from the geometric center of the atoms at the interface of the receptor protein to that of the scaffold protein. After the translocation, the buried solvent accessible surface area at the interface and the rms difference of key interaction atoms between the native ligand and scaffold proteins are calculated. If the buried solvent accessible surface is less than 1200 Å² or the rms difference is larger than tolerance value, the scaffold protein is rejected. Indeed, for the 19 complexes in Table I, only deoxyhemoglobin molecule1–molecule2 contacts have buried solvent accessible surface at the interface

less than 1200 \AA^2 and the binding free energy is only $-4.8 \text{ kcal mol}^{-1}$.¹⁰ The translocation is continued until the interface packing density falls in the permissible range. When the interface packing density is below the lower limit, the moving direction is reversed.

Then the complementarity of each side chain at the scaffold protein binding site is evaluated against its surrounding atoms and the side chain is mutated to different rotamers in the rotamer library of Maeyer except Cys, Pro, and four charged residues: Asp, Glu, Arg, and Lys.¹³ Charged residues are avoided due to their high desolvation energies. We define a side chain at the binding site when it has buried solvent accessible surface upon association. The complementary score for each rotamer is calculated correspondingly. Again, grid-based molecular representations are used. The side-chain atoms are digitized on a 3D grid with a grid step of 1 \AA . Grid nodes inside of the side chain atoms are assigned a value of -10 and two layers of grid nodes on the side chain surface are assigned a value of 1 . Others are set to zero. The surrounding atoms of the side chain are projected onto a similar grid. Grid nodes inside of the receptor atoms are assigned a value of 2 and those inside of the scaffold protein are assigned a value of 1 . Others are set to zero. This projection pattern aims to have a bias on the receptor protein. Products of the numbers assigned to the two grids at each coinciding grid point are accumulated as complementary score. The rotamer with the highest complementary score is saved. If the score is 200 higher than original side chain, or it is more than 100 and 100 higher than the original side chain, the original residue is mutated to the saved rotamer. Such an operation eliminates not only atom overlapping but also large holes at the interface. To reduce desolvation energy, charged residues, which are totally buried at the interface and do not form any salt bridge, are also mutated to the saved rotamer. If none of O/N atoms of a charged residue have more than 10 \AA^2 solvent accessible surface, we consider that residue is totally buried. When the distance between two charged O/N atoms is under 3.2 \AA , the pair is accounted as a salt bridge.

Finally, the scaffold protein in complex with the receptor protein is scrutinized using the interactive computer graphics tools in Quanta (Molecular Simulation, Inc., San Diego, CA). Mutations except the grafted key residues may be reversed or additional mutations can be introduced according to visual analysis. Attention should be paid to maintain original structure and optimize the interactions between molecules. Indeed, solvent accessible residue mutations are avoided, when possible, because adjusting side-chain conformation of these residues is a much simpler process and is less risky to devastate original structure. Side-chain dihedral angles at the interface are regulated manually to achieve more hydrophobic interactions or form hydrogen bonds.

Minimization and Evaluation

The mutated scaffold protein is then minimized with the receptor protein by Charmm.^{17,18} Distance constraints are used to mimic the strong atom–atom interactions between

two molecules in the complex of the native ligand and receptor proteins. The constraint potential is used as follows:

$$E(r) = 10 \times (r - r_0)^2$$

where r_0 is the distance (\AA) between two atoms involved in strong interactions in the native complex structure; r is the distance (\AA) between relevant atoms in the complex of scaffold and receptor proteins. At first all dihedral angles are constrained with a force constant of $5 \text{ kcal mol}^{-1} \cdot \text{rad}^{-2}$ to avoid overminimization. One hundred cycles of the steepest descents minimizer are used. Then the dihedral angle constraints are removed and the complex is minimized by 50 steps of adopted-basis Newton–Raphson minimizer (ABNR). A distance-dependent dielectric constant is used in calculating electrostatic interactions. Deviations from target distances are recorded for each atom pair that has been assigned a constraint. Under such minimization conditions, the rms difference between minimized, and preminimized protein backbone atoms is usually no more than 0.3 \AA .

