The MPSim-Dock Hierarchical Docking Algorithm: Application to the Eight Trypsin Inhibitor Cocrystals

ART E. CHO,* JOHN A. WENDEL, NAGARAJAN VAIDEHI, PETER M. KEKENES–HUSKEY, WELY B. FLORIANO, PRABAL K. MAITI, WILLIAM A. GODDARD, III
Materials and Process Simulation Center, California Institute of Technology, Pasadena, California 91125

Received 20 January 2004; Accepted 25 June 2004
DOI 10.1002/jcc.20118
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: To help improve the accuracy of protein–ligand docking as a useful tool for drug discovery, we developed MPSim-Dock, which ensures a comprehensive sampling of diverse families of ligand conformations in the binding region followed by an enrichment of the good energy scoring families so that the energy scores of the sampled conformations can be reliably used to select the best conformation of the ligand. This combines elements of DOCK4.0 with molecular dynamics (MD) methods available in the software, MPSim. We test here the efficacy of MPSim-Dock to predict the 64 protein–ligand combinations formed by starting with eight trypsin cocrystals, and crossdocking the other seven ligands to each protein conformation. We consider this as a model for how well the method would work for one given target protein structure. Using as a criterion that the structures within 2 kcal/mol of the top scoring include a conformation within a coordinate root mean square (CRMS) of 1 Å of the crystal structure, we find that 100% of the 64 cases are predicted correctly. This indicates that MPSim-Dock can be used reliably to identify strongly binding ligands, making it useful for virtual ligand screening.

 J Comput Chem 26: 48–71, 2005

Key words: protein–ligand docking; clustering; conformational search; enrichment; trypsin cocrystals

Introduction

An ultimate goal of the bioinformatics revolution is biofunctionomics, the prediction of the function of every protein in the genome.[1] The first step in predicting function is to predict where and how strongly do various ligands bind to the protein. There has been a great deal of progress in using theory and computation to dock small ligands to the 3D structures of proteins;[2–8] however, there remains concern as to whether the current methods are sufficiently accurate to locate the proper binding site in the absence of a cocrystal protein structure. We have shown that the HierDock hierarchical docking strategy can be used to scan the entire protein to identify the putative binding region[9–14] and subsequently target the binding region sufficiently thoroughly to identify the binding conformation.[9–15] This has made it practical to predict ligand-binding site in proteins and to calculate relative binding energies for VLS procedures function of proteins.[16] However, a docking procedure when used to dock ligands to the apo-protein conformation, should ensure that the conformational search is comprehensive, and the use of HierDock has not been validated for the conformation selection part of this procedure.

In this article we consider the validation of the MPSim-Dock method for searching the target putative binding region to locate the binding conformation. We will assume that the target region has been identified by other methods such as ScanBindsite[16] or Pass[17] or by having a cocrystal with some other ligand and we will consider how reliably we can identify the exact binding conformation and calculate binding energy. MPSim-Dock combines the matching method, in which trial ligand conformations are matched onto the ligand binding site on the receptor as in DOCK4.0[18] with the more accurate all-atom force field (FF) energy expression and continuum solvation methods in MPSim.[19] The method also ensures that the ligand conformations generated are diverse by a diversity factor, to make the conformational search comprehensive. MPSim-Dock method is tested for the case of eight trypsin inhibitors cocrystallized with trypsin. Each cocrystal has a slightly different protein structure to accommodate the shape of each ligand. In some cases the ligand–protein binding is mediated by tightly bound waters and in other cases it is not. There are experimental inhibition constants K_is available for seven of the eight

Correspondence to: W. A. Goddard, III; e-mail: wag@wag.caltech.edu

**Present address:* Center for Biomolecular Simulations, Columbia University, New York, NY 10027

Contract/grant sponsor: NIH; contract/grant numbers: RO1-GM62523-01, RO1-AI40567, and RO1-CA85779

inhibitors,[20,21] thus allowing us to compare both structure and binding energy.

