PROTEINS: Structure, Function, and Genetics 17:266–278 (1993)

Orientational Sampling and Rigid-Body Minimization in Molecular Docking

Elaine C. Meng,[1] Daniel A. Gschwend,[1] Jeffrey M. Blaney,[2] and Irwin D. Kuntz[1]
[1]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143-0446; and [2]Chiron Corporation, Emeryville, California 94608

ABSTRACT **The biological activities of proteins depend on specific molecular recognition and binding. Computational methods for predicting binding modes can facilitate the discovery and design of ligands and yield information on the factors governing complementarity. The DOCK suite of programs has been applied to several systems; here, the degree of orientational sampling required to reproduce and identify known binding modes, with and without rigid-body energy minimization, is investigated for four complexes. There is a tradeoff between sampling and minimization. The known binding modes can be identified with intensive sampling alone (10,000 to 20,000 orientations generated per system) or with moderate sampling combined with minimization. Optimization improves energies significantly, particularly when steric clashes are present, and brings many orientations closer to the experimentally observed position. Whether or not minimization is performed, however, sampling must be sufficient to find at least one structure in the vicinity of the presumed true binding mode. Hybrid approaches combining docking and minimization are promising and will become more viable with the use of faster algorithms and the judicious selection of fewer orientations for minimization.**

Key words: molecular recognition, ligand binding, complementarity, interaction energy, automated prediction

INTRODUCTION

Ligand binding is a crucial aspect of protein function. Biological activity for most proteins, whether or not they are receptors in the classical sense, depends on the specific recognition of one or more binding partners.

Computational approaches to predicting binding geometries are of interest not only for their potential utility in the design of bioactive compounds, but also for insights they may yield on the thermodynamics of molecular recognition. In general, these methods "dock" molecules together in many ways and "score" or evaluate each orientation. Ligand discovery applications typically involve searching through a database of compounds, with the idea that those which score well should be more likely to bind to the target macromolecule. Docking can also be useful in ligand optimization; knowledge of a compound's preferred mode of binding is needed so that one can suggest structural modifications intended to form, enhance, or disrupt specific interactions with the receptor.

No matter what the ultimate goal of a docking study is, both orientational sampling and energy evaluation must be adequate for useful results to be obtained. Sufficient sampling of orientation space is necessary because favorable geometries must be found before they can be recognized; conversely, without a good scoring function, "correct" binding modes cannot be recognized even when they have been generated. With the DOCK suite of programs, it has been shown that experimental geometries can be reproduced quite accurately, and in many cases, identified by different scoring functions.[1–3] The degree of sampling required to achieve this goal has not been examined systematically, however. It is also unclear whether it is more efficient to sample thoroughly and select the most favorable orientations from the resulting population, or to generate fewer orientations and optimize them within the context of the receptor site. These issues are addressed below, using systems employed in previous tests of DOCK.[3]

TEST SYSTEMS AND COMPUTATIONAL METHODS

Four well-determined structures of ligand/receptor complexes were selected from the Brookhaven Protein Data Bank[4] (Fig. 1): 4dfr[6] (dihydrofolate reductase/methotrexate), 6rsa[7] (ribonuclease A/uridine vanadate), 2gbp[8] (periplasmic binding protein/glucose), and 3cpa[9] (carboxypeptidase A/glycyltyrosine). As described previously,[3] DOCK 3.0 was used to generate multiple orientations of each ligand in the corresponding receptor site. Briefly, in each system, all crystallographic waters and ions

Received April 26, 1993; revision accepted July 21, 1993.
Address reprint requests to Dr. Irwin Kuntz, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446.

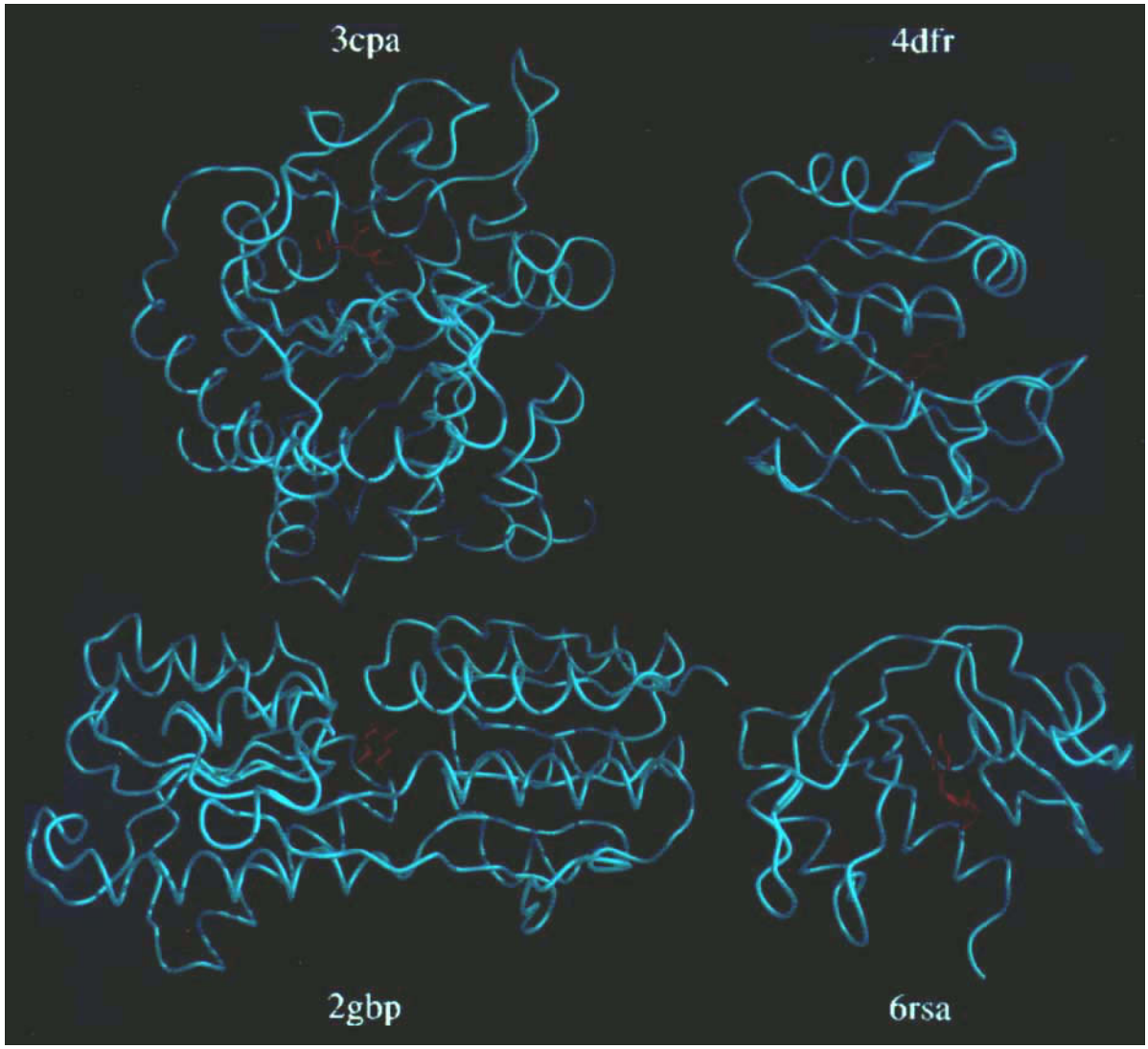

Fig. 1. Ligands (red) and protein Cα traces (cyan) for the test systems 4dfr, 6rsa, 2gbp, and 3cpa, shown clockwise starting from the upper right. Figure generated with UCSF MidasPlus.[5]

