

## Reaching the global minimum in docking simulations: A Monte Carlo energy minimization approach using Bezier splines

JEAN-YVES TROSSET AND HAROLD A. SCHERAGA\*

Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301

Contributed by Harold A. Scheraga, April 21, 1998

**ABSTRACT** The docking problem faces two major challenges: the global optimization of a multivariable function, such as the energy, and the ability to discriminate between true and false positive results, i.e., native from nonnative structures based on the input energy function. Among all energy evaluation tools, only a local energy-minimization method using an accurate enough potential function is able to discriminate between native and nonnative structures. To meet these requirements, a Monte Carlo with energy-minimization method has been incorporated into a new ECEPP/3 docking program. The efficiency of the simulation results from the use of an energy-grid technique based on Bezier splines and from a simplification of the receptor by switching on the energy of only important residues of the active site. Simulations of a thrombin-inhibitor complex show that the global minimum of the energy function was reached in every independent run within less than 3 min of time on an IBM RX 6000 computer. For comparison, 10 standard independent Monte Carlo simulations with  $10^6$  steps in each were carried out. Only three of them led to a conformation close to the x-ray structure. The latter simulations required an average of 24 min and about 10 hr with and without the grid, respectively. Another important result is that the Bezier spline technique not only speeds up the calculation by reducing the number of operations during the energy evaluation but also helps in reaching the global minimum by smoothing out the potential energy surface.

Determination of the structure of a receptor-ligand complex is a prerequisite for estimating the binding affinity or binding free energy between the two molecules. In a typical drug design project, this energy estimation should be accomplished for a large number of candidate molecules. The problem thus amounts to determining the structure of the bound ligand accurately, for the binding free energy calculation to be correct, and fast enough for screening a large molecular database in a reasonable amount of time. Much effort has been devoted to developing efficient automatic procedures to calculate the structures of ligands bound to their receptors.

This task faces two major challenges. The first one is to find the global minimum (GM) of a multivariable scoring or energy function that measures the fit between the ligand and the receptor. This is the multiple-minima problem that depends strongly on the size of the system. However, for docking simulations, the problem can be simplified by assuming the structure of the receptor and/or the ligand to be fixed or partially flexible. The second challenge is to find an energy function or an energy-evaluation tool that is able to discriminate between native and nonnative structures. Because the "energy" of the scoring function is the only criterion available for selecting the correct bound structure, it should have the

property that its GM value corresponds to the native conformation of the complex.

Many approaches have been proposed for docking a ligand onto its receptor. They are based mostly on geometric or energetic criteria or on a combination of the two. Geometric criteria use the idea of shape complementarity between the ligand and the receptor (1–22). This can be implemented in different ways, by least-squares fitting the positive and negative images of the ligand and the receptor, respectively (4, 7, 8), by using three-dimensional (3D) correlation techniques (6, 13, 17, 21, 22), or by using computer vision-based methods (10, 11). These methods have proven to be efficient for selecting a reasonable number of candidates for the bound structure of the ligand. These candidates then are re-evaluated according to some energetic criteria. This consists of calculating the energy of the bound ligand complex with an heuristic potential function or with a standard mechanical force field. The second approach makes use only of the energetic criteria without geometrical evaluation of the shape complementarity between the two molecules. The energy function involved refers to the standard types of molecular interactions such as hydrogen bond, electrostatic interaction, van der Waals contact, hydrophobic forces, solvation energy, etc. Currently, the methods used for sampling the conformational space are based on combinatorial or random search using a graph-theoretical approach (8, 9, 15, 20), genetic algorithms (23–27), Monte Carlo (MC) (28–39), or molecular dynamics (40–43).