An empirical method is used to calculate protein–protein binding free energy. The formula to calculate binding free energy ΔG_{cal} was deduced by a least square fit.¹⁰

$$\Delta G_{\text{cal}} = -0.00516 * S_{\text{pho}} + 0.0152 * S_{\text{phi}} \\ - 1.04 * N_{\text{pair}} - 3.97$$

where S_{pho} , S_{phi} , and N_{pair} are buried solvent accessible hydrophobic surface, hydrophilic surface at the interface, and the number of hydrophilic pairs between two molecules, respectively.

RESULTS

Selection of Ligand and Scaffold Structures

Barnase is 110-residue extracellular ribonuclease of *Bacillus amyloliquefaciens*. The same organism produces an 89-residue polypeptide inhibitor, barstar. Barnase and barstar form a very tight complex, with a K_d of 1.3×10^{-14} , which corresponds to a ΔG of $-18.9 \text{ kcal mol}^{-1}$ at 25°C .¹⁹ A great deal of research work has been done upon these two molecules and they have been ideal model for protein folding and interaction studies.^{19–24} Mutagenesis studies have shown that the residues in barstar, which contribute most to the tight interaction, are Asp39, Asp35, and Tyr29. These three residues were grafted in this study. OD1, OD2, and CG atoms of the two aspartic acid residues were identified as key interaction atoms. We included CG atoms to make sure that the angles of $\text{CG—O} \cdots \text{H}$ were favorable in the expected hydrogen bonds. CD1, CE1, and CE2 were considered as

key interaction atoms for Tyr29, which will determine the position of the tyrosine aromatic ring.

We chose scaffold proteins with similar dimension to that of barstar. A total of 60 crystal or averaged nmr structures with residue number in the range of 86–95 were selected from the Protein Data Bank (PDB) including averaged nmr structure of barstar (PDB code 1bta), which is treated as a positive control. The rms difference of backbone atoms between 1bta and barstar in barnase–barstar complex is 0.90 Å.

Searching for Target Sites

The barnase–barstar complex structure (PDB code 1brs) was used as template structure. The key interaction residues of barstar are on a convex binding interface and the calculated solvent accessible surfaces of C_{β} atoms of Tyr29, Asp35, and Asp39 in barstar are all larger than 50 Å². We believe the candidate residues in scaffold protein should also be located on a convex surface and with solvent accessible C_{β} atoms to make favorable interactions with concave barnase interface. So the lower limit for the C_{β} solvent accessible surface of candidate residues was set to 20 Å². For 1bta, 410 potential binding sites, which mimic the geometric relationships between any two barstar key interaction residues, were selected. In this study, 2048 rotamer combinations were checked for each potential binding site. The computation took about half an hour on a SGI O2 workstation for 1bta. When the low limit was set to zero, 6723 potential binding sites were selected and it is extremely CPU intensive to check all of them. However, when the key interaction residues are located in a narrow cave, it is more reliable to search all residues in a scaffold protein. The C_{β} atom of Asp189 in bovine trypsin, which is located in a narrow cave and strongly interacts with bovine pancreatic trypsin inhibitor, is not solvent accessible under the same conditions as described in Methods and Materials. Fortunately, we are interested in transferring the convex binding epitope of inhibitors in most cases.

The scaffold protein was superposed onto barstar based upon the coordinates of key interaction atoms. If the superposing rms difference was below 1 Å, the complementarity between the scaffold protein and barnase was evaluated. The lower limit of the geometric complementary score between a scaffold protein and barnase was set to 300. We determined the acceptable range of interface packing density according to statistical analysis. Nineteen protein–protein complexes were used (Table I). For most of them, the packing density falls in the range of 0.67–0.76, which was considered as permissible interface packing den-

sity of the modeled complexes. The search took 33 h on a SGI O2 workstation and 20 binding sites were selected. Some of them belong to the homologous proteins or the same protein. For 1bta, two binding sites were predicted. One of them, which has a lower superposing rms difference and higher complementary score, is very similar to barstar native binding epitope in complex structure. The candidate residues were correctly predicted. The rms difference of backbone atoms, which are less than 8 Å far from barnase, is only 0.99 Å.

We believe it is quite safe to set the lower limit of C_{β} solvent accessible area of candidate residues to 20 Å² in this study. During a test search when the lower limit was set to 30 Å², 18 binding sites were selected and the computational time reduced to one half of the original search.

