Protein Engineering vol.9 no.1 pp.37–41, 1996

Long-distance potentials: an approach to the multiple-minima problem in ligand–receptor interaction

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 multiple-minima problem is a classical problem in molecular structure prediction. For ligand–receptor systems, a possible direction to alleviate this major obstacle is to simplify the objective function (intermolecular energy) and smooth its profile. We introduce long-distance atom–atom potentials for ligand–receptor interactions. The longer ranges result in averaging of the energy potential at a given point. Our simplified force field is based on a trivial empirical representation of interatomic interactions as a step function. We demonstrate that the intermolecular energy calculation by a systematic search with such a simplified long-distance force field delivers the global minimum (crystallographically determined position of the ligand) by radically suppressing local minima (or false-positive fits). The effectiveness of the approach is demonstrated on different molecular complexes of known structure.

Keywords: docking algorithms/drug design/energy minimization/molecular recognition/protein structure

Introduction

Interactions between biological molecules are a complicated process determined by a variety of structural and physicochemical factors. Recent advances in theoretical approaches to this problem are based on the analysis of energy in macromolecular (ligand–receptor) associations [for an overview, see Janin (1995)] as well as on sometimes more phenomenological, docking techniques [for a review of presently available approaches, see Blaney and Dixon (1993), Cherfils and Janin (1993), Kollman (1994) and Kuntz *et al.* (1994)]. Many of these techniques implement a search for 'the best fit' between molecular surfaces (geometrical and/or physicochemical match). This is justified by the existing experimental evidence on the structures of co-crystallized complexes. However, the difference between the approaches based on explicit intermolecular energy minimization and those based on the 'surface fit' considerations is often just a difference in terms. Indeed, a geometric fit between molecular surfaces is usually a match between van der Waals representations of the surface atoms, which corresponds to the minimum in van der Waals atom–atom energy. Similar considerations may apply to the physicochemical complementarity. Of course, the energy of ligand–receptor associations is much more complicated than a score of a simple steric or physicochemical fit. There is, however, a contradiction between a more faithful, 'realistic' description of intermolecular interaction and the feasibility (in reasonable computational time) of the global minimum determination for macromolecular systems. Both the elaborate character of the objective function and the multiplicity of minima contribute to the complexity of the problem.

The multiple-minima problem is a classical problem for intramolecular structure prediction. The same applies, of course, to the intermolecular energy calculations. A possible direction to alleviate this major obstacle is to simplify the objective function (intermolecular energy) and smooth its profile. Many of the existing docking approaches, in fact, are designed specifically for this purpose. Development of the procedures based on molecular surface fit could often be interpreted as an attempt (sometimes successful) to substitute the 'real' intermolecular energy by a rough simplification of some of its components (e.g. steric, electrostatic or hydrophobic). However, the multiple-minima problem (or, in alternative terms, the multiplicity of false-positive fits) still presents a serious obstacle for docking studies. This is especially true in 'the general case', in which no binding sites are predefined on both the ligand and receptor, and is further complicated by structural inaccuracies and conformational changes upon complex formation, both local and, possibly, global.

An important direction towards the solution of the multi-minima problem in conformational analysis of macromolecules has been developed by Scheraga and co-workers (Piela *et al.*, 1989). The approach suggests the use of the diffusion equation formalism to smooth the energy profile radically. Such a smoothing process is correlated with the expansion of ranges for interatomic potentials (Wawak *et al.*, 1992). In our work, we explicitly introduce long-distance atom–atom potentials for ligand–receptor interactions. The longer ranges result in 'averaging' of the energy potential at a given point. We do not make use of mathematical formalism to determine the parameters of the expanded potentials. Our simplified 'force field' is based on a trivial empirical representation of interatomic interactions as a step function, which corresponds to our previous description of low-resolution discrete molecular images (Vakser, 1995). We demonstrate that the intermolecular 'energy' calculation by a systematic search with such a simplified long-distance force field delivers the global minimum (crystallographically determined position of the ligand) by radically suppressing local minima (or false-positive fits). The effectiveness of the approach is demonstrated on different molecular complexes of known structure.

