Protein Engineering vol.9 no.9 pp.741–744, 1996

Main-chain complementarity in protein–protein recognition

Ilya A.Vakser

Center for Molecular Design, Washington University, Box 1099, St Louis, MO 63130, USA

Present address: The Rockefeller University, Box 270, 1230 York Avenue, New York, NY 10021, USA

The existing theoretical approaches to protein–protein recognition concentrate on the details of the molecular surface at atomic resolution, while a possible role of the main chain in complex formation has been largely unexplored. To address this problem, we represented the molecules by Cα atoms and applied the step-function potentials for intermolecular energy calculations. Since our goal was not to predict, as accurately as possible, the structure of a protein–ligand complex, but to reveal the role of the backbone in the formation of such a complex, all the potentials were identical and Cα centered. Thus, for the specific purposes of our study, we do not simulate the difference in the side chains at the molecular surface. The structures were taken from known co-crystallized complexes. The intermolecular energy calculation was performed by a systematic 6-D search on a grid. The results revealed that in all cases tested (except antigen–antibody) the positions of the ligand at the binding site on the receptor corresponded to the lowest-energy configurations of the complex. The complementarity between the backbones, in general, may facilitate the initial placement of the ligand at the binding site of the receptor. At the same time, the identity and the specific conformation of the surface side chains play a crucial role in the next stage of the complex formation.

Keywords: docking algorithm/ligand–receptor interaction/long-range potentials/molecular complexes/protein structure

Introduction

The functions of proteins are determined by their interactions with other molecules. Thus, it is important to understand the principles of complex formation at the structural level. A possible role of various structural elements in protein–protein recognition may be considered as one of such principles. The concept of surface complementarity at protein–protein interfaces is presently well established. It is supported by the observation of X-ray structures of protein complexes (see, for example, Lawrence and Colman, 1993), as well as by the successful application of geometry-based docking procedures, when the steric fit is a prerequisite of the physicochemical complementarity [for example, see Shoichet and Kuntz, 1991; Katchalski-Katzir *et al.*, 1992; Vakser and Aflalo, 1994; Fischer *et al.*, 1995; for a review of the available docking techniques, see Blaney and Dixon (1993), Cherfils and Janin (1993), Kollman (1994), Kuntz *et al.* (1994) and Lybrand (1995)]. Obviously, there are two structural factors which create the complementarity between the molecular surfaces: (i) the main-chain fold and (ii) the identity and the conformation of the side chains on the surface. These factors may be correlated to a certain degree; however, the backbone conformation is believed to be determined mostly by the core residues (Matthews, 1993). The existing theoretical approaches concentrate on the second factor in protein–protein recognition (details of the molecular surface at atomic resolution), while a possible role of the main chain in complex formation has been largely unexplored.

Our recent docking approach (Vakser, 1995, 1996b), specifically designed for low resolution (~7 Å) structures, suggested that the elements of the general fold are important components in protein–protein recognition. In that study we demonstrated that the systematic grid search for possible binding modes of molecules, deprived of any structural details below the 7 Å level, still retrieves most of the structural features (position of the ligand and orientation of its binding site) of the correct configuration of the complex. Molecules, represented by non-hydrogen atoms, were projected on a sparse grid (with the grid-step of ~7 Å), which guaranteed that the details of the structure below the step of the grid were eliminated. The docking was performed between these low-resolution molecular images by a correlation technique, with a scan of the ligand's orientations, which is equivalent to a systematic search in six dimensions (three translations and three rotations of the ligand). Later, we showed that this procedure is formally equivalent to an intermolecular energy calculation with long-range step-function potentials (Vakser, 1996a).