were removed, the ligand and receptor were separated, and hydrogens were added in standard geometries. A molecular surface was created for the receptor region of interest with the MS algorithm,[10] then used in the program SPHGEN[1] to calculate spheres for docking. The CHEMGRID and DISTMAP modules of DOCK 3.0 were used to create grids for force field (FF) and contact scoring, respectively.

DOCK FF scores are approximate interaction energies, a sum of van der Waals (VDW) and electrostatic components:[3]

$$FF\ score = \sum_{i=1}^{lig}\left[\sqrt{A_{ii}}\sum_{j=1}^{rec}\frac{\sqrt{A_{jj}}}{r_{ij}^{12}} - \sqrt{B_{ii}}\sum_{j=1}^{rec}\frac{\sqrt{B_{jj}}}{r_{ij}^{6}} + 332.0q_i\sum_{j=1}^{rec}\frac{q_j}{Dr_{ij}}\right].$$

Each term is a double sum over ligand atoms i and receptor atoms j, A and B are VDW repulsion and attraction parameters, r_{ij} is the distance in angstroms between atoms i and j, q_i and q_j are the point charges on atoms i and j, D is the dielectric function, and 332.0 is a factor that converts the electrostatic energy into kcal/mol. In CHEMGRID, the sums over receptor atoms are evaluated at each point on a three-dimensional grid and stored; later, during docking, rapid scoring is performed by multiplying the appropriate ligand atom descriptors by the nearby grid point values (usually with interpolation). AMBER united-atom parameters[11] were used for the receptors; for the ligands, steric parameters were based on the AMBER all-atom force field[12] and partial charges were derived as described previously.[3] A 10.0-Å cutoff and a dielectric function of D

TABLE I. Docking Variables

Run	Ligand bins		Sphere bins		Atoms[†]	Spheres[‡]	Found[§]	Minscore**	Written[††]
	Width*	Overlap*	Width*	Overlap*					
4dfr									
High	1.0	0.2	1.0	0.2	13	86	67,234	60[‡‡]	2,406
Medium	0.4	0.1	0.8	0.2	13	86	17,354	60[‡‡]	869
Low	0.2	0.0	1.0	0.0	13	86	7,158	60[‡‡]	337
6rsa									
High	1.0	0.5	1.0	0.5	21	47	77,184	40[‡‡]	3,492
Medium	0.4	0.1	0.8	0.2	21	47	4,785	0	266
Low	0.2	0.0	1.0	0.0	21	47	1,693	0	105
2gbp									
High	1.0	0.4	1.0	0.4	12	75	196,752	60[‡‡]	1,680
Medium	0.4	0.1	0.8	0.2	12	75	21,037	0	849
Low	0.2	0.0	1.0	0.0	12	75	7,773	0	389
3cpa									
High[§§]	1.5	0.5	1.5	0.5	17	44	794,541	100[‡‡]	2,684
Medium	0.4	0.1	0.8	0.2	17	44	7,946	0	518
Low	0.2	0.0	1.0	0.0	17	44	4,053	0	301

*Angstroms.
[†]Number of ligand atoms, excluding hydrogens.
[‡]Number of spheres describing the receptor site.
[§]Total number of orientations found.
**Minimum contact score for writing out an orientation.
[††]Number of orientations written out, those which do not violate the close contact limits set in DISTMAP and have a contact score no less than minscore and a FF score no greater than 100.0 kcal/mol.
[‡‡]Set greater than zero to limit the number of orientations written.
[§§]The distance-matching tolerance was 2.0 Å; in all other cases, a tolerance of 1.5 Å was used.

$= 4r$ were used, where r is the distance in angstroms between interacting atoms.

Grids for contact scoring are produced by DISTMAP.[2] For each grid point, the number of receptor atoms within an acceptable distance range is stored; points making close contacts are penalized with a large negative score. Hydrogens are ignored. During docking, each ligand atom receives the score of the nearest grid point, and the total contact score is the sum over the ligand atoms. Close contact limits were set to 2.3 and 2.8 Å for receptor polar and nonpolar atoms, respectively, and the cutoff distance was 4.5 Å.

Docking parameters were as described previously[3] for the runs with high sampling (over 50,000 orientations generated per system); however, two additional runs were performed for each system at lower levels of sampling. The docking parameters for each trial are listed in Table I.

The ligand-site matching algorithm is purely geometric and identical to that used in DOCK 2.0.[2] Sets of sphere centers are matched to sets of ligand atoms based on pairwise internal distances and used to generate ligand orientations. The sphere–sphere and atom–atom distances are first sorted into bins; a sphere–atom pairing is considered only when these points are in the corresponding bins. In the present work, an orientation is found when the distances among four spheres are equivalent to the distances among four ligand atoms, within a tolerance. Together with the distance-matching tolerance, the ligand bin width and overlap and the receptor (sphere) bin width and overlap control the total number of orientations found. Since four atom–sphere pairs are required, each point in the three longest-distance ligand bins is tried versus each point in the corresponding sphere bin. Enlarging the bins increases the combinatorial search and the number of orientations generated, since there will be more points per bin. Increasing the total number of ligand atoms or spheres has a similar effect. Only a subset of the orientations found is written out, subject to minimum contact score and maximum FF score criteria (Table I) and the close contact limits set in DISTMAP.