Methods

A molecular recognition algorithm based on a correlation technique, which takes advantage of the Fourier transformation to reduce the computational time drastically, was introduced earlier by Katchalski-Katzir *et al.* (1992). It implements a search for the maximum geometric surface overlap for high-resolution structures. Since then, we have developed new approaches designed to simplify the objective function and improve the signal-to-noise ratio in ligand–receptor matching. One of our methods (Vakser and Aflalo, 1994) takes advantage

of the solvation properties of the molecular structures. It leaves only hydrophobic regions on the molecular surfaces, which improves the relative weight of the real match. The other method (Vakser, 1995) goes much further by eliminating all structural details below 7 Å resolution. This radical step demonstrated the possibility of predicting the configuration of the complex between low-resolution structures. All these approaches are based on phenomenological concepts of surface complementarity or atomic density at the ligand–receptor interface. Matching molecules by the correlation technique allows, however, a clear interpretation in terms of energy potentials. This definitely has to provide a more adequate understanding of the docking approach. What is even more important is that it may yield better clues for future developments and applications, according to the natural properties of intermolecular interactions. Calculation of interatomic energy by a correlation technique and Fourier transformation was described by Harrison *et al.* (1994). Our approach (Vakser, 1995) of averaging atomic contributions by a sparse-grid representation of structures, when redesigned in terms of energy potentials, provides an important opportunity of addressing directly the problem of eliminating local minima in a search for the global energy minimum.

From a formal, mathematical point of view the procedure for energy calculation is similar to our 'low-resolution docking' technique. This approach is described elsewhere (Vakser, 1995), along with details of the algorithm and its implementation. Here we present a brief description of this technique, as applied to the calculation of ligand–receptor interaction energy. The approach is illustrated in Figure 1. We use step-function potentials for the energy of atom–atom interactions (Figure 1a). As in 'regular' energy calculations, this potential is additive (the value at any point is the sum of contributions of all atoms within the interaction range). The potential has two parameters: R is the width of the negative energy well and the range of repulsion (positive energy) and U is the energy of repulsion (the energy of attraction is always -1). Everywhere beyond the $2R$ distance from an atom the energy is 0. A ligand's atoms are represented by unity values. The potentials for all atoms of the receptor are projected, one by one, onto a 3-D grid with a grid step of R. At the beginning of the projection procedure all values on the grid are zero. During the projection, if the distance between the center of an atom and a given point on the grid is shorter than or equal to R, the value of this grid point is increased by U. If this distance is greater than R but shorter than $2R$, the added value is -1. The ligand molecule is projected in the same way, on a similar grid. In this case, if the distance between the center of an atom and a grid point is shorter than or equivalent to R, the added value is 1.

The correlation technique allows a rapid calculation of intermolecular energy at a given ligand's orientation. The 3-D grids for the receptor and the ligand are identical (the same interval of the grid and the same number of grid points). At each step of the correlation procedure, the values in the equivalent points of the first and second grids are multiplied and the resulting values of all these multiplications are summed up. This gives a number which is the energy of the ligand–receptor interaction for a given ligand position. Figure 1b and c illustrates the main principle of this calculation, using our step-function potentials. There are cases where the positive or the negative energy contribution of an atom is represented by more than one grid point. Thus, the attraction or repulsion