We consider each of the eight trypsin cocrystals to define a fixed protein to which the binding structure and energy for all eight ligands is predicted. Because the root-mean-square deviation in coordinates (CRMS) difference between the eight crystals ranges up to 1.02 Å, this provides a best-case example of how useful the docking techniques will be for predicted structures (either by homology to known structures or by first principles predictions, which are often be much less accurate). The criterion used for success in conformational sampling is that the ligand conformation predicted to have the lowest energy or the conformation with an energy very close to the best (say within 2 kcal/mol of the best) should lead to a structure within a CRMS error of 1.0 Å from the cocrystal structure. With this level of accuracy we expect that practical subsequent steps of simulation in which the ligand–protein complex is optimized, (perhaps with explicit water and molecular dynamics at the appropriate temperature) will lead to accurate structures and energetics. Considering the most realistic case of docking with NO WATERS in the protein, the MPSim-Dock methods discussed herein, is successful for 100% of the $8 \times 8 = 64$ cases considered in this article.

Section 2 describes the methods, while Section 3 presents the results and discussions. The summary is in Section 4.

Methods

Force Fields

We used the Dreiding all-atom FF[22] for both the protein and the ligand. This includes the internal energy of the ligand along with the nonbond energy of interaction with the protein. The energy expression includes valence energies (bonds, angles, torsion, and inversion) and nonbond interactions (Coulomb, van der Waals, and hydrogen bond) within the ligand and with the protein. The charges for the protein were taken from Charmm22 FF[23] while the ligand charges were from Gasteiger.[24] The crystallographic waters were described with the TIP3P FF[25] and charges. The combination of Dreiding FF with Charmm22 charges has been validated elsewhere for studying ligand binding to proteins.[9–16] The nonbond interactions were calculated using the Cell Multipole method[26] to calculating the nonbond interactions. The molecular dynamics (MD) and minimization calculations were all performed using MPSim. The energy scores in MPSim-Dock are based on the all atom ligand FF energy with protein fixed. The Analytical Volume Generalized Born (AVGB) continuum solvation method[27] has been used to calculate solvation energies of the ligand–protein complex (and the separated protein and ligand). AVGB includes both the electrostatic energy and a cavity term derived from solvent–solute interactions.[27]

Preparation of Structures

Starting Structures

This article considers the eight inhibitors for the serine protease, trypsin for which there are high-quality crystal structures for the ligand–trypsin cocrystal. In each case the target protein is the same, but the binding of the ligand leads to a slightly different conformation. The protein conformations in the eight cocrystal structures differ by CRMS from 0.27 to 1.02 Å. Thus, these eight systems validate the docking methods both for Virtual Ligand Screening (VLS) and for target protein conformation selection. Because each of the eight protein conformations serves as a target structure to dock eight ligands there are a total of $8 \times 8 = 64$ cases for the docking tests.

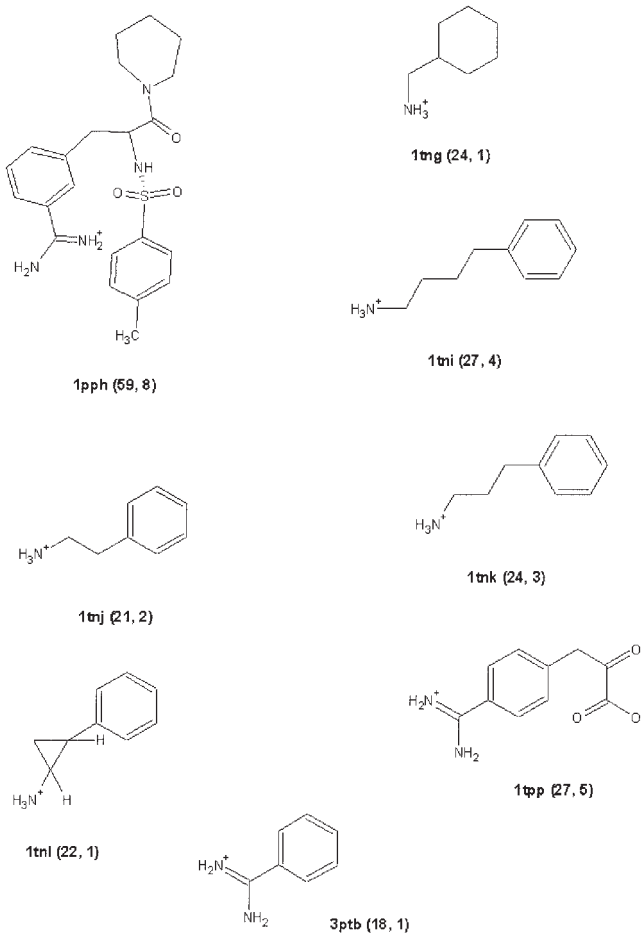


Figure 1. Eight ligands were used for our investigation. PDB codes are for the native cocrystal complexes. Crossdocking was done without modification on these ligands. The two numbers in parentheses are the number of atoms and the number of rotatable bonds, respectively.