Rigid-body minimization was performed on each orientation written from DOCK, using an adaptation of the program RGDMIN.[13] In this algorithm, the receptor molecule is kept stationary while the six degrees of freedom of the rigid ligand, three Euler angles and three translational vectors, are manipulated to minimize the calculated interaction energy. Numerical derivatives are used within a modified Davidon–Fletcher–Powell routine.[14] RGDMIN was adapted to incorporate the DOCK 3.0 FF scoring function and parameters,[3] resulting in the program RGDMIN3. The only difference between the FF scores and the RGDMIN3 interaction ener-

TABLE II. Timings*

Run	DOCK (s)	RGDMIN3 (hr:min)	Written†
4dfr			
High	103	10:39	2,406
Medium	43	3:51	869
Low	30	1:30	337
6rsa			
High	146	9:26	3,492
Medium	34	0:39	266
Low	28	0:15	105
2gbp			
High	178	6:37	1,680
Medium	60	2:25	849
Low	48	1:04	389
3cpa			
High	766	23:24	2,684
Medium	33	3:00	518
Low	25	1:42	301

*All calculations performed on a Silicon Graphics Iris 4D/35 workstation.
†Number of orientations written by DOCK and minimized using RGDMIN3.

gies is that the former use precalculated grid values while the latter are based on the exact distances among ligand and receptor atoms. The two sets of values correspond closely (data not shown). Minimizations were performed with a 10.0-Å nonbonded cutoff, a maximum rotational step size of 10°, and a maximum translational step size of 0.5 Å. Minimizations were continued until restarts decreased the energy by less than 0.5 kcal/mol.

All calculations were carried out on a Silicon Graphics Iris 4D/35 workstation; timings are given in Table II. Although the input and results for the high-sampling docking runs are the same as described previously,[3] the runs were faster, for two reasons: time-saving changes were incorporated into DOCK between the two sets of runs, and a faster workstation was used for the newer calculations.

RESULTS AND DISCUSSION

For each system and at each level of sampling, the root-mean-square deviation in atomic position (RMSD) of every orientation of the ligand relative to the experimentally observed orientation was calculated. Hydrogens were not included in the calculation. RMSD is plotted versus DOCK FF score, before and after rigid-body minimization. Pre- and post-minimization RMSD values are compared. Table III summarizes the results in terms of the top-scoring orientations before and after minimization, and Table IV shows the effects of optimizing the crystal structure orientations.

Dihydrofolate Reductase

N1-protonated 2,4-diamino-6-methylpteridine, the rigid part of methotrexate, was used for docking.[3] Members of the lowest-RMSD family of orientations receive the most favorable scores at all levels of sampling, even before minimization (Fig. 2 and Table III). Other families of orientations also score well: the 2.8-Å structures are barrel-rolled and angled slightly relative to the crystallographic position, the 4.0-Å structures are angled approximately 90°, and the 4.8-Å structures are flipped end-to-end. The 9.0- and 13.0-Å structures do not overlap the experimentally observed orientation.

The most prominent effect of minimization is to collapse the high FF scores down without, for the most part, causing large changes in RMSD (Figs. 2 and 3). Apparently, steric conflicts can be resolved by slight rigid-body adjustments of the ligand, as expected from the steepness of the VDW potential. Similar numbers of orientations move farther from and closer to the experimental geometry during minimization. This is not surprising, since many local minima are being explored; there is no reason to expect a linear or even monotonic relationship between RMSD and interaction energy over such a large region of orientation space. When only the lowest-RMSD structures are considered, however, the majority of the orientations move toward the experimentally observed position (Fig. 3), suggesting that the X-ray orientation lies near a local minimum.

Ribonuclease A

Uridine 3′-phosphate was constructed from the crystallographic ligand, uridine vanadate, and used for docking.[3] The RMSD values are more diffusely distributed than in the dihydrofolate reductase test case; significant translations relative to the observed binding mode as well as opposite head-to-tail alignments are found. At the highest level of sampling, the lowest-RMSD family of orientations is identified as the most favorable both before and after minimization (Fig. 4 and Table III). At the intermediate level of sampling, just two orientations resembling the known binding mode are found; they receive the best scores, but only after minimization are they clearly the most favorable. At the lowest level of sampling, members of the lowest-RMSD family of orientations again receive the best scores, but minimization *decreases* their separation in score from orientations with RMSD values close to 5.0 Å. The latter have their phosphate groups in essentially the correct position but are at an angle of about 60° relative to the experimentally observed orientation. This weak discrimination between "correct" and "incorrect" modes by the FF score may reflect an overemphasis on electrostatics, especially when interactions among formal charges are involved.[3]

Interestingly, a greater number of low-RMSD orientations is found with low sampling than with intermediate sampling. This highlights an important

TABLE III. The Top-Scoring Orientations Before and After Minimization

	Before minimization				After minimization			
	"Correct"*		"Incorrect"*		"Correct"*		"Incorrect"*	
Run	FF score†	RMSD‡	FF score†	RMSD‡	FF score†	RMSD‡	FF score†	RMSD‡
4dfr								
High	−30.828	0.64	−24.998	3.75	−32.869	0.52	−26.655	3.99
Medium	−30.773	1.27	−24.269	2.81	−32.869	0.52	−27.439	2.83
Low	−30.828	0.64	−23.126	4.05	−32.869	0.52	−25.180	2.92
6rsa								
High	−58.437	0.56	−52.256	5.51	−63.736	0.55	−57.381	4.83
Medium	−43.017	1.96	−38.122	12.23	−61.381	0.96	−47.926	5.36
Low	−48.551	0.84	−27.459	8.48	−62.683	0.34	−57.381	4.83
2gbp								
High	−21.485	0.29	−13.752	12.34	−23.007	0.38	−14.358	12.58
Medium	−8.537	0.84	−13.327§	12.43	−19.267	0.30	−14.156	12.40
Low	42.252	0.62	−12.341§	12.70	−20.630	0.25	−14.156	12.40
3cpa								
High	−41.303	1.17	−30.995	7.03	−46.929	0.64	−34.154	7.20
Medium	−25.775	1.78	−30.652§	7.66	−38.935	0.76	−36.055	2.04**
Low	−22.925	1.84	−24.795§	7.63	−36.221	1.93	−31.594	6.87

*"Correct" and "incorrect" refer to RMSD values below and above 2.0 Å, respectively.
†kcal/mol.
‡Angstroms.
§Note that an "incorrect" orientation receives the best score.
**Arguably the correct binding mode; the top-scoring *obviously* incorrect orientation has an FF score of −33.342 kcal/mol and an RMSD of 8.24 Å.

TABLE IV. The Effects of Minimizing the Crystallographic Ligand Orientations

System	Ligand	FF score*	FF score,* min	RMSD,† min
4dfr	2,4-Diamino-6-methylpteridine	−29.085	−31.484	0.5263
6rsa	Uridine 3′-phosphate	−61.389	−61.821	0.1499
2gbp	β-D-Glucose	−16.995	−21.904	0.2755
3cpa	Glycyl-L-tyrosine	−25.167	−45.224	0.4909

*kcal/mol.
†Angstroms.

point: results from a higher level of sampling, as quantified by the total number of orientations found, will not necessarily include all of the results from a lower level of sampling. Although a great deal of overlap is to be expected, the orientations actually obtained depend on the sphere–sphere distances, the atom–atom distances, and the docking parameters used in each run.

As noted previously, minimization may move orientations closer to or farther from the crystallographic position; in this case, however, there is a preponderance of orientations for which the RMSD decreases. This is particularly true for orientations that start out close to the observed binding mode (Fig. 3).