All of the methods and approaches mentioned above face the same problem: the existence of what might be called "false-positive" results among the set of solutions, i.e., native-like structures with higher energy than some nonnative structures. Shoichet and Kuntz (3) have tested the reliability for differentiating the native conformation from nonnative ones with different energy evaluation methods such as buried surface area, free energy of solvation, mechanical constraints, packing, electrostatic complementarity, and energy minimization using standard mechanical force fields. They found that only the energy-minimization technique was able to discriminate, in most cases, the native from nonnative conformations. Electrostatic complementarity using a DELPHI type of calculation showed significant improvement (3). Guida *et al.* (30) used the energy-minimization technique coupled with a MC method to calculate the bound structure of four inhibitors of thermolysin. They found that, in each case, the crystallographically observed conformations were among the low-energy conformers discovered, with three of them having the lowest energy. Caflisch *et al.* (31) also applied the local energy-minimization technique, but within a different scheme, to the HIV type 1 aspartic acid proteinase complex. Their results also did not lead to any false-positive results; their lowest-energy conformation was the closest to the x-ray structure.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/958011-5\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

Abbreviations: MC, Monte Carlo; MCM, MC energy minimization; FPR, NH<sub>2</sub>-D-Phe-Pro-Arg-COOH; rmsd, rms deviation; GM, global minimum; 3D, three-dimensional.

\*To whom reprint requests should be addressed. e-mail: has5@cornell.edu.

It is necessary to use a local energy-minimization technique to discriminate between a native and a nonnative structure because a structure close to the GM but with some atomic clashes will have a high energy value, higher than some unrelated nonnative structure. In that case, the native-like structure will be ranked very high unless this structure is refined by energy minimization, in which case the GM is obtained. The reason the energy-minimization method is not used extensively for screening potential candidates is its computational cost. The computational time varies, of course, with the system, but for a typical complex with a small rigid ligand and a rigid receptor and with a distance cutoff of 8–10 Å, one energy minimization might take several minutes of computer time on a current workstation. Screening many candidates or implementing the minimization technique into a MC scheme would take hours of calculations. This frustrates the accepted goal of obtaining the docked structure in less than an hour or if possible within a few minutes of simulation per ligand molecule.

In this paper, we present two techniques that make the MC energy minimization (MCM) method competitive, especially for docking simulations. The first one is an approximation of the molecular system and turns out to be well adapted for the rigid docking problem. It consists of docking the ligand onto a small part of the active site: only certain residues of the active site of the receptor are considered during the simulation. The rest of the atoms, including those blocking accessibility to the active site, are not taken into account. For example, instead of considering all of the active site residues, only those lying at the "bottom" are seen by the ligand. When an approximate bound structure of the ligand is obtained, additional residues of the active site then are added to the system. This approach has the advantage of lowering the number of interactions and also finding the bound structure of the ligand by starting from the open form of the receptor when the closed form is not available.

The second improvement is the incorporation of a grid-based energy evaluation technique using a Bezier splines interpolation scheme developed recently in our laboratory (44). This technique enables one to evaluate, on a 3D grid, not only the energy but its gradient as well. Full advantage of the grid now can be applied to an energy-minimization procedure. The speed-up increases with the size of the rigid part of the system that is incorporated into the grid. This technique, together with the simplification of the receptor, lowers the time of one energy minimization to less than 1 s for a rigid ligand and a rigid receptor, compared with several minutes when the grid is absent and the whole rigid receptor is taken into account in the energy calculation, i.e., without a cutoff distance.

This paper presents an adaptation of the MCM method (45, 46) for the docking simulation. Calculations were carried out for the noncovalent binary complex of human  $\alpha$ -thrombin with the tripeptide NH<sub>2</sub>-D-Phe-Pro-Arg-COOH (FPR). This substrate is analogous to PPACK (D-Phe-Pro-Arg-chloromethylketone), which is an inhibitor of the fibrinogen interaction site of thrombin. The structure of the covalent complex, thrombin-PPACK, has been determined by x-ray crystallography at 2.3 Å resolution (47, 48). Besides its biological relevance, thrombin presents interesting structural characteristics for the active site: this is a serine protease that cleaves the Arg-16–Gly-17 peptide bond of the A $\alpha$ -chain of human fibrinogen. The arginine side chain of the ligand is inserted into a deep narrow pocket of the enzyme making the docking difficult by a pure MC technique (see *Results*). Two insertion loops flank each side of the thrombin pocket. They play the role of a lid that opens and closes during the binding of fibrinogen to thrombin (47, 48).