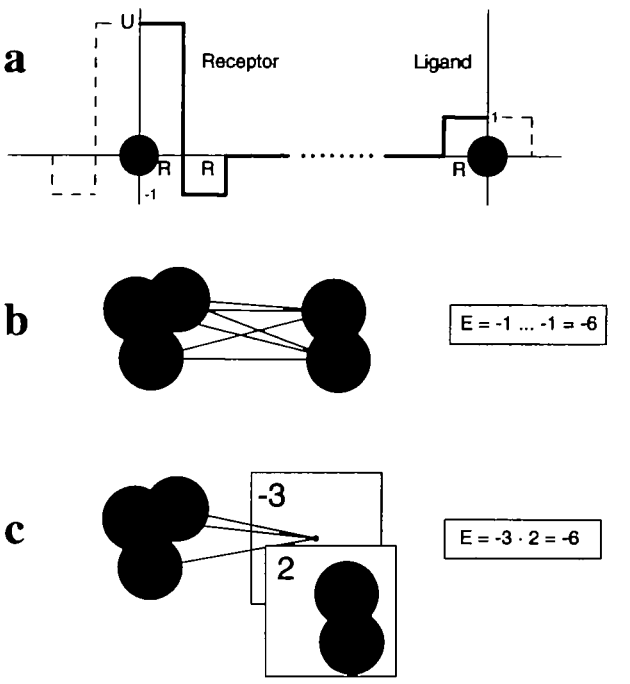

Fig. 1. Intermolecular energy calculation by the correlation technique. (**a**) The energy of the atom–atom interaction is decomposed into a step-function potential around the receptor atoms and a similar distribution around the ligand atoms. Both functions are digitized on a 3-D grid with grid spacing R. Shadowed circles schematically represent atoms. As an illustration of the technique, (**b**) the 'regular' calculation of intermolecular energy with the step-function potential is compared with (**c**) the energy calculation by the correlation technique. The squares are the cross-sections through single elements of the 3-D grids. According to the algorithm, the point on the ligand's grid (center of the white square) is given a value of 2 (two atoms within the range of R). The point on the receptor's grid (center of the gray square) is given a value of -3 (three atoms within the range between R and $2R$). When multiplied (for a description of the correlation, see the text) these values give the same energy E as in the regular energy calculation.

between some atom pairs is accounted for more than once, which contradicts the principles of the atom–atom empirical energy calculations. However, the distribution of such 'double' representations in the molecular structure is random. Thus, given the very limited accuracy of our potential function, which is an extremely rough approximation of 'regular' atom–atom potentials, this discrepancy can be neglected. At the next step of the procedure, the ligand is moved by one grid step further and the same process is repeated. Eventually all ligand positions on the grid have to be assessed. This could be done very efficiently by Fourier transformation [for mathematical details, see Katchalski-Katzir *et al.* (1992)]. All possible ligand orientations are sampled with a predefined angular step. Thus, the whole procedure is equivalent to a systematic search for low-energy configurations of the ligand–receptor complex, through all six degrees of freedom (three translational and three rotational). The full-search c.p.u. time on an SGI workstation decreases from more than 12 h for the short-range potentials (R = ~2 Å), to less than 1 min for the long-range potentials (R = ~7 Å).

Results and discussion

The parameters R and U of the potential (Figure 1a) provide natural means of varying the range of the atom–atom interactions. We performed a series of calculations on ligand–receptor systems (co-crystallized complexes from the PDB;