Periplasmic Binding Protein

Three families of orientations are produced when β-D-glucose is docked to periplasmic binding protein (Fig. 5). The lowest-RMSD structures reproduce the known geometry, in which glucose occupies the center of a tunnel traversing the protein. The intermediate-RMSD orientations overlap the crystal structure ligand but are flipped or rotated to various degrees. The high-RMSD structures are at the ends of the tunnel; apparently, constrictions prevent orientations from being evenly distributed throughout the tunnel.

At the highest level of sampling, the lowest-RMSD family of orientations is identified as the most favorable both before and after minimization; at the intermediate and low levels of sampling, at least one orientation resembling the observed binding mode is generated, but does not receive the best score until after minimization (Fig. 5 and Table III). The best-scoring orientation from each level of sampling prior to minimization is compared to the crystal structure in Figure 6.

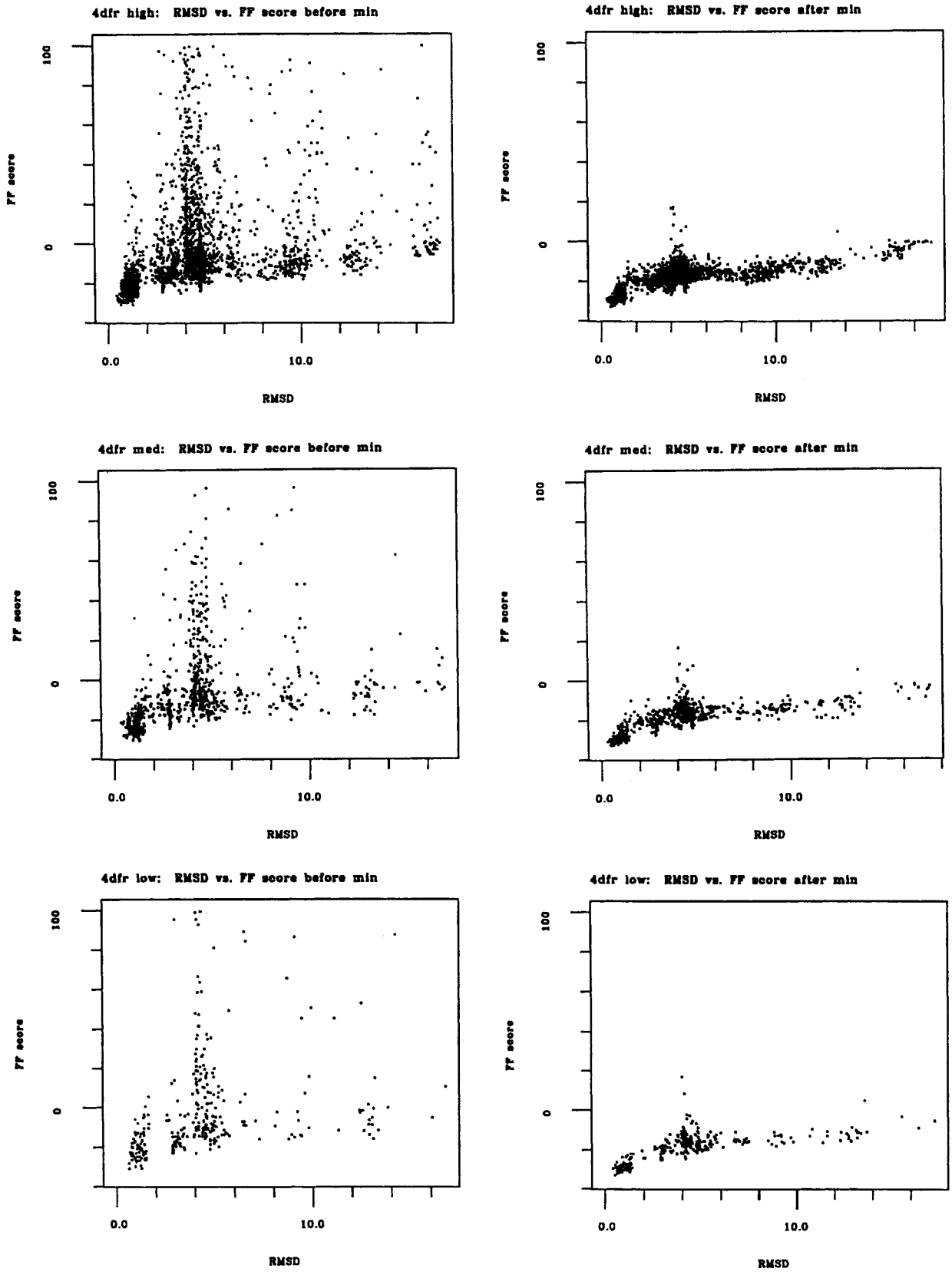

Fig. 2. 4dfr system: RMSD versus force field score at three levels of sampling, before and after rigid-body minimization.