Both molecules were considered as rigid in this study. To test the ease or difficulty of the docking of FPR onto thrombin, we

carried out comparative calculations using the standard Metropolis MC without energy minimization but using our grid-based technique. This approach is similar to the approach used by many authors (12, 28, 29, 32, 38). The efficiency of the grid technique has been measured by the savings in computational time and also in the number of MC steps necessary to reach the GM. The results show that the Bezier spline not only decreases the simulation time by lowering the number of operations per energy evaluation but also has the nice property of smoothing out the potential energy surface, which is particularly helpful in searching for the GM (49). This program is part of a new modeling ECEPP/3 package developed recently in our laboratory.

## METHODS

**Regularization.** The coordinates of human  $\alpha$ -thrombin were taken from the file 1ppb in the Protein Databank (50). The reference structure of the complex that will be used for comparison with the calculated structure was prepared as follows: the x-ray structures of the ligand and the A and B chains of human  $\alpha$ -thrombin were regularized independently, using the protocol that will be described elsewhere; i.e., the structures were forced to have the standard bond lengths and bond angles of ECEPP/3. After regularization, the structure of the whole complex (internal degrees of freedom) was energy-minimized in several stages by using harmonic distance constraints with decreasing weights. At each stage, 2,000 minimization iterations were carried out except for the last stage in which minimization was carried out until convergence (rms-gradient = 10<sup>-4</sup> kcal/mol.rad). The values of the weight, associated with each minimization stage were: 100.0, 10.0, 1.0, 0.1, and 0.0. The purpose of the constraint energy was to prevent the complex from departing significantly from its crystal structure during the early stages of minimization. The last energy minimization was carried out without x-ray constraints. No cutoff distance was used, which is an important point. If a cut-off distance is introduced during minimization, the final structure after complete minimization will not be at the energy minimum when a new list of nonbonded interactions is recalculated from the final structure. Because of this, convergence cannot be attained with a cutoff distance, no matter how many successive energy minimizations are carried out. At the end of the complete minimization, the rms deviation (rmsd) between the regularized structure and the x-ray structure was 1.9 Å for thrombin and 1.0 Å for the FPR portion of PPACK.

**Docking.** At the beginning of the simulation, the orientation and the position of the ligand were generated randomly in such a way that the Pro2-C $\alpha$  atom of FPR was located uniformly inside a sphere of 3 Å or 15 Å radius (computational experiments I and II, respectively) centered at point P (6.2, 9.5, 13.5 Å) near the catalytic residue Ser-195 (with respect to the coordinate system of the Protein Databank structure). A harmonic distance constraint was applied when the ligand atom Pro2-C $\alpha$  moved outside of the sphere. The docking thus consists of tumbling the ligand inside the active site region. Only seven degrees of freedom were considered, three for translation and four quaternions for rotation. The constraint concerning the norm of the quaternion was incorporated in the energy function (unpublished work). The step sizes for the quaternions were large enough for two successive rotations to be uncorrelated. The step size for the translation move was chosen in such a way that the displacement of the center of mass of the ligand was about 0.1–0.3 Å per MC move.

To facilitate the docking, a large part of the receptor was disregarded in the computation. For experiment I, the only residues taken into account were the residues of the binding pocket of thrombin (Fig. 1). They correspond to segment nos. 14 and 16 of the B chain of thrombin (Table 1). All of the other