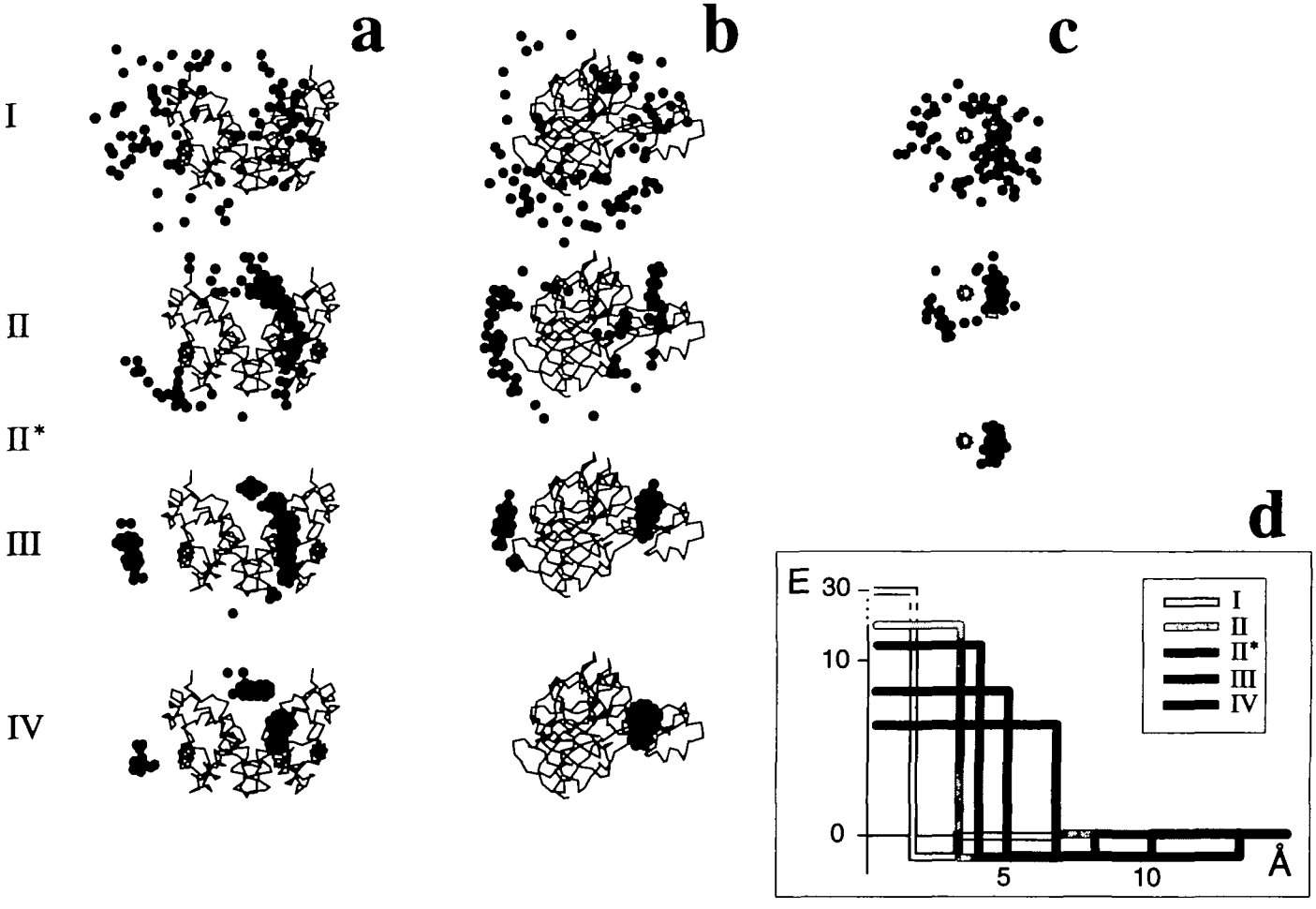

Fig. 2. Low-energy configurations of molecular complexes. The first 100 lowest-energy positions of the ligand are shown for (**a**) hemoglobin α- and β-subunits, (**b**) trypsin and BPTI and (**c**) helices 1 and 2 from the L-subunit of the photosynthetic reaction center. Each pair of molecules is shown in the crystallographically determined configuration. The black circles represent positions of the ligand's center of gravity. The ligand systematically appears closer to the receptor when long-range potentials are applied, due to the character of the energy function. In cases when several low-energy ligand positions (with different orientations) were found at the same point, the corresponding circles were separated by ~0.3 Å for better visualization. The potentials (d) for the examples were varied from R = 1.7 Å to R = 6.8 Å.

Abola *et al.*, 1987), with ranges gradually expanded from those characteristic of Lennard–Jones potentials (R = ~2 Å) up to long ranges of R beyond 7 Å. We tried a multisubunit protein (2HHB; Fermi *et al.*, 1984), enzyme–inhibitor complexes (2PTC, Marquart *et al.*, 1983; 1CHO, Fujinaga *et al.*, 1987; 2SNI, McPhalen and James, 1988; 3APR, Suguna *et al.*, 1987), antigen–antibody complexes (2HFL, Sheriff *et al.*, 1987; 1GGI, Rini *et al.*, 1993) and transmembrane helices from the L-subunit of the photosynthetic reaction center (1PRC, Deisenhofer and Michel, 1989). The ligands were placed in an arbitrary position and intermolecular energy evaluation was performed by the systematic search algorithm described above. The low-energy configurations of the complex were compared with those known from the X-ray data.