Analysis of the Selected Binding Sites

We chose four representative scaffold proteins for visual analysis, which have a binding site of high complementary score and minimal mutated residues. The four accepted scaffold proteins included the barstar averaged nmr structure (1bta), cell adhesion protein tanascin (1ten), T-plasminogen activator F1-G (1tpg), and sex-lethal protein (2sxl). No mutations were suggested for 1bta by the program. Except for the grafted key residues, three additional mutations were suggested for 1ten. They were T817M, P865M, and D866Q. After visual inspection, only two mutations, T817G and D866Q, were made. The mutation of T817G could further reduce the bad steric contacts with barnase than T817M. The mutation of P817M was reversed since the penetration between the Pro817 and barnase was not heavy compared with the risk of devastating original structure by the mutation. Two additional mutations, Y33L and H44M, were suggested by the program for 1tpg. The mutation of H44M was reversed since His44 is solvent accessible and the bad steric contacts with barnase could be avoided by regulating the side-chain dihedral angles. Three additional mutations were suggested for 2sxl, D14M, E19A, and Y21M. After visual analysis, the mutation of D14M was reversed since the buried carboxyl of the ASP14 could become solvent accessible by regulating the side-chain dihedral angles. The mutations of E19Q and Y21Q were made instead of E19A and Y21M. The introduction of two glutamines resulted in 5 hydrogen bonds across binding interface. Additional mutation of R25S was introduced to avoid unfavorable electrostatic interaction with Lys27 in barnase. The introduced Ser25 also forms a hydrogen bond with Lys27 of barnase mimic Thr42 of barstar.

Table II Overview of Interfaces Between Barnase and Scaffold Proteins

Scaffold Protein	Buried Surface (Å ²) ^a	Compl. Score ^b	Packing Density	Candidate Positions	RMSD (Å) ^c	Mutated Residues ^d
1bta	1520	642	0.75	29, 35, 39	0.94	
1ten	1570	545	0.72	845, 862, 864	0.89	T817G D866Q
1tpg	1570	420	0.76	20, 67, 69	0.97	Y33L
2sxl	1600	492	0.75	71, 18, 22	0.98	E19Q Y21Q R25S

^a Solvent accessible surface was calculated using the program NACCESS with a 1.4 Å probe radius.¹¹

^b Geometric complementary score between a scaffold protein and barnase was calculated as described in Methods and Materials.

^c The rms difference between key interaction atoms of barstar in complex structure and those of a scaffold protein.

^d Mutations were made according to the suggestions of the program and visual analysis.

Table II outlines the interfaces between the four scaffold proteins and barnase.

The mutated scaffold proteins were cominimized with barnase by Charmm. Distance constraints were used on the complexes between the scaffold proteins and barnase during energy minimization to mimic strong atom–atom interactions between barstar and barnase. Hydrogen bonds and salt bridges between key interaction residues of barstar and barnase were considered to be involved in strong interactions.¹⁰ The reason for the strong interaction between barstar Tyr29 and barnase His102 is still unknown.²⁴ It cannot be explained as base-to-aromatic interaction.²⁵ We assigned the strong atom–atom interactions to neighboring atoms: CD1Tyr29–ND1His102, CE2Tyr29–CD2His102, and CE1Tyr29–NE2His102. The crystal structure of the barnase–barstar complex (PDB entry 1brs) was minimized by 200 ABNR steps by Charmm. A harmonic constraint with a force constant of 5 kcal mol⁻¹ Å⁻² was applied to all heavy atoms and distance-dependent dielectric constant was used in calculating electrostatic interactions. The distances between atoms that were considered to be involved in strong interactions were recorded and used as target distances for strongly interacted atoms in the modeled complexes. After minimization, the binding affinities of the scaffold proteins were evaluated by the scoring function. Table III lists calculated binding free energies for the four accepted scaffold proteins. 2sxl has the lowest calculated binding free energy among the accepted scaffold proteins except for 1bta. Interestingly, 2sxl also has a similar secondary structure to barstar at the interface. Barstar contacts with barnase by a helix and a loop. So does 2sxl. Figure 1 displays the complex structures of the four scaffold proteins and barnase.