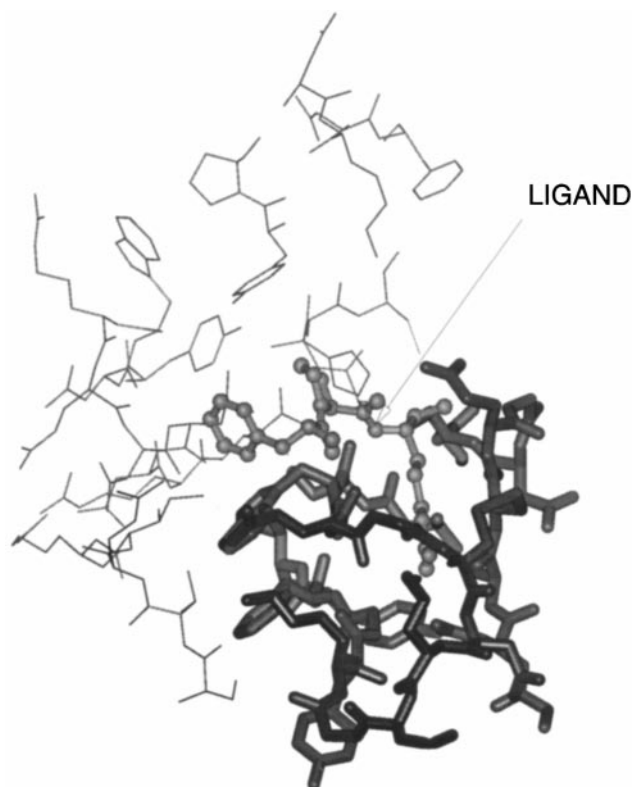


FIG. 1. Active site residues of human  $\alpha$ -thrombin that were considered during the simulation. Thick line: residues taken into account in experiment I. Thin line: additional residues added for experiment II. The ligand FPR is indicated by an arrow.

segments of thrombin were turned off in the energy calculation, in particular, the surrounding loops “Trp-148” and “Trp-60D” that are located at the entrance of the active site (47, 48). For experiment II, all of the active site segments were turned on during the simulation (see Table 1). In our simulation, all residues were considered in their neutral charge state in order not to bias the simulations, because Arg-3 of FPR can make an electrostatic bridge with Asp-109 of thrombin, which is located at the bottom of the active site at about 2.5 Å from the guanidinium group of Arg-3. The calculations were carried out on a single node of the IBM-SP2 RX 6000 Super-Computer at the Cornell Theory Center.

**Grid Calculation.** In this study, the contribution of the nonbonded and electrostatic field of all of the atoms of the rigid segments of the receptor that were considered during the simulation (switched on) was incorporated into a 3D grid. The principle and the manner in which the grid is calculated have been presented elsewhere (44). The grid was defined as a cube of  $56 \text{ \AA}^3$  volume with a constant spacing of  $0.666 \text{ \AA}$ , whose center is located at point P1 = (6, 20, 13 Å) somewhere inside the active site. The 3D grid was actually split up into 15 different grids: one for the electrostatic field and 14 for the nonbonded energy associated with each of 14 atom types. The value of the field at a given point of the active site was estimated by using two layers of neighbors, i.e., the 64 nearest points of the grid. The interpolation of the value of the field at the position of any atom of the ligand, from these 64 grid points, was carried out with Bezier splines. The major advantage of this interpolation scheme is that the splines are smooth, i.e., continuous and differentiable everywhere on the grid. The energy as well as the gradient could be estimated from the grid points, making the energy minimization very efficient. The calculation of the grid was carried out only twice, once for each experiment (I and II).

Table 1. List of segments for which the energy has been switched on or switched off during the rigid docking of FPR onto human  $\alpha$ -thrombin

Segment number	Residue number in the sequence	Chymotrypsin numbering	Energy turned off or turned on for experiment I	Energy turned off or turned on for experiment II
FRP				
1.	1-3	1-3	ON	ON
Thrombin A chain				
1.	1-36	1-15	OFF	OFF
Thrombin B chain				
1.	1-40	16-54	OFF	OFF
2.	<b>41-44</b>	<b>55-58</b>	<b>OFF</b>	<b>ON</b>
3.	45-46	59-60	OFF	OFF
4.	<b>47-48</b>	<b>60A-60B</b>	<b>OFF</b>	<b>ON</b>
5.	<b>49-51</b>	<b>60C-60D</b>	<b>OFF</b>	<b>OFF</b>
6.	<b>52-54</b>	<b>60E-60H</b>	<b>OFF</b>	<b>ON</b>
7.	55-89	60I-93	OFF	OFF
8.	<b>90-99</b>	<b>94-102</b>	<b>OFF</b>	<b>ON</b>
9.	100-140	103-140	OFF	OFF
10.	<b>141-156</b>	<b>141-151</b>	<b>OFF</b>	<b>OFF</b>
11.	157-175	152-170	OFF	OFF
12.	<b>176-180</b>	<b>171-175</b>	<b>OFF</b>	<b>ON</b>
13.	181-197	176-187	OFF	OFF
14.	<b>198-206</b>	<b>188-196</b>	<b>ON</b>	<b>ON</b>
15.	207-224	197-212	OFF	OFF
16.	<b>225-241</b>	<b>213-229</b>	<b>ON</b>	<b>ON</b>
17.	242-259	230-247	OFF	OFF

The residues of the active site of thrombin are listed in bold type, even though the energies of all of them were not turned on.