The eight crystal structures with pdb codes 3ptb, 1tng, 1tnj, 1tni, 1tnk, 1tnl, 1tpp, and 1pph were downloaded from the protein databank. The ligands for these cocrystal structures are shown in Figure 1. The ligands used in this study range from small (3ptb) to medium-large (1pph) in size having from one to eight rotatable bonds. Except for 1tnl ligand, they are all positively charged. Missing side chains were added using SCWRL,[28] and all hydrogens were added explicitly using PolyGraf. The charge on the side chains of Arg, Lys, and Asp, Glu were neutralized with Cl^- and Na^+ respectively, except that we ensured that this was not done for salt bridges and that no such ions would interfere with the ligand

Table 1. Comparison of the Force-Field Minimized Structures of Eight Cocrystals of Trypsin with the Structure from X-ray Crystallography.

	1tng	1tni	1tnj	1tnk	1tnl	1tpp	1pph	3ptb
All atom	0.65	0.73	0.56	0.67	0.54	0.56	0.79	0.73
Binding site	0.61	0.68	0.48	0.32	0.35	0.4	0.84	0.39

Coordinates RMS error is in Å. The first row is comparison of all atoms of proteins and the second row is for only the binding site atoms.

in the binding pocket. The binding pocket of each protein was analyzed for the protonation states of histidines using What-If code.[29] The various His residues were protonated at the Nε, in accordance with the What-If results. The eight cocrystals were then optimized by minimizing the energy while including the forces due to solvation by water. This was accomplished using the Surface Generalized Born (SGB) continuum solvation method.[30] These SGB calculations used an internal dielectric constant of 2.5 and an external dielectric of 80.37.

The eight cocrystal structures all contain well-defined crystallographic waters, some of which are in or near the binding site, and in some cases the water plays an important role in determining structure and energy. To explore the best strategies, we considered three cases:

1. *No waters* (eliminating all water from the binding site): this is the only realistic level for docking studies because the waters suitable for one ligand would generally be different for others.
2. *Conserved waters* (keeping only those waters present in all eight cocrystals). These waters are likely the most strongly bound, and hence the ones that might well be predicted without crystal structure information. This is mainly useful for comparing how well the predicted binding energies compare with experimental values for all eight ligands.
3. *All waters* (keeping every water molecule in the crystal structure). This is not useful for docking but it is useful in validating how well the predicted binding energies compare with experimental values over the eight ligands.

Minimization of the Target Cocrystal Structures

The first step was to minimize the energy of the ligand–protein complex taken from the cocrystal structures using the Dreiding FF and charges as described above. These calculations used MPSim to carry out conjugate-gradient minimization to a convergence criterion of 0.1 kcal/mol/Å. The first step optimized all hydrogen atoms without allowing any other atoms to move. This includes all hydrogens on the protein, hydrogens on the water molecules (if present), and hydrogens on the ligand. This is considered an important step because the experimental structural data does not provide information about the hydrogens.

The second step of minimization kept all atoms of the ligand fixed while the rest of the protein structure was allowed to optimize around it. All protein atoms, water molecules (if present), and ions were allowed to adjust to the crystallographic coordinates of the ligand. Again, a convergence of 0.1 kcal/mol/Å or was used. Because the waters being eliminated from the structure may change the binding mode, we want to find the no-water-present structure most compatible with the ligand. This allows us to assess our accuracy by comparing to this structure.

Finally, the crystal structures of the whole protein with the ligand and counterions Na^+ and Cl^- present, were minimized. This was done for all three cases: (1) all crystallographic waters, (2) only conserved waters, and (3) no waters.

FF Validation Studies

The energy-minimized structures for the cocrystals are compared to the starting crystal structures in Table 1. The CRMS for corresponding atoms was calculated for structures before and after minimization and for the whole protein and also for the binding site (defined as those residues within 5 Å of the ligand, which includes 23 residues of protein).

Consider first the case in which all crystallographic waters are included. Here the CRMS error for the whole ligand–protein complexes varies from 0.56 to 0.79 Å. This represents the errors due to minimization (rather than dynamics at room temperature), the FF and charges, crystal packing effects in the experimental structure, and error in the crystal structure. This good correspondence to the crystal structure validates that the FF preserves the global minimum defined by the crystal structure. Thus, we consider that an accuracy of 0.6 Å represents agreement with experiment.