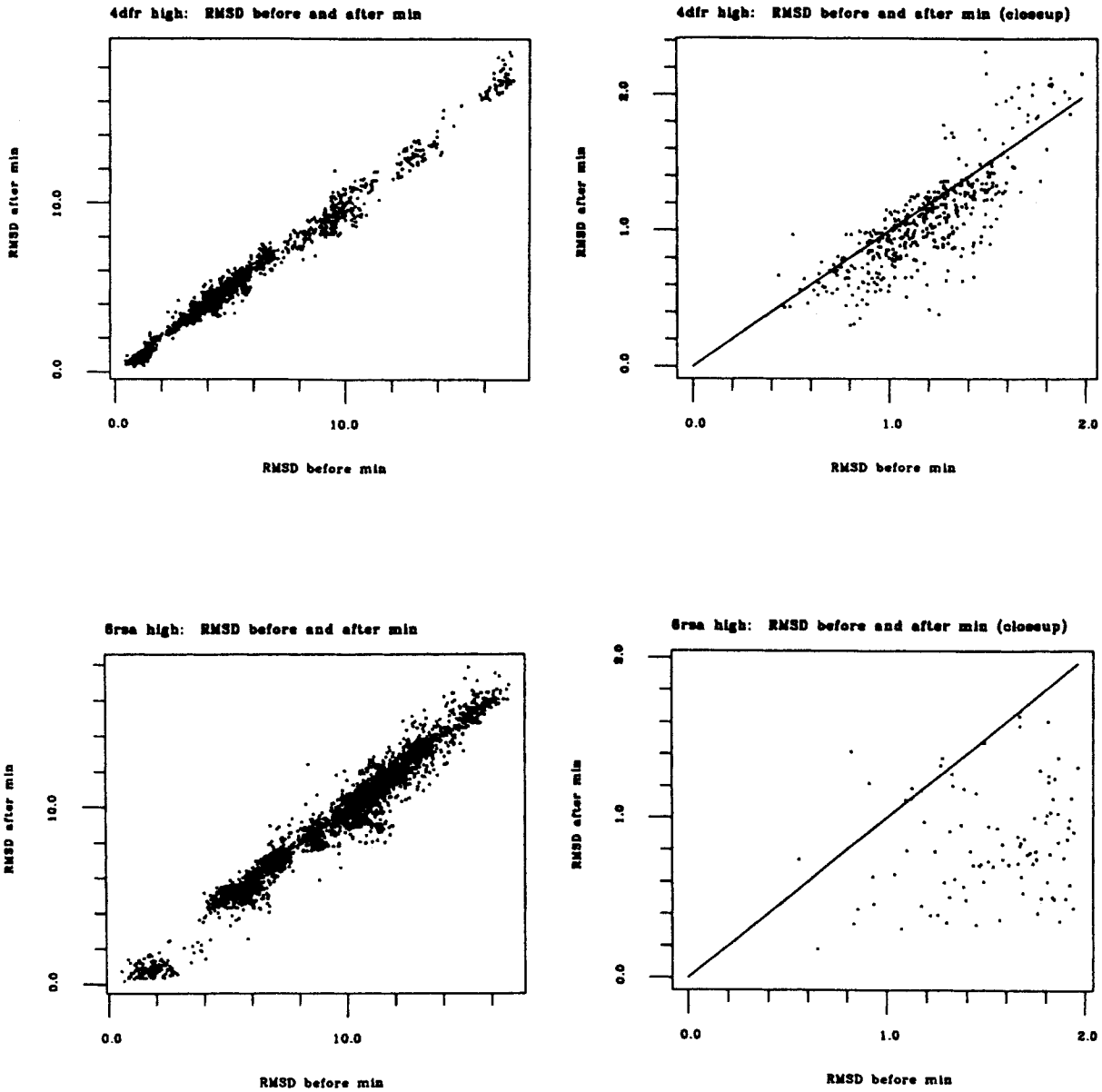

Fig. 3. RMSD values before and after rigid-body minimization. Results are for the 4dfr and 6rsa high-sampling runs, with closeups on the low-RMSD regions of the data (the solid lines pass through the origin and have a slope of 1).

Carboxypeptidase A

The crystallographic ligand glycyl-L-tyrosine was used for docking. The structure of the complex contains close contacts which hinder its regeneration by DOCK: the ligand phenolic oxygen is only 2.55 Å from the α-carbon of glycine-253 and one of the ligand carboxylate oxygens is only 2.23 Å from the side chain oxygen of tyrosine-248. Even so, RMSD values below 2.0 Å describe essentially the known binding mode. Structures with RMSD values of 3.0–4.0 Å are barrel-rolled and translated approximately a bond length along the long axis of the molecule, and structures with RMSD values greater than 6.0 Å are flipped end-to-end. As in the periplasmic binding protein system, members of the lowest-RMSD family of orientations receive the best scores both before and after minimization when sampling is intensive, but only after minimization when intermediate or low sampling is performed (Fig. 7 and Table III). The low-sampling run generates just one orientation similar to the crystallographic mode; if sampling had been slightly less thorough, the "correct" orientational family may have been missed altogether. Minimization can be extremely helpful if

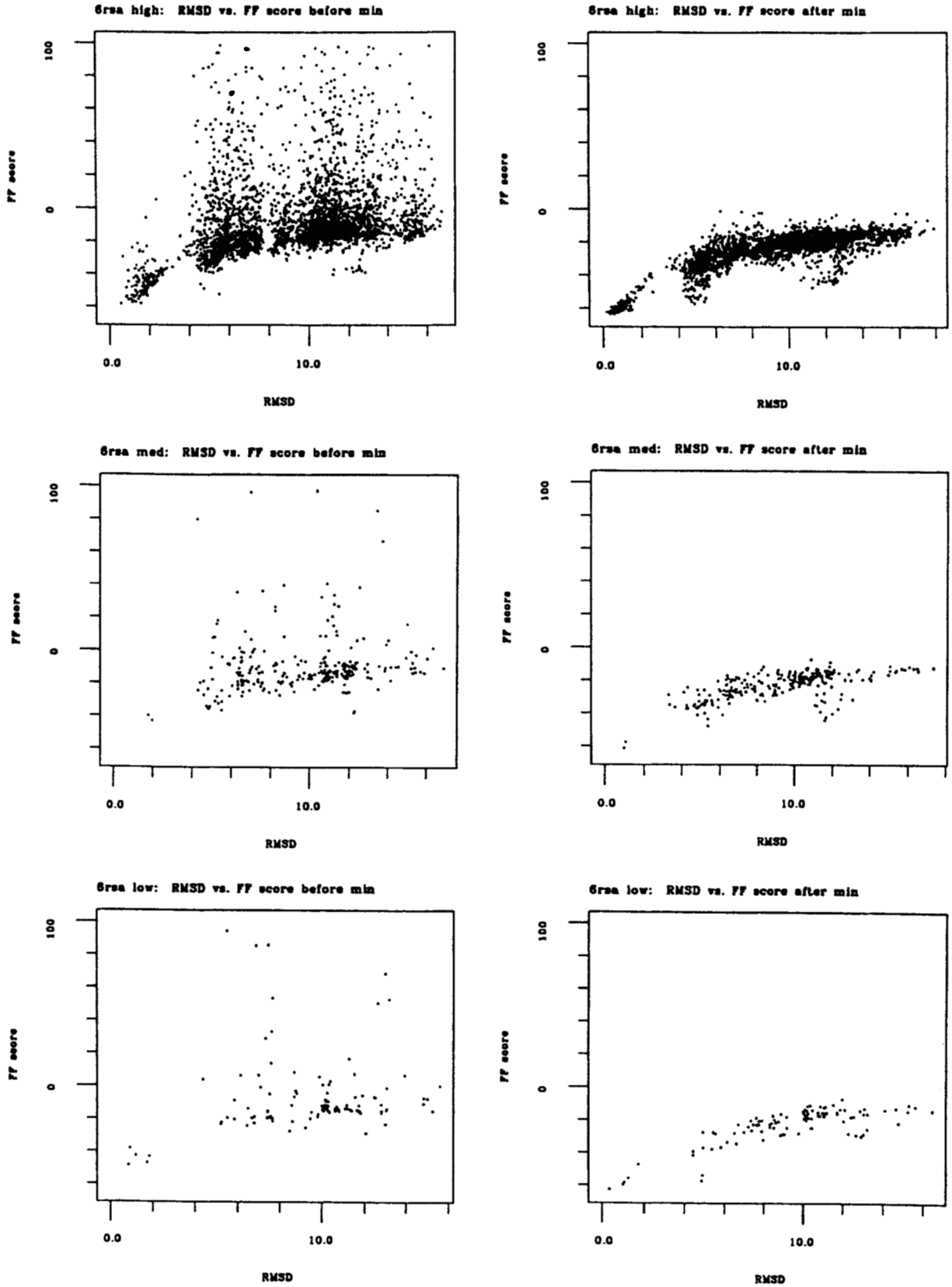

Fig. 4. 6rsa system: RMSD versus force field score at three levels of sampling, before and after rigid-body minimization.