## RESULTS AND DISCUSSION

The first series of experiments consisted of docking the FPR ligand onto a small part of the active site consisting of two segments, nos. 14 and 16. The ligand was generated inside a sphere of radius 3 Å. Twenty independent rigid-docking simulations using the MCM method and the 3D grid energy evaluation were carried out. The simulation stopped when the GM was reached, i.e., when the energy was less than 0.4 kcal/mol above the GM ( $-40.7 \text{ kcal/mol}$ ); this corresponds to a rmsd from the reference structure of the ligand of less than 0.1 Å. Our results showed that all 20 simulations reached the GM. On average,  $49 \pm 34$  MC steps were required to reach the GM corresponding to an average simulation time of 35 s with a SD of 20 s.

The same types of simulations also were carried out without the grid. The grid calculation was replaced by the calculation of all of the atomic interactions between the ligand and the two segments. The GM was reached again in all 20 simulations. The average number of MC steps to reach the GM was  $126 \pm 108$  with a time of  $411 \pm 347$  s, where the error corresponds to one SD. The computational time is thus about 12 times longer than simulations with the grid. This factor depends, of course, on the size of the system. The result is that an average of 2–3 times more MC steps are required to reach the GM when the simulations are carried out without the grid. This suggests that it is easier to locate the GM when the original function is replaced by Bezier splines. This is because of the smoothing property of the Bezier splines. By smoothing out the potential energy surface, it is easier to overcome the energy barriers and reach the GM.

To assess the complexity imposed by the narrowed active site of thrombin, we compared our MCM procedure with a current MC procedure used for docking (12, 28, 29, 32, 38). Simulations were carried out using the same procedure as previously, i.e., with the ligand initially placed randomly in a sphere of 3 Å radius around the active site, and the grid-based energy

Table 2. Results of the 10 independent MC simulations (without energy minimization) of experiment I

Energy*	rmsd†, Å
-37.7	0.4
-29.8	0.6
-22.6	7.6
-22.3	1.2
-22.2	0.8
-21.7	5.9
-21.5	6.8
-21.3	7.8
-21.3	6.8
-21.2	7.4

\*The GM is defined to be the set of energies for which the rmsd is less than 0.1 Å, i.e.  $E$  is in the range  $[-40.7; -40.3]$  kcal/mol.

†The rmsd is calculated between the lowest-energy structure of the ligand FPR and its reference structure, which is the structure obtained after energy-minimizing the x-ray structure of the complex, human  $\alpha$ -thrombin-FPR.

calculation but with standard MC without minimization. Ten independent standard MC simulations of  $10^6$  steps were carried out. The results showed that none of them reached the GM (Table 2). However, the lowest energy value, which is about 3 kcal/mol above that of the GM corresponds to a structure close to the native (rmsd = 0.4 Å). The average time per simulation was about 24 min. The chance of reaching the GM thus can be estimated at 10% (1/10) for the standard MC procedure. The MCM protocol is about 40 times faster with a chance of success of 100% for this system. This series of standard MC simulations also showed false-positive results (Table 2). Structures with low energy values have comparatively high rmsd values and, conversely, high-energy structures have low rmsd values. This shows the importance of reaching the GM in a docking simulation, because, after a certain threshold above the GM, the energy and the rmsd value are no longer correlated, i.e., it is no longer possible to predict the native conformation from the energy value.

To test the robustness of the method, a second series of computer experiments was carried out. This time, almost all of the residues of the active site were taken into account, and the ligand was generated randomly inside a sphere of 15 Å centered in the middle of the active site. As before, 20 independent MCM simulations, using the grid, were carried out. We again obtained 100% success in reaching the GM. The average number of steps was  $208 \pm 202$ , corresponding to an average time of  $140 \pm 128$  s. Without the grid, 19 of 20 simulations led to the GM within an average time of 2,284 s. The computer time savings with the grid in this case was about 16-fold. For comparison, 10 independent MC simulation of  $10^6$  steps, without energy minimization and without the grid, were carried out. Each simulation took about 10 hr of computer time. None of them led to the GM. All of the lowest energy conformations of the complex were 10–20 kcal/mol higher in energy than that of the GM with rmsd values ranging from 6 to 8 Å.