Table 1 also shows the CRMS deviation from the crystal structures for the residues in the binding site. These CRMS differences range from 0.32 to 0.84 Å. This suggests that 0.6 Å difference between the predicted structure and experiment is reasonable. The protein conformations from these minimized structures will be used as target protein for docking studies. These crystal-minimized structures were also used as reference structures for all calculation of CRMS to docked structure.

Validation of the All-Atom Energy Scoring Function

Absolute Binding Energies

A major goal in docking studies is to obtain the relative binding energies for different ligands. Thus, an important but stringent test for the scoring function used in the docking algorithm is to compare the calculated binding energies using the scoring function to the experimentally measured binding constants for each ligand. To this end the binding energy for the minimized crystal structures for each of the eight trypsin inhibitor cocrystals was calculated.

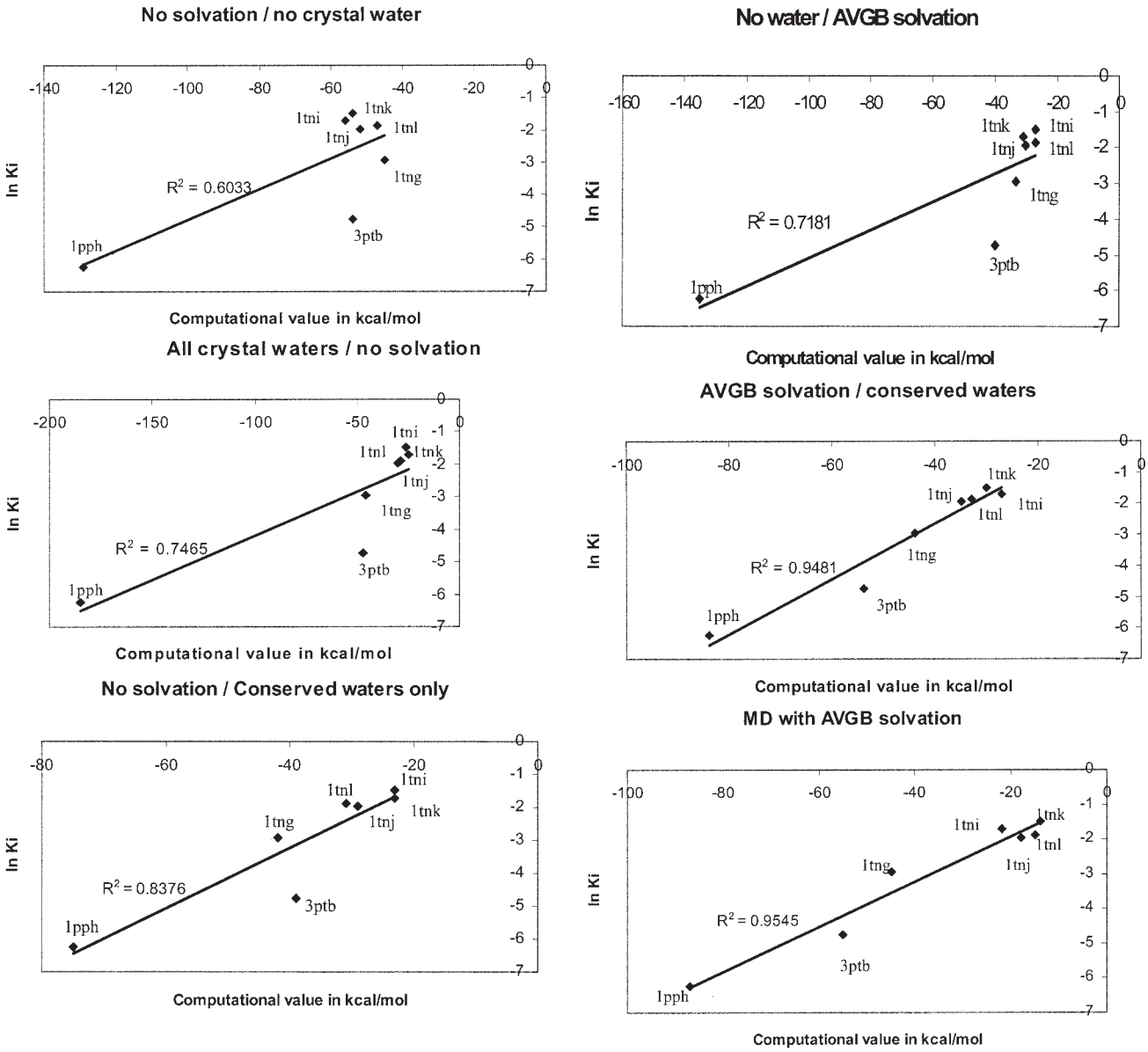


Figure 2. Comparison of calculated binding energies for eight trypsin/inhibitor complexes (diagonal cases) to the experimentally measured binding constants. (a) Case with no implicit solvent effect with no crystal water. (b) Case with no implicit solvent effect with all the crystal waters. (c) Case with no implicit solvent effect with only conserved crystal waters. (d) Case with analytic volume generalized Born (AVGB) solvation with no crystal water. (e) Case with AVGB solvation with only conserved crystal waters. (f) Case with molecular dynamics with AVGB solvation with no crystal water.

These calculations feature different setting for inclusion of crystallographic waters and continuum solvation. The binding energy (BE) was calculated using

$$\mathrm{BE} = \mathrm{PE(protein)} + \mathrm{PE(ligand)} - \mathrm{PE(protein/ligand\ complex)} \quad (1)$$

where PE is the total (potential) energy from the FF. In these calculations the free protein and free ligand were also solvated and their structures were minimized using SGB, but the side-chain conformations of the protein were not recalculated. For each of the three components in eq. (1), the final binding energy was calculated including solvation effects using the accurate AVGB continuum solvation method.[27]

The calculated binding energies are compared to the measured binding affinities for the trypsin case in Figure 2 and Table 10. This table and figure show a few different settings for our calculations. The best correlation factor for these eight cases is 0.95, in

excellent agreement with the experimental binding affinities. This validates the use of our FF for calculating relative binding energies of various ligands. Despite this excellent agreement, it should be noted that the present calculations do not take into account the fluctuations in the ligand conformation with room temperature dynamics. Including these effects would tend to narrow the difference between the best and worst binding energies. Nor have we included the entropic contributions that are important at room temperature. Including this would make all of the binding energies much weaker.