In our present paper we investigate the role of the molecular backbone structure in protein–protein recognition. To address this problem directly, we represented the molecules by Cα atoms only and applied the Cα-centered potentials for intermolecular energy calculations. Similar simplified interactions have become quite common in protein structure prediction (Wodak and Rooman, 1993). For the protein–ligand interactions, Levitt's (1976) residue–residue potentials were applied by Wodak and Janin (1978). These potentials were designed to distinguish between different side chains to approximate the full atom–atom energy function. Since our goal was not to predict as accurately as possible, the structure of the protein–ligand complex, but to reveal the role of the backbone in the formation of such a complex, we made all the potentials identical and Cα centered. Thus, for the specific purposes of our study, we do not simulate the difference in the side chains at the molecular surface. A systematic 6-D search for complementarity between the ligand and receptor backbone structures revealed that, in most cases, the low-energy configurations of the complexes are non-randomly related to their crystal structures.

Methods

In our procedure, all atoms, except Cαs, were deleted from the molecular structures. The principal component in energy

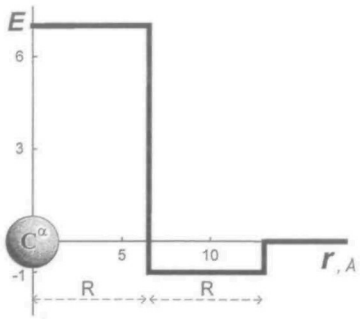

Fig. 1. A step-function potential between Cα atoms of the ligand and the receptor. The units of the potential are arbitrary. *R* is the range of both repulsive and attractive parts.

calculations between the backbone structures is the form of the potential function. The application of functions similar to the Lennard–Jones potential, which is equivalent to a close contact between Cα atoms of the ligand and the receptor, corresponds to a random distribution of the ligand around the receptor (I.A.Vakser, unpublished). The range of the repulsion part of the potential has to be long enough to keep the molecules apart, at a distance which corresponds approximately to the presence of side chains. At the same time, long ranges of both repulsion and attraction parts smooth the energy profile by averaging the contributions of neighboring Cα atoms. For our potential we chose a simple step-function form (Figure 1). The optimal values for *E* (repulsion) and *R* were determined as 7.0 and 6.6 respectively. Both the decrease and the increase in these values corresponded to a more scattered (eventually random) distribution of the low-energy configurations.

The systematic search procedure for the intermolecular energy calculation is based on our correlation algorithm for protein surface recognition (Katchalski-Katzir *et al.*, 1992) which was later extended for partial molecule representations (Vakser and Aflalo, 1994), low-resolution structures (Vakser, 1995, 1996b) and reinterpreted in terms of energy potentials (Vakser, 1996a). In our present study of backbone structures we use the same formalism. Briefly, the potential is digitized on a 3-D grid (with the grid-step of *R*) around the receptor (Figure 2). The ligand (at a given orientation) is shifted relative to the receptor in three spatial coordinates, with an interval of *R*. At each given ligand's position, the intermolecular energy is calculated according to the Cα–Cα step-function potential. Technically, this is done by projecting the ligand onto a similar grid, with unity values in the grid points around each Cα atom. Thus, if a point on the ligand's grid is closer than *R* to *N* ligand's Cα atoms, its value is assigned as *N*. Then a 3-D correlation procedure is applied to the receptor's and the ligand's grids and the values in the resulting 3-D correlation matrix are retrieved. Numerically, these values are equivalent to intermolecular energy values in regular atom–atom calculations (see Vakser, 1996a). Thus, the ligand–receptor intermolecular energy is systematically evaluated in the *x*, *y* and *z* coordinates (Figure 2). In addition to that, the three angular coordinates α, β and γ (Figure 2) are tried with a pre-set angular step. The value for this step was chosen as 20°, since smaller values did not improve the results. This angular step corresponded to 2142 uniformly distributed ligand's orientations. Since every orientation was combined with the systematic *x*, *y* and *z* search, the procedure is equivalent to an exhaustive grid search in six dimensions. The correlation technique for the intermolecular energy evaluation is much faster than the regular atom–atom