### CONCLUDING REMARKS

The MCM method using Bezier splines for the grid interpolation scheme meets the two major requirements for docking ligands onto proteins: robustness and accuracy. Robust enough to reach the GM with great reproducibility (in our case, 100%) and accurate because no false-positive results were obtained. The lowest energy value corresponds to the native structure. The reasons for the absence of false-positive results are, first, the use of local energy minimization after each MC step. Therefore, only energy-refined structures without atomic clashes are considered. Second, the ECEPP/3 force field with

an all atom model is sufficiently accurate to describe the x-ray structure of the complex as the GM. In this study, the energy-minimized complex was the reference system. If the reference structure were not energy-minimized, then conformations with lower energy could be obtained by MCM simulation.

The results show that it is sufficient to consider only a small part of the active site to determine the bound conformation of the ligand. When the ligand is rigid, the GM of the reduced molecular system turns out to correspond to the native bound structure. This means that the orientation and location of the ligand are determined entirely by a few particular active site residues, such as those of the active site pocket. The other residues do not influence the orientation and position of the ligand. In flexible docking, this might not be true. Even if some part of the ligand is correctly localized in the active site, a portion of the ligand might change its internal conformation to adapt to the perturbation of the active site. This is demonstrated in a study with a flexible ligand and/or a flexible receptor (M. C. Maurer, J.-Y.T., C. C. Lester, E. E. DiBella, and H.A.S., unpublished results).

We thank J. Kostrowicki for helpful discussions, and the Association Française pour la Recherche Thérapeutique and the Centre National de la Recherche Scientifique IMABIO for fellowship support to J.-Y.T. This work was supported by National Science Foundation Grant MCB95-13167, by National Institutes of Health Grants GM-14312 and HL-30612, and by the Cornell Biotechnology Center. The calculations were carried out on the IBM SP2 supercomputer at the Cornell Theory Center (CTC), which is funded in part by the National Science Foundation, New York State, the IBM corporation, the National Institutes of Health National Center for Research Resources (P41 RR-04293), and the CTC Corporation Partnership Programs.

- DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. & Venkataraghavan, R. (1988) *J. Med. Chem.* **31**, 722–729.
- Jiang, F. & Kim, S.-H. (1991) *J. Mol. Biol.* **219**, 79–102.
- Shoichet, B. K. & Kuntz, I. D. (1991) *J. Mol. Biol.* **221**, 327–346.
- Bacon, D. J. & Moulton, J. (1992) *J. Mol. Biol.* **225**, 849–858.
- Kasinos, N., Lilley, G. A., Subbarao, N. & Haneef, I. (1992) *Protein Eng.* **5**, 69–75.
- Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. & Vakser, I. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2195–2199.
- Leach, A. R. & Kuntz, I. D. (1992) *J. Comp. Chem.* **13**, 730–748.
- Shoichet, B. K., Bodian, D. L. & Kuntz, I. D. (1992) *J. Comp. Chem.* **13**, 380–397.
- Shoichet, B. K. & Kuntz, I. D. (1993) *Protein Eng.* **6**, 723–732.
- Helmer-Citterich, M. & Tramontano, A. (1994) *J. Mol. Biol.* **235**, 1021–1031.
- Fischer, D., Lin, S. L., Wolfson, H. L. & Nussinov, R. (1995) *J. Mol. Biol.* **248**, 459–477.
- Knegtel, R. M. A., Antoon, J., Rullmann, C., Boelens, R. & Kaptein, R. (1994) *J. Mol. Biol.* **235**, 318–324.
- Meyer, M., Wilson, P. & Schomburg, D. (1996) *J. Mol. Biol.* **264**, 199–210.
- Peters, K. P., Fauck, J. & Frömmel, C. (1996) *J. Mol. Biol.* **256**, 201–213.
- Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996) *J. Mol. Biol.* **261**, 470–489.
- Sobolev, V., Wade, R. C., Vriend, G. & Edelman, M. (1996) *Proteins* **25**, 120–129.
- Vakser, I. A. (1996) *Protein Eng.* **9**, 37–41.
- Vakser, I. A. (1996) *Biopolymers* **39**, 455–464.
- Wallqvist, A. & Covell, D. G. (1996) *Proteins* **25**, 403–419.
- Ewing, T. J. A. & Kuntz, I. D. (1997) *J. Comp. Chem.* **18**, 1175–1189.
- Friedman, J. M. (1997) *Protein Eng.* **10**, 851–863.
- Gabb, H. A., Jackson, R. M. & Sternberg, M. J. E. (1997) *J. Mol. Biol.* **272**, 106–120.
- Judson, R. S., Jaeger, E. P. & Treasurywala, A. M. (1994) *J. Mol. Struct.* **308**, 191–206.
- Clark, K. P. & Ajay, (1995) *J. Comp. Chem.* **16**, 1210–1226.
- Jones, G., Willett, P. & Glen, R. C. (1995) *J. Mol. Biol.* **245**, 43–53.