A regular practice to limit a set of low-energy structures is to include those with energies within a certain interval from the lowest energy. However, because of the schematic, empirical character of our potentials, in particular at longer ranges, we preferred to use the criterion of an absolute number of low-energy structures. For comparative purposes, in all complexes and for all ranges of the potential, we selected the first 100 lowest-energy configurations. The range R in the step-function potential, which is identical with the corresponding interval R of the grid, imposes the limit of accuracy for the calculated positions of a ligand. Moreover, the 'smoother' the energy profile (at values of R beyond the characteristic van der Waals radii), naturally, the fewer restrictions apply to the ligand's orientation. For most of the tested complexes, at R values close to 7 Å, the ligands displayed a distinct preference for correct reorientation of their binding sites towards that of the receptor. At the same time, even when placed at the exact, crystallographically determined position, these ligands were capable of virtually free rotation around the axis which connects the centers of gravity of the molecules through their binding sites. Because of these factors, we did not apply the criterion of r.m.s. deviation from the experimental position of the ligand. This is a limitation of the method for longer ranges of the potentials, imposed by the smoothed intermolecular energy function. Certainly, our long-range potentials are applicable for an initial screen of a ligand's binding preferences only, which may be followed by more accurate calculations with 'regular' potential functions. Thus, instead of the r.m.s. estimates, we examined the positions of the ligand's center of gravity. Those corresponding to the crystal structure were further analyzed for ligand orientation preferences.

In all cases tested, the short-range potentials (R = ~2 Å),

which more closely approximate the characteristic Lennard–Jones potentials, produced a disperse distribution of local minima. A gradual increase in R, accompanied by a decrease in U values, invariably resulted in a more dense distribution, which eventually converged to the area of the global minimum (the experimental position of the ligand) at large values of R. In all cases, except the transmembrane helices of the photosynthetic reaction center, the final clustering of the ligand's low-energy positions at the global minimum area occurred at R = 6.8 Å. For the helix–helix interactions, the convergence was observed earlier, at R = 4.1 Å. A further increase in the corresponding values of R resulted in the destruction of position clusters and led to a homogeneous, random distribution of low-energy configurations.

Figure 2 illustrates three examples of the molecular complexes tested. For the α- and β-subunits of human deoxyhemoglobin (Figure 2a), the low-energy positions of the ligand (subunit β) eventually separated into three clusters. The largest one (on the right) corresponded to the α1–β1 contact, where the strongest interaction between the hemoglobin subunits occurs in the crystal structure. The second one (on top) coincided with the α1–β2 contact in the crystal structure of the heterodimer. The third cluster (on the left) corresponded to the strongest α–β contact between heterodimers in the crystal packing (Fermi *et al.*, 1984). In the case of enzyme–inhibitor complexes (a trypsin–BPTI complex is shown in Figure 2b), full convergence to the area of the global minimum was observed. The low-energy positions of antigens (lysozyme and a peptide), calculated with the variable region of Fab, consolidated into two distinct clusters (data not shown). The smaller one corresponded to the antigen position in the crystal structure. The larger one was a clear artifact of the variable region separation from the rest of the antibody in our calculations (it was located at the opposite side of the variable domain, which in the full molecule is buried inside the structure). In all cases tested, in low-energy configurations with longer ranges of the potential, the ligand's binding site showed a preference to be directed towards the binding site on the receptor. This is equivalent to the determination of two orientational degrees of freedom. The distribution of the third angle values (usually the spin around the axis connecting the centers of gravity of the molecules through their binding sites) was sparse. Along with the intermolecular interactions of soluble ligand–receptor systems, we tried transmembrane helices 1 and 2, as well as helices 4 and 5 of the photosynthetic reaction center (Figure 2c shows the helix 1–helix 2 system). For both pairs, all low-energy configurations converged to the global minimum area at shorter ranges of the potentials (R = 4.1 Å). The orientation of the main axis of the 'ligand' helix in these configurations corresponded to that of the crystal structure. However, the spin angle around this axis was much less defined.