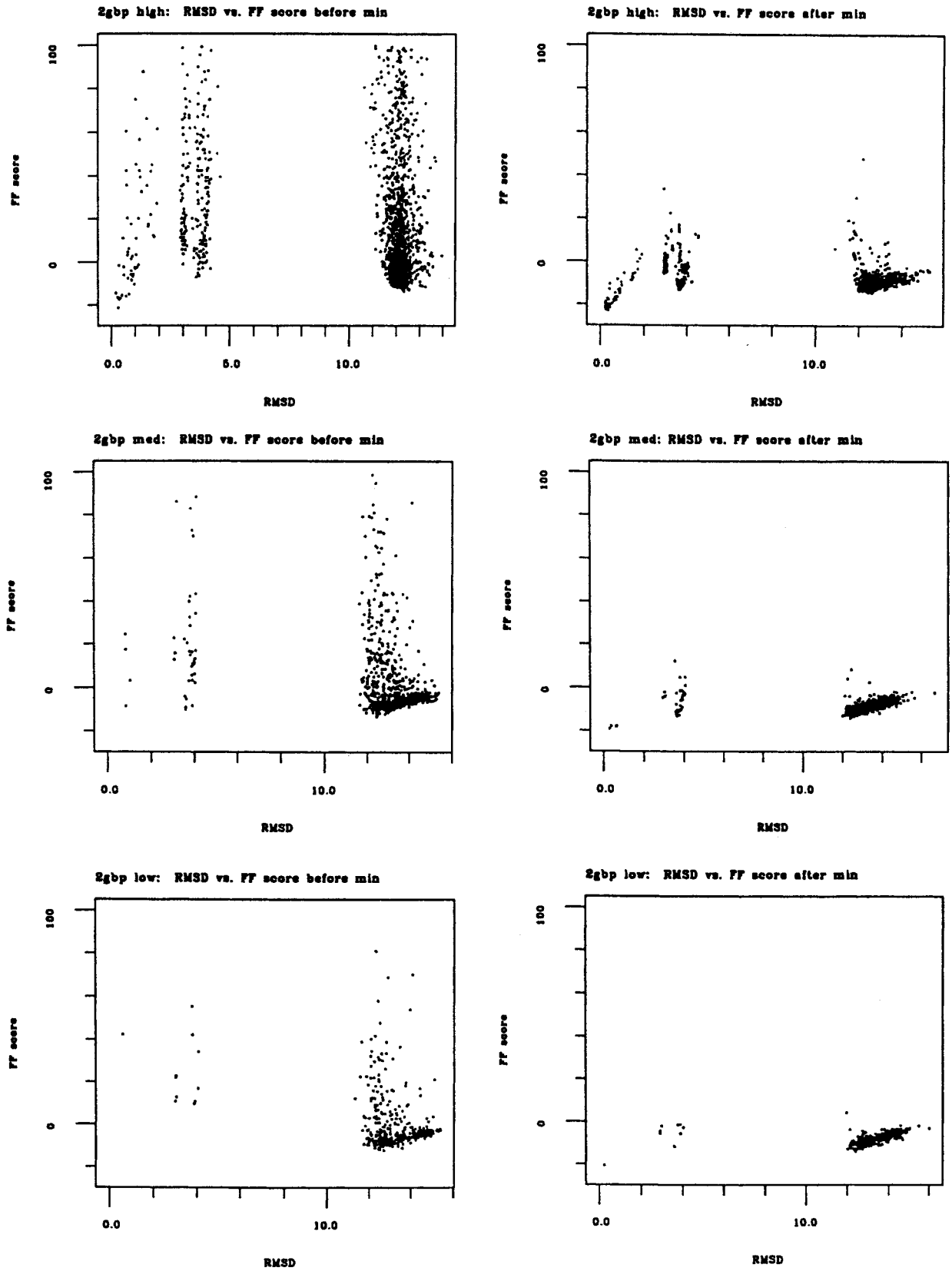

Fig. 5. 2gbp system: RMSD versus force field score at three levels of sampling, before and after rigid-body minimization.

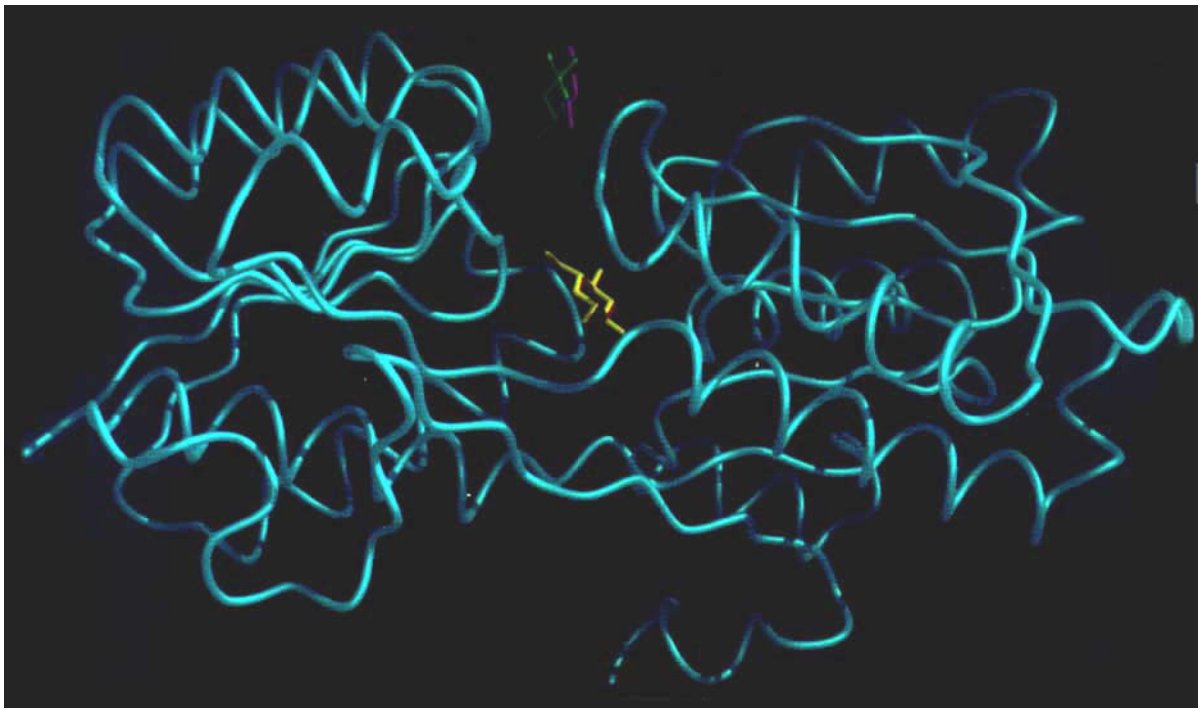

Fig. 6. The best-scoring orientations of glucose in periplasmic binding protein prior to minimization: magenta, low sampling; green, medium sampling; yellow, high sampling; red, crystallographic position. The Cα trace of the protein is shown in cyan. Figure generated with UCSF MidasPlus.[5]

sampling is rather sparse, but it cannot salvage a situation in which no orientations are close to the presumed true binding mode.