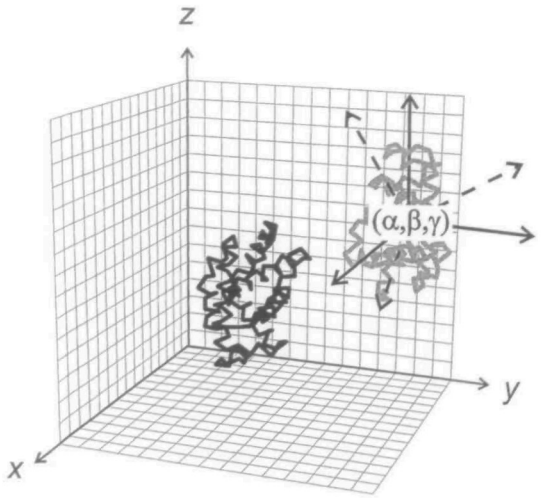

Fig. 2. The system of coordinates for the docking of backbone structures. The molecules shown are the α-subunit (receptor, black) and the β-subunit (ligand, gray) of human hemoglobin. For illustrative purposes, the grid is shown only in *xy*, *yz*, *zx* planes. The size and the spacing of the grid (relative to the size of the molecules) are characteristic of those actually employed in the docking procedure.

calculations; thus the full 6-D search for a pair of molecules took less than 1 min of c.p.u. time on an SGI workstation.

Results and discussion

We tested the backbone structures from co-crystallized complexes taken from the Protein Data Bank (Abola *et al.*, 1987). The structures were α- and β-subunits of human hemoglobin (2HHB; Fermi *et al.*, 1984), trypsin and BPTI (2PTC; Marquart *et al.*, 1983), subtilisin and chymotrypsin inhibitor (2SNI; McPhalen and James, 1988), acid proteinase and peptide inhibitor (3APR; Suguna *et al.*, 1987), α1–α2 subunits of MHC I and a peptide (1HSA; Madden *et al.*, 1992), the variable region of Fab and lysozyme (2HFL; Sheriff *et al.*, 1987) and the variable region of Fab and a peptide (1GGI; Rini *et al.*, 1993). Each pair of the backbone structures was subject to a systematic energy evaluation on a grid, using the Cα–Cα potential, as described above.

An example of such an evaluation for the hemoglobin subunits is shown in Figure 3. Since the search is performed in six dimensions (*x*, *y*, *z*, α, β and γ coordinates), we can show only a 2-D cross-section through the actual 6-D grid. For illustrative purposes, the actual backbone structure of the α-subunit (receptor) is overlapped with the grid. The intermolecular energy values on the grid are in the positions of the ligand's gravity center. At longer distances from the receptor, the energy values are zero, which corresponds to the nature of our Cα–Cα step-function potential. Closer to the receptor, the energy becomes negative. Large positive values correspond to severe overlaps between the ligand and the receptor backbone structures. As can be seen, the global energy minimum is quite close (within the accuracy of our discrete space representation) to the actual position of the β-subunit gravity center in the co-crystallized complex. A systematic shift of the global minimum position (compared with its 'experimental' position in the gravity center of the co-crystallized ligand) towards the receptor, was detected for all the complexes. We attribute this to a simple step-function character of our Cα–Cα potential. This effect is similar to what was observed earlier for all-atom molecules (Vakser, 1996a).