DISCUSSION AND CONCLUSIONS

We have developed a strategy for grafting protein–protein interaction epitope based on molecular docking.¹⁰ At first, scaffold proteins are docked onto the receptor protein. Only matches with high docking scores are saved. For each saved match, a scaffold protein is accepted if it has suitable positions for grafting key interaction residues of the ligand protein. In this study, the searching order was reversed. The candidate residues in scaffold protein were identified at first. Unlike molecular docking, this strategy does not involve a 6-D search and the grid step is set to be very small (1 Å) in calculating complementary score between scaffold and receptor proteins.

Charged residues are paid more attention in our method due to their high desolvation energies. They usually form salt bridges between ligand and receptor proteins at the interface and are considered as key interaction residues. We believe it is important to maintain these salt-bridges during grafting. Even though a salt bridge is unfavorable in protein association, removing one charged residue will be more unfavorable since the large desolvation energy of the other residue can no longer be compensated by the

Table III Evaluations of the Minimized Complexes

Scaffold Protein	Calculated Binding Free Energy (kcal/mol)	Mean Distance Constraint Violations (Å) ^a
1bta	−20.1	0.05
1ten	−12.4	0.07
1tpg	−14.9	0.07
2sxl	−17.3	0.07

^a Mean distance violations from target distances for strongly interacted atom–atom pairs. The target distances were derived from native barstar–barnase complex structure.