26. Judson, R. S., Tan, Y. T., Mori, E., Melius, C., Jaeger, E. P., Treasurywala, A. M. & Mathiowetz, A. (1995) *J. Comp. Chem.* **16**, 1405–1419.
27. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. (1997) *J. Mol. Biol.* **267**, 727–748.
28. Goodsell, D. S. & Olson, A. J. (1990) *Proteins* **8**, 195–202.
29. Yue, S.-Y. (1990) *Protein Eng.* **4**, 177–184.
30. Guida, W. C., Bohacek, R. S. & Erion, M. D. (1992) *J. Comp. Chem.* **13**, 214–228.
31. Caflisch, A., Niederer, P. & Anliker, M. (1992) *Proteins* **13**, 223–230.
32. Hart, T. N. & Read, R. J. (1992) *Proteins* **13**, 206–222.
33. Abagyan, R., Totrov, M. & Kuznetsov, D. (1994) *J. Comp. Chem.* **15**, 488–506.
34. Bohacek, R. S. & McMartin, C. (1994) *J. Am. Chem. Soc.* **116**, 5560–5571.
35. Cherfils, J., Bizebard, T., Knossow, M. & Janin, J. (1994) *Proteins* **18**, 8–18.
36. Friedman, A. R., Roberts, V. A. & Tainer, J. A. (1994) *Proteins* **20**, 15–24.
37. Totrov, M. & Abagyan, R. (1994) *Nat. Struct. Biol.* **1**, 259–263.
38. Cummings, M. D., Hart, T. N. & Read, R. J. (1995) *Protein Sci.* **4**, 885–899.
39. Nakajima, N., Higo, J., Kidera, A. & Nakamura, H. (1997) *Chem. Phys. Lett.* **278**, 297–301.
40. Di Nola, A., Roccatano, D. & Berendsen, H. J. C. (1994) *Proteins* **19**, 174–182.
41. Luty, B. A., Wasserman, Z. R., Stouten, P. F. W., Hodge, C. N., Zacharias, M. & McCammon J. A. (1995) *J. Comp. Chem.* **16**, 454–464.
42. Wasserman, Z. R. & Hodge, C. N. (1996) *Proteins* **24**, 227–237.
43. Gorse, A.-D. & Gready, J. E. (1997) *Protein Eng.* **10**, 23–30.
44. Oberlin, D., Jr. & Scheraga, H. A. (1998) *J. Comp. Chem.* **19**, 71–85.
45. Li, Z. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
46. Li, Z. & Scheraga, H. A. (1988) *J. Mol. Struct.* **179**, 333–352.
47. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989) *EMBO J.* **8**, 3467–3475.
48. Bode, W., Turk, D. & Karshikov, A. (1992) *Protein Sci.* **1**, 426–471.
49. Kostrowicki, J. & Scheraga, H. A. (1996) *DIMACS Ser. Discrete Math. Theor. Comp. Sci.* **23**, 123–132.
50. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.