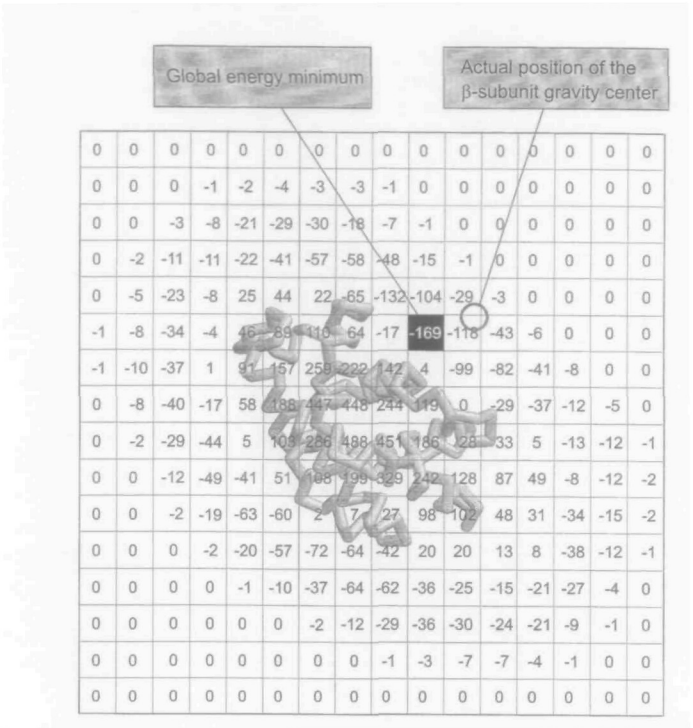

Fig. 3. The results of the systematic intermolecular energy evaluation for the α and β subunits of human hemoglobin. Only a two-dimensional cross section through the full six-dimensional table of energy values is shown. Each energy value corresponds to a certain position of the ligand and is placed in the ligand's gravity center. The cross section contains the global minimum of the intermolecular energy and the gravity centers of the ligand and the receptor backbone structures. For illustrative purposes, the 3-D receptor backbone structure (in the actual orientation and with the correct relative size) is overlapped with the energy values. For better visualization, the numbers which represent the intermolecular energy are 10 times smaller than the actual values (in arbitrary units) obtained according to the Cα–Cα potential. The details of the search procedure are described in the text.

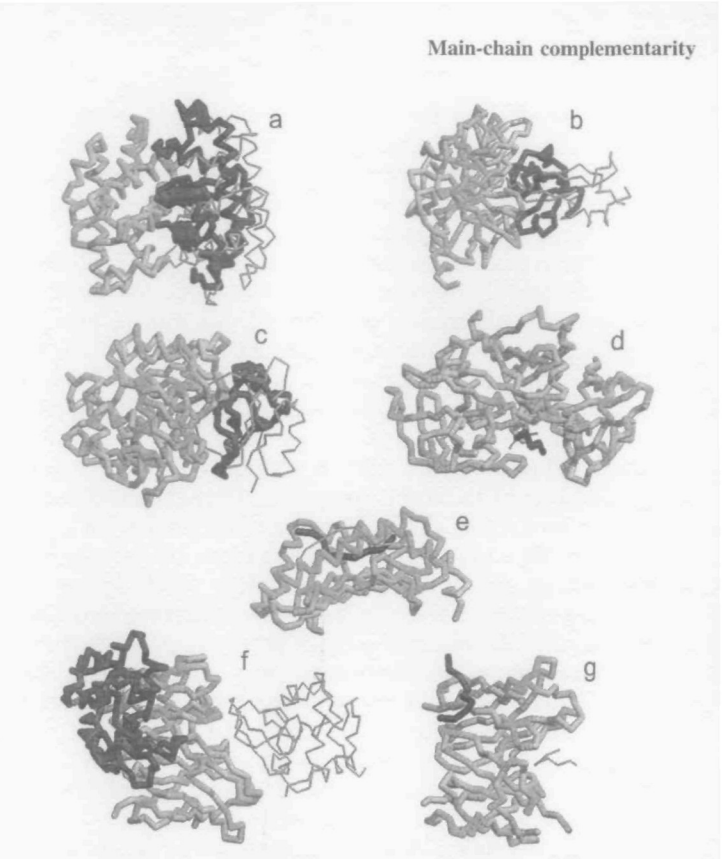

Fig. 4. The lowest-energy complexes between the backbone structures. The molecular pairs are: α and β subunits of human hemoglobin (**a**), trypsin and BPTI (**b**), subtilisin and chymotrypsin inhibitor (**c**), acid proteinase and peptide inhibitor (**d**), α1–α2 subunits of MHC I and a peptide (**e**), the variable region of Fab and lysozyme (**f**), and the variable region of Fab and a peptide (**g**). The thick cylinders represent the receptor (light gray) and the ligand (dark gray). The backbone of the ligand in the crystal structure is shown by the thin cylinders. The Figure was prepared using the program RasMol (R.Sayle, Greenford, UK).