Overview

Although case-specific features are evident, general principles govern the numbers of orientations found and written (Table I), their RMSD distributions, the relationships between RMSD and score, and the effects of rigid-body minimization. The total number of orientations generated is a function of the number of ligand atoms and their distribution in space, the number of spheres describing the site and their distribution in space, and the DOCK distance-matching parameters. A majority of the orientations found are eliminated because they violate the close contact limits set in DISTMAP. Score cutoffs also decrease the number of orientations written, but their effects on RMSD distributions are quantitative rather than qualitative. RMSD values are diffusely distributed in the dihydrofolate reductase and ribonuclease A systems (Figs. 2 and 4). In these systems, a fairly continuous spread of values is possible because the docked ligand is small relative to the active site volume; a range of displacements in several directions is sterically allowed. Symmetry also affects the RMSD distributions; although there is not much extra volume in the periplasmic binding protein and carboxypeptidase A sites, rough ligand symmetry results in alternate binding modes that occupy the same volume as the known mode but have fairly large RMSD values (Figs. 5 and 7).

The most obvious and general result of rigid-body minimization is the marked decrease in the highest energies (Figs. 2, 4, 5, and 7). Virtually all of the positive FF scores drop to values below zero during rigid-body minimization; only a few stragglers, typically 1% of the orientations, are trapped in positive local minima. These significant changes in score, ranging up to more than 100 kcal/mol, are accompanied by small changes in RMSD (Fig. 3). The average absolute change in RMSD is 0.5 Å or smaller for each of the high-sampling data sets. The major effect of rigid-body optimization is clearly the relief of short-range steric repulsions. When a particular minimum has been well-sampled, optimization only lowers the best FF scores by a few kcal/mol; this is true of the lowest-RMSD family of orientations in all of the high-sampling runs (Table III). If a particular minimum has not been well-sampled, however, the change in score can be much larger; an extreme example is the single "correct" orientation from the 2gbp low-sampling run, for which the score drops more than 60 kcal/mol (Table III). It is in these situations that energy optimization is the most beneficial, allowing a successful identification of the known binding mode when the docking scores alone would have been misleading. Whether or not mini-

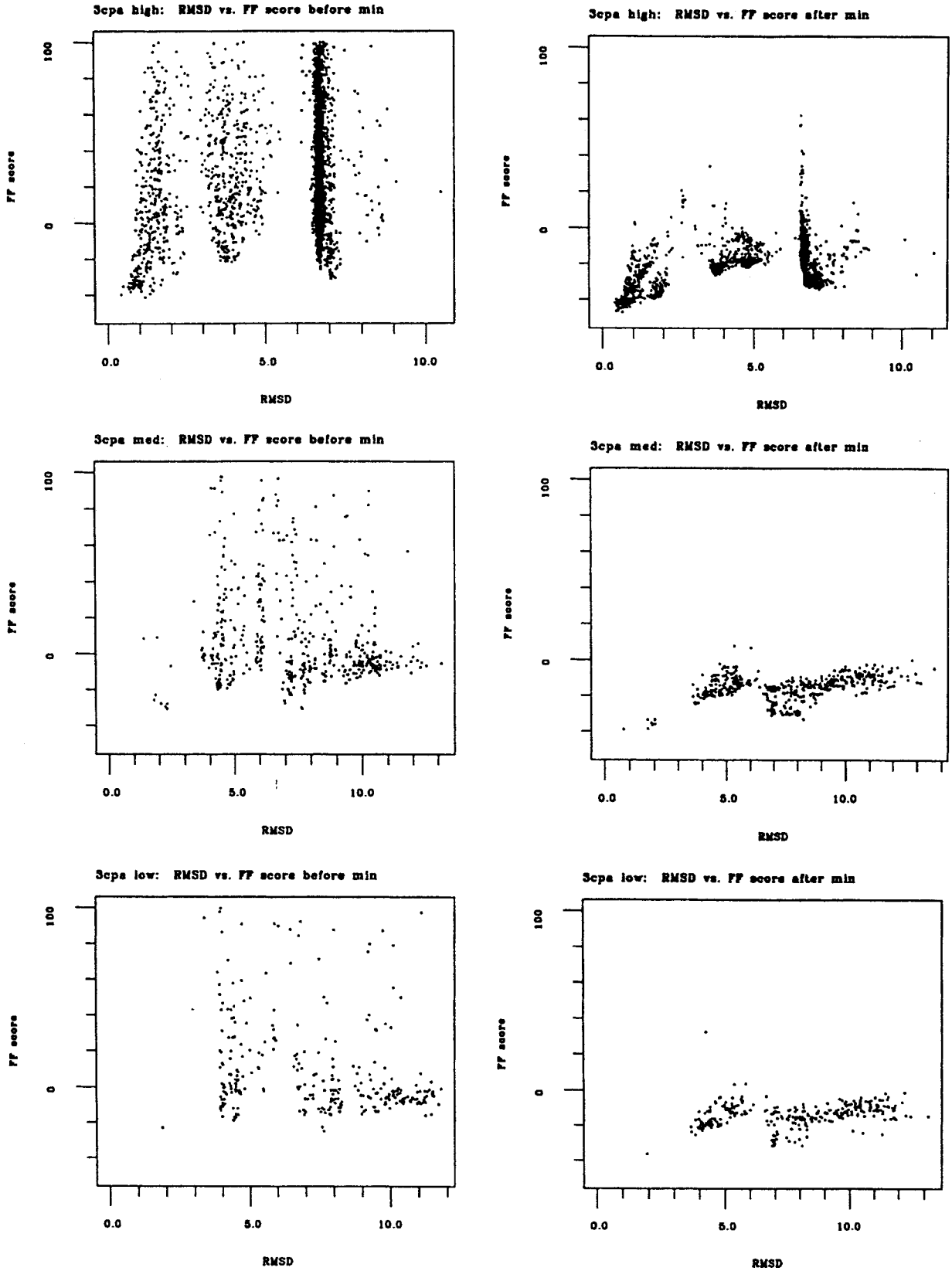

Fig. 7. 3cpa system: RMSD versus force field score at three levels of sampling, before and after rigid-body minimization.