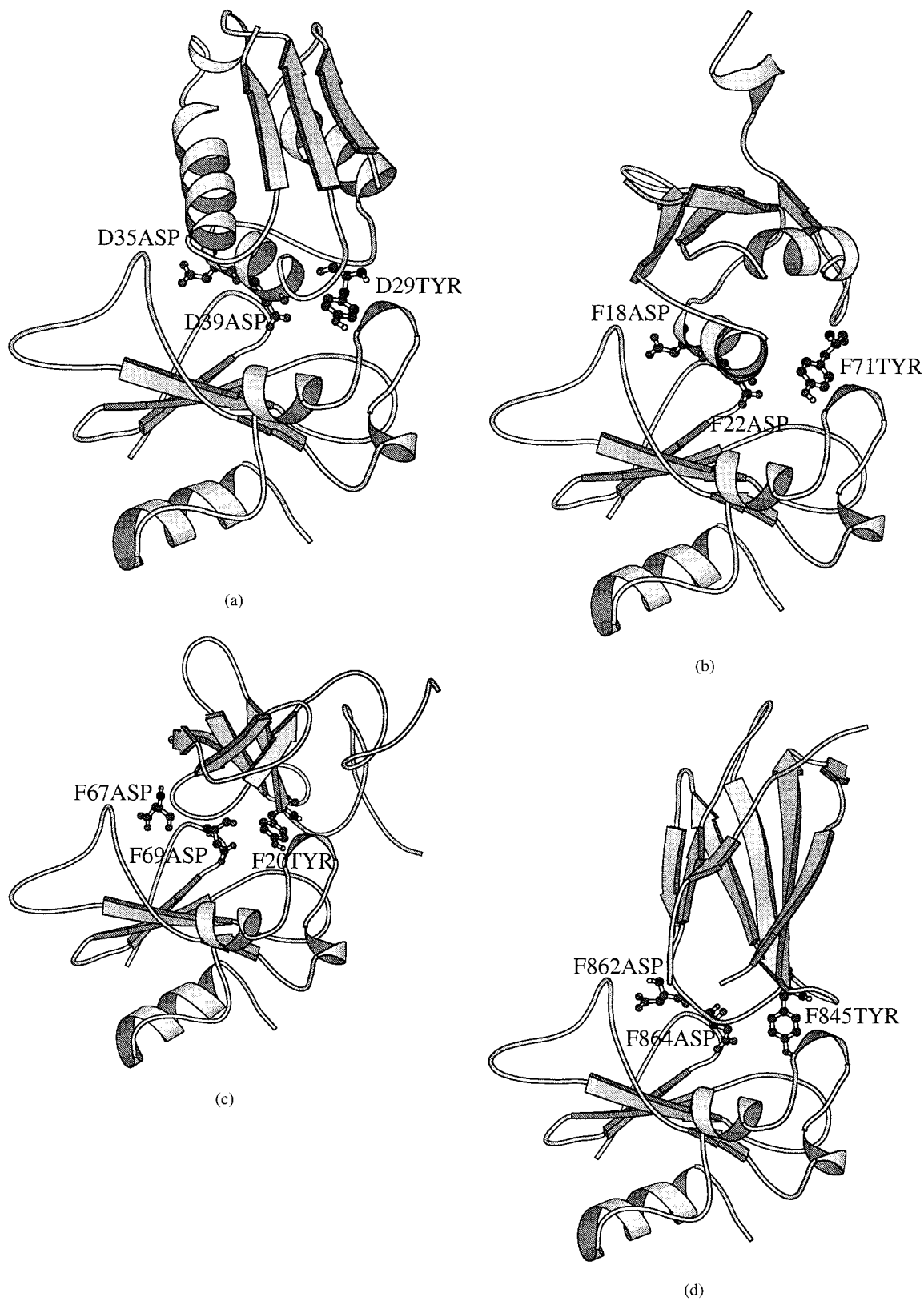


FIGURE 1 The modeled complex structures (a) barnase-1bta, (b) barnase-2sxl, (c) barnase-1tpg, and (d) barnase-1ten.

bridge energy. During refinement stage, buried charged residues of a scaffold protein, which do not form any salt bridge, are considered for mutation to reduce desolvation energy upon binding. However, noncharged hydrophilic atoms are not involved in such situation. Indeed, an analysis of a collection of 319 protein–protein interfaces showed that 17.4% of fully buried hydrophilic atoms do not form any hydrogen bond.²⁶ The geometry of hydrogen bonds across the protein interface is generally not optimal and has a wide distribution. Eisenberg et al. estimated atomic solvation parameters of amino acid residues as follows²⁷:

$$\Delta\delta(C) = 16 \text{ cal}\text{\AA}^2\text{mol}^{-1}$$

$$\Delta\delta(N/O) = -6 \text{ cal}\text{\AA}^2\text{mol}^{-1}$$

$$\Delta\delta(O^-) = -24 \text{ cal}\text{\AA}^2\text{mol}^{-1}$$

$$\Delta\delta(N^+) = -50 \text{ cal}\text{\AA}^2\text{mol}^{-1}$$

$$\Delta\delta(S) = 21 \text{ cal}\text{\AA}^2\text{mol}^{-1}$$

During protein association, the effect of the buried noncharged hydrophilic atoms, which do not form any hydrogen bond, may only weaken the favorable contribution of the neighboring buried hydrophobic atoms. The large buried surface at the protein–protein interface can accumulate favorable interactions and strengthen binding force. Thus, we only accept scaffold proteins with more than 1200 \AA^2 solvent accessible area at the interface and the effect of buried noncharged hydrophilic atoms was not considered. Indeed, most hydrophilic atoms at the interface of modeled complexes, which are impossible to form any hydrogen bond, may become solvent accessible just by regulating side chain dihedral angles.

It is very important that the interface of the complex between scaffold and receptor proteins has a reasonable packing density. When the two molecules contact too closely, the original structure may be destroyed during energy minimization. The number of buried hydrophilic pairs between scaffold and receptor proteins is also affected by the interface packing density, which plays an important role in the scoring function. Although it is nearly impossible to find a scaffold protein with an extremely complementary interface, reasonable packing density allows atoms to make small shifts and form a more complementary interface.

The program here may miss suitable scaffold proteins. In order to save computational time, residues

located on the surface and with an outward $C_\alpha \rightarrow C_\beta$ vector are selected at first. However, it is difficult to judge if a residue is solvent accessible. Some qualified residues may be omitted. The definition of two couples of residues as the same geometric relationship is very stringent in regard to the side-chain length of some large residues such as Arg, Lys, Glu, Trp, Phe, and Tyr. Discrete rotamers were used during molecular superposing. The conformations of the rotamers affect not only the superposing rms but also the relative position and the complementary score between scaffold and receptor proteins. Despite the disadvantage, this strategy is time-saving and practicable for searching data bases. Hellinga and Richards noted that their program, which was suitable for grafting small ligand binding sites, intended to search for some good scaffold proteins but not the best one in all possibility.²⁸ So does our strategy.