For all the complexes, we analyzed a set of 10 low-energy configurations. The lowest-energy structures are shown in Figure 4. Hemoglobin (Figure 4a) was taken as an example of multisubunit proteins. From 10 low-energy positions of the β-subunit, six (with different orientations) were found at the binding site of the α-subunit (α1–β1 contact). It is worthwhile to note that the overall number of possibilities for placing the β-subunit in contact with the α-subunit, using our grid, is more than 10^6 (the same applies to other complexes, too). The lowest-energy configuration of the complex is the closest to the X-ray structure. In comparison with the crystal structure coordinates, the β-subunit is positioned within the binding region of the α-subunit and rotated by ~35°.

Trypsin–BPTI and subtilisin–chymotrypsin inhibitor are examples of enzyme–inhibitor complexes where the ligand (inhibitor) is a small protein. In both cases, all low-energy configurations corresponded to the ligand position within the binding site of the receptor, with different orientations. In the lowest-energy configuration of the trypsin–BPTI complex (Figure 4b), the backbone structure of BPTI is rotated by approximately 70° relative to the orientation from the crystal structure of the complex. In the case of the subtilisin–chymotrypsin inhibitor complex, the discrepancy in orientation is even larger (Figure 4c).

For the acid proteinase–peptide inhibitor complex, in all low-energy configurations of the complex between the backbone structures, the ligand was found in the receptor binding site. The main axis of the ligand (peptide in stretched conformation) coincided approximately (with angular difference close to 0 or 180°) with the axis of the ligand in the crystal structure. However, the spin angle around this axis was different. In the lowest-energy configuration (Figure 4d), the ligand is shifted by ~10 Å along the main axis.

A similar behavior of the backbone structures was observed in the case of a peptide and the recognition part of MHC I. In all low-energy configurations, the ligand was found in the binding groove of MHC I. In the lowest-energy configuration (Figure 4e), the main axis of the peptide coincided with that of the crystal structure (the direction is the opposite). The spin angle around this axis was different.

The results of backbone matching for antigen–antibody pairs were quite different from the results for the other complexes. In all low-energy configurations, the ligand (antigen) was found outside the antibody's binding site (the lowest-energy complexes of antibody–lysozyme and antibody–peptide are shown in Figure 4f and g). Since the Fc domains were excluded from the receptor structure, we considered a possibility that these ligand positions were an artifact of the Fc separation. The interface with Fc created low-energy false-positive matches in our previous study with full-atom representation of molecules (Vakser, 1996a). In the case of Cα structures, however, the graphical analysis did not support this possibility, since the ligands did not overlap with the Fc domains in the corresponding X-ray structures.

As we indicated earlier, our procedure is a blind, exhaustive search on a grid, through all six degrees of freedom of

the ligand relative to the receptor. In each grid point, the intermolecular energy is calculated according to the identical Cα–Cα step-function potentials. The number of possible configurations of a complex in our grid search, where the ligand and the receptor are in close contact, is more than 10^6. If we consider the ligand regardless of its orientation, the number of possibilities will still be more than 10^3. If we assume that the backbone conformation does not play a role in protein recognition, we may expect the predicted ligand positions to be random. Thus, the probability of finding the ligand within two grid steps from the crystallographically determined position must be very small (~0.03). A statistically significant deviation of this probability from zero would indicate a certain recognition role of the main chain. Our computer experiment revealed, however, that all backbone structures (except antigen–antibody) in all 10 low-energy configurations (in the case of hemoglobin, in six out of 10), were found within less than two grid steps from the crystallographically determined position in the complex. Taking into account such a remarkably non-random character of the results, we may conclude that the main-chain fold plays an important role in protein recognition.

At the same time, the results show that the role of the main chain in antigen–antibody complexes is, probably, less significant than in the other cases of protein complexes. A possible reason may be that the antibody molecules, with basically the same main-chain fold, have to recognize different antigens. This means that the backbone cannot be a recognition factor in this case. The conformational differences in the main chain of the recognition loops in the variable domain of Fab may just facilitate the specific arrangement of the side chains, which could reflect certain differences in the principles of complex formation.