TABLE V. Movement During Rigid-Body Minimization; RMSD From Starting Position*,†

System	Mean	Standard deviation	Minimum	Maximum	Median
4dfr	0.754	0.501	0.000	4.481	0.685
6rsa	1.022	0.708	0.000	5.083	0.869
2gbp	0.880	0.538	0.000	4.767	0.757
3cpa	0.766	0.387	0.001	3.103	0.700

*Angstroms.
†Statistics for orientations from high-sampling runs.

mization will change the rankings of families of orientations depends on how well they have been sampled, that is, whether or not sterically optimal members have already been found. If two families are both well-sampled, their separation in score may not change much during minimization (Table III; see "correct" versus "incorrect" in all of the high-sampling runs).

To round out the comparisons, the experimental ligand orientations were also subjected to rigid-body minimization (Table IV). Diaminomethylpteridine in dihydrofolate reductase moved to an RMSD value of 0.5 Å and its score decreased approximately 2 kcal/mol; some movement was expected since it is just a part of the larger crystallographic ligand, methotrexate. Uridine phosphate in ribonuclease A was essentially unchanged. Glucose moved very slightly (to an RMSD value of 0.3 Å) within periplasmic binding protein, but its score dropped 5 kcal/mol. Finally, the score of glycyltyrosine in carboxypeptidase A dropped significantly even though it only moved to an RMSD value of 0.5 Å; this small rigid-body adjustment relieved the close contacts present in the crystal structure (see above).

It is encouraging that in all cases the experimentally observed orientation is close to a minimum, likely the global minimum, in the FF scoring function. This conclusion is reinforced by the behavior of orientations roughly similar to the crystallographic orientation; nearly all move closer to the experimentally determined position during minimization. The average changes in RMSD for orientations with starting values of 2.0 Å or less from the crystal structures are −0.1, −0.7, −0.2, and −0.3 Å for the 4dfr, 6rsa, 2gbp, and 3cpa high-sampling data sets, respectively. With rounding, identical averages are obtained when all RMSD values are relative to the *minimized* crystallographic orientations. These average changes may be considered rough "radii of convergence," or indications of how much an orientation will move toward a nearby minimum. Another measure is the RMSD of each minimized orientation from its starting position; statistics are given in Table V.

In considering the time-efficiency of minimization, it should be noted that RGDMIN3 has not been optimized for speed, nor have alternate methods of minimization been explored. It is clear that pending further investigation, minimization will become much more cost-effective. A cutoff of 10.0 Å was used to allow a direct comparison with the DOCK FF scores. This could probably be halved without degrading the results, since the primary effect of minimization appears to be the resolution of short-range steric conflicts. Decreasing the cutoff decreases time requirements significantly. Speed can be increased 10- to 50-fold by using grid-based scoring during minimization. With the proper choice of parameters, the results are qualitatively identical to those from continuum minimization. Work continues on increasing the efficiency of minimization and identifying suitable subsets of orientations from DOCK to be minimized.

It is important to test scoring and sampling in systems where the binding geometries are known. In most applications, DOCK will be used in the absence of such knowledge, with the added uncertainty of which molecular conformations are the most appropriate. Furthermore, it is harder to make quantitative comparisons among different candidate ligands than among different orientations of the same ligand. Nonetheless, guidelines for sampling should be developed in systems where the results can be compared directly to experiment. Barring structural errors, any failures that occur in rigid-body docking of the known bound conformations must be due to sampling or scoring. It has been shown that the DOCK FF scoring function, among others, can identify the known binding modes in a number of test systems.[3] Thus, a primary goal of this work is to provide guidelines for the amount of orientational sampling that should be performed within DOCK. The sets of parameters in Table I are somewhat arbitrary, and indeed hardly begin to span parameter space. In general, increasing bin widths will increase sampling, as quantified by the number of orientations found. Setting the parameters to give an average of 10,000–20,000 orientations per compound is recommended for searching databases of candidate ligands; less intensive sampling may be acceptable if minimization is performed. More thorough sampling is recommended for single-molecule docking, even in combination with minimization, as it is computationally inexpensive.

CONCLUSION

There is a tradeoff between sampling and minimization; for the correct orientations to receive the best force field scores, intensive sampling is required, or low-to-moderate sampling combined with minimization. Minimization brings unfavorable energies down sharply, moves many orientations closer to the experimentally determined position, and allows identification of the known geometry when docking

alone would have been misleading. The tradeoff is not complete, however. Whether or not minimization is performed, sampling must be sufficient to find at least one structure in the vicinity of the true binding mode.

Although in the present work it is more efficient to sample orientations thoroughly than to combine docking with minimization (Table II), the hybrid approach is promising. Progress is being made on coupling more efficient algorithms with the selection of fewer orientations for minimization.

ACKNOWLEDGMENTS

We gratefully acknowledge support from NIH Grants GM-31497 (I.D.K.) and GM-39552 (G.L. Kenyon) and DARPA Grant MDA-91-J-1013 (F.E. Cohen). J.M.B. thanks Prof. Terry Lybrand for providing several critical subroutines and helping with the development of RGDMIN.

REFERENCES

1. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161:269–288, 1982.
2. Shoichet, B.K., Bodian, D.L., Kuntz, I.D. Molecular docking using shape descriptors. J. Comp. Chem. 13:380–397, 1992.
3. Meng, E.C., Shoichet, B.K., Kuntz, I.D. Automated docking with grid-based energy evaluation. J. Comp. Chem. 13:505–524, 1992.
4. 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. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
5. Ferrin, T.E., Huang, C.C., Jarvis, L.E., Langridge, R. The MIDAS display system. J. Mol. Graph. 6:13–27, 1988.
6. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., Kraut, J. Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Angstroms resolution. I. General features and binding of methotrexate. J. Biol. Chem. 257:13650–13662, 1982.
7. Borah, B., Chen, C.-W., Egan, W., Miller, M., Wlodawer, A., Cohen, J.S. Nuclear magnetic resonance and neutron diffraction studies of the complex of ribonuclease A with uridine vanadate, a transition-state analogue. Biochemistry 24:2058–2067, 1985.
8. Vyas, N.K., Vyas, M.N., Quiocho, F.A. Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. Science 242:1290–1295, 1988.
9. Rees, D.C., Lipscomb, W.N. Crystallographic studies on apocarboxypeptidase A and the complex with glycyl-L-tyrosine. Proc. Natl. Acad. Sci. U.S.A. 80:7151–7154, 1983.
10. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science 221:709–713, 1983.
11. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Jr., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
12. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 7:230–252, 1986.
13. Blaney, J.M. Ph.D. dissertation, University of California, San Francisco, 1982.
14. Fletcher, R. "Practical Methods of Optimization." New York: Interscience, 1960.