We had also tried another strategy.²⁹ It first identifies candidate residues in a scaffold protein. Then the scaffold protein is superposed onto the native ligand protein based on the C_α and C_β atoms correspondingly. Since this method does not need to check many rotamer combinations at each potential binding site, it runs very quickly. However, even a subtle difference in the orientations of the ligand and target $C_\alpha \rightarrow C_\beta$ vectors lead to a large deviation of atoms at side-chain terminal, which usually are supposed to strongly interact with the receptor protein. The large rms difference between key interaction atoms of ligand and scaffold proteins cannot always be reduced to an acceptable level by regulating side-chain dihedrals. On the other hand, suitable scaffold proteins may be skipped because of a low complementary score. A scaffold protein could be put at an unfitting position in order to achieve a minimal rms difference during superposition though a larger superposing rms difference is acceptable in some cases. So in this study, we superposed scaffold proteins onto the ligand protein based on key interaction atoms.

This work was supported by the Department of Science and Technology of China and the Natural Science Foundation of China.

REFERENCES

1. Wharton, R. P.; Ptashne, M. *Nature* 1985, 316, 601–605.
2. Jones, P. T.; Dear, P. H.; Foote, J.; Neuberger, M. S.; Winter, G. *Nature* 1986, 321, 522–525.
3. Hedstrom, L.; Szilagyi, L.; Rutter, W. *Science* 1992, 255, 1249–1253.

4. Janin, J. *Proteins* 1995, 21, 30–39.
5. Cunningham, B. C.; Wells, J. A. *J Mol Biol* 1993, 234, 554–563.
6. Molina, M. A.; Marino, C.; Oliva, B.; et al. *J Biol Chem* 1994, 269, 21467–21472.
7. Clackson, T.; Wells, J. A. *Science* 1995, 267, 383–386.
8. Björk, I.; Brieditis, I.; Raub-Segall E.; Pol, E.; et al. *Biochemistry*, 1996, 35, 10720–10726.
9. Covell, D. G.; Wallqvist, A., *J Mol Biol* 1997, 269, 281–297.
10. Liang, S. D.; Li, W. Z.; Xiao, L.; Wang, J. S.; Lai, L. H. *J Biomol Struct Dynam* 2000, 17, 821–828.
11. Hubbard, S. J.; Thornton, J. M. NACCESS, computer program, Department of Biochemistry and Molecular Biology, University College London, 1993.
12. Brint, A. T.; Willett, P. *J Mol Graph* 1987, 5, 49–56.
13. Maeyer, M. D.; Desmet, J.; Lasters, I. *Folding Design* 1997, 2, 53–66.
14. Ferro, D. R.; Hermans, J. *Acta Cryst* 1977, A33, 345–347.
15. Katchalski-Katzir, E.; et al. *Proc Natl Acad Sci USA* 1992, 89, 2195–2199.
16. Richards, F. M. *J Mol Biol* 1974, 82, 1–14.
17. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; et al. *J Comput Chem* 1983, 4, 187–217.
18. Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; et al. *Biophys J* 1992, 61, A143.
19. Schreiber, G.; Fersht, A. R. *Biochemistry* 1993, 32, 5145–5150.
20. Fersht, A. R. *FEBS Lett* 1993, 325, 5–16.
21. Hartley, R. W. *Biochemistry* 1993, 32, 5978–5984.
22. Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* 1994, 33, 8878–8889.
23. Schreiber, G.; Fersht, A. R. *Structure* 1994, 2, 945–951.
24. Schreiber, G.; Fersht, A. R. *J Mol Biol* 1995, 248, 478–486.
25. Burley, S. K.; Petsko, G. A. *FEBS Lett* 1986, 203, 139–143.
26. Xu, D.; Tsai, C. J.; Nussinov, R. *Protein Eng* 1997, 10, 999–1012.
27. Eisenberg, D.; McLachlan, A. *Nature* 1986, 319, 199–203.
28. Hellinga, H. W.; Richards, F. M. *J Mol Biol* 1991, 222, 763–785.
29. Liang, S. D.; Xiao, L.; Mao, F. L.; Jiang, L.; Han, Y. Z.; Lai L. H. *Chinese Sci Bull* 2000, in press.