We may conclude that the complementarity between the backbones, in general, may facilitate the initial placement of the ligand at the binding site of the receptor. At the same time, the identity and the specific conformation of the surface side chains play a crucial role in the next stage of the complex formation.

Acknowledgements

I wish to thank Professor Garland R.Marshall for his support and helpful discussions. I am grateful to Professor Denise D.Beusen for her valuable comments.

References

Abola,E.E., Bernsein,F.C., Bryant,S.H., Koetzle,T.L. and Weng,J. (1987) In Allen,F.H., Bergerhoff,G. and Sievers,R. (eds), *Crystallographic Databases—Information Content, Software Systems, Scientific Applications.* Data Commission of the International Union of Crystallography, Bonn, pp. 107–132.

Blaney,J.M. and Dixon,J S. (1993) *Perspect. Drug Disc. Des*, **1,** 301–319.

Cherfils,J. and Janin,J. (1993) *Curr. Opin. Struct. Biol.*, **3,** 265–269.

Fermi,G., Perutz,M.F., Shaanan,B. and Fourme,R. (1984) *J. Mol. Biol.*, **175,** 159–174.

Fischer,D., Lin,S.L., Wolfson,H.L. and Nussinov,R. (1995) *J Mol. Biol.*, **248,** 459–477.

Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) *Proc. Natl Acad. Sci. USA*, **89,** 2195–2199.

Kollman,P.A. (1994) *Curr. Opin. Struct. Biol.*, **4,** 240–245.

Kuntz,I.D., Meng,E.C. and Shoichet,B.K. (1994) *Acc. Chem. Res.*, **27,** 117–123.

Lawrence,M.C. and Colman,P.M. (1993) *J. Mol. Biol*, **234,** 946–950.

Levitt,M. (1976) *J. Mol. Biol.*, **104,** 59–107.

Lybrand,T.P. (1995) *Curr. Opin. Struct. Biol.*, **5,** 224–228.

McPhalen,C.A. and James,M.N.G. (1988) *Biochemistry*, **27,** 6582–6598.

Madden,D.R., Gorga,J.C., Strominger,J.L. and Wiley,D.C. (1992) *Cell*, **70,** 1035–1048

Marquart,M., Walter,J., Deisenhofer,J., Bode,W. and Huber,R. (1983) *Acta Crystallogr., Sect. B*, **39,** 480–490.

Matthews,B.W. (1993) *Annu. Rev. Biochem.*, **62,** 139–160.

Rini,J.M., Stanfield,R.L., Stura,E.A., Salınas,P.A., Profy,A.T. and Wilson,I.A. (1993) *Proc. Natl Acad. Sci. USA*, **90,** 6325–6329.

Sheriff,S., Silverton,E.W., Padlan,E.A., Cohen,G.H., Smith-Gill,S.J., Finzel,B C. and Davies,D.R. (1987) *Proc. Natl Acad. Sci. USA*, **84,** 8075–8079.

Shoichet,B.K. and Kuntz,I.D. (1991) *J. Mol. Biol.*, **221,** 327–346.

Suguna,K., Bott,R.R., Padlan,E.A., Subramanian,E., Sheriff,S., Cohen,G.H. and Davies,D.R. (1987) *J. Mol. Biol.*, **196,** 877–900.

Vakser,I.A. (1995) *Protein Engng*, **8,** 371–377.

Vakser,I.A. (1996a) *Protein Engng*, **9,** 37–41.

Vakser,I.A. (1996b) *Biopolymers*, in press.

Vakser,I.A. and Aflalo,C. (1994) *Proteins*, **20,** 320–329.

Wodak,S.J. and Janin,J. (1978) *J. Mol. Biol.*, **124,** 323–342.

Wodak,S.J. and Rooman,M.J. (1993) *Curr. Opin. Struct. Biol.*, **3,** 247–259.

Received February 4, 1996; revised April 28, 1996; accepted May 7, 1996