The results clearly prove the feasibility of the global minimum determination in ligand–receptor systems, by simplifying and smoothing the energy profile. The pronounced character of the global minimum for the helix–helix interaction is certainly an important fact in itself. The distribution of low-energy configurations at various ranges of the potential also helps to reveal the role of different elements of macrostructure in intermolecular interaction. When short-range potentials (R = ~2 Å) are applied, the major role may be attributed to individual atom–atom interactions. This results in an elaborate energy profile with multiple local minima distributed around the 'receptor'. The longer ranges of the potential (R = ~4 Å) correspond to an averaging of atom–atom interactions at the level of individual residues. The best illustration of this could be the results of the helix–helix energy calculation, where the side chains are the largest elements of the structure. Further extension of the potential's range in this case results in a random distribution of the second helix (the 'ligand') positions, which corresponds to the absence of larger structural elements of recognition for helix–helix interactions. In contrast to this, for globular macromolecules and their ligands, such an extension leads to a further consolidation of low-energy configurations. The best clustering in the global minimum area was reached at ranges with R = ~7 Å. This has to be related to the role of elements of the general fold with larger characteristic sizes.

Conclusions and perspectives

The long-range step-function potentials, combined with an exhaustive grid search for the energy minima, present a practical approach to the multiminima problem in ligand–receptor interactions. The longer ranges for potential functions result in an averaging of individual atom–atom interactions. This also gives larger weight to the elements of the macro-structure, while filtering out the details of local atomic configurations. This dramatic smoothing of the energy profile comes along with a substantial simplification of the energy by a step function. A systematic, exhaustive grid search for low-energy values guarantees the complete determination of all minima (within the accuracy of the energy profile and the grid step). Application of this approach to various molecular complexes (multisubunit proteins, enzyme–inhibitor and antigen–antibody complexes) showed that the long-range potential functions eliminate all (in some cases, most) local minima and leave a distinct global minimum. For helix–helix interactions, similar results were obtained at intermediate ranges. The very existence of the pronounced global minimum in this case offers important possibilities for future modeling of integral membrane receptors.

We plan to expand our approach in several important directions. We will further investigate the role of elements of the general fold in ligand–receptor interactions. To determine the limits of applicability of our methodology, we will perform computer experiments on molecular complexes of known configuration, with distorted structures of their components. An important application of the approach will be the problem of transmembrane helices packing in integral membrane receptors as well as structural studies of ion channels and protein-folding algorithms. A natural consequence of our approach is the development of a technique, in which low-energy minima are determined using long-range potentials, followed by a gradual reduction of the potentials to the regular ranges. The contraction of the potential's range will be linked to the incorporation of more realistic features of real intermolecular interactions, which will contribute to an accurate structure prediction of ligand–receptor complexes.

Acknowledgements

I wish to thank Professor Garland R.Marshall for his support and helpful discussions. I am grateful to Professors Gregory V.Nikiforovich and Denise D.Beusen for their 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.
Deisenhofer,J. and Michel,H. (1989) *Science*, **245**, 1463–1473.
Fermi,G., Perutz,M.F., Shaanan,B. and Fourme,R. (1984) *J. Mol. Biol.*, **175**, 159–174.
Fujinaga,M., Sielecki,A.R., Read,R.J., Ardelt,W., Laskowski,M.,Jr and James,M.N.G. (1987) *J. Mol. Biol.*, **195**, 397–418.
Harrison,R.W., Kourinov,I.V. and Andrews,L.C. (1994) *Protein Engng*, **7**, 359–369.
Janin,J. (1995) *Proteins*, **21**, 30–39.
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.
McPhalen,C.A. and James,M.N.G. (1988) *Biochemistry*, **27**, 6582–6598.
Marquart,M., Walter,J., Deisenhofer,J., Bode,W. and Huber,R. (1983) *Acta Crystallogr.*, **39B**, 480–490.
Piela,L., Kostrowicki,J. and Scheraga,H.A. (1989) *J. Phys. Chem.*, **93**, 3339–3346.
Rini,J.M., Stanfield,R.L., Stura,E.A , Salinas,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.
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. and Aflalo,C. (1994) *Proteins*, **20**, 320–329.
Wawak,R.J., Wimmer,M.M. and Scheraga,H.A. (1992) *J. Phys. Chem.*, **96**, 5138–5145.

Received April 5, 1995; revised September 29, 1995; accepted October 14, 1995