Monotonicity of the Scoring Function

The all-atom FF energy would be used to select the best ligands among all the docked ligand conformations. For this purpose it is imperative that the scoring function can discard the low binding affinity ligand–protein structures (say worse than 2 Å CRMS) while recognizing the good structures (say better than 1 Å CRMS). Thus, to have a reasonably monotonic relation between binding energy and CRMS for all ligand conformations with CRMS <2 Å would be desirable. To test this 500 docked ligand conformations were generated with DOCK4.0, and the corresponding binding energies for each conformation was calculated using the MPSim-Dock scoring function and plotted it (solid symbols in Fig. 2) as a function of CRMS (compared to the crystal-minimized structure). This shows that for docked conformations having a CRMS greater than 1.5 Å the energy function does not reliably select the smallest CRMS. Thus, to reliably identify good ligand conformations, the Dock procedure must be sufficiently robust that it finds some conformations with smaller CRMS (say <1 Å) so that the energy be useful in identifying the good ligand conformations.

To explore the relation between energy and CRMS for various ligand conformations with CRMS <1 Å, a number of near-native ligand conformations were generated by performing annealing MD using MPSim. Here we started from the minimized cocrystal structures and heated the system from 50 and 600 K with 1 ps for each 10-K step. Then the system was cooled from 600 K back down to 50 K. The snapshots of the ligand conformation were extracted every 10 fs from the annealing MD trajectory. This generated a number of ligand conformations in the protein that have CRMS values <0.5 Å. Figure 3 shows that the energy function is nearly monotonic below a CRMS value of about 0.75 Å. This suggests that if the docking procedure generates ligand/protein complexes within 0.75 Å of the native structures, we can probably identify them as good structures based on the MPSim all-atom FF energy.

Preparation of Trypsin Off-Diagonal Cases

Tests in which the ligand is docked to a protein with exactly the right structure to bind to the ligand are not adequate for testing the docking protocol. In practice the protein structure will always be somewhat off from the best one for the ligand. To obtain some measure of this effect, one of the eight protein conformations was used as the target protein for binding the other seven ligands. These cases are referred to as *off-diagonal cases,* while *diagonal case* is used to refer to docking of the ligand to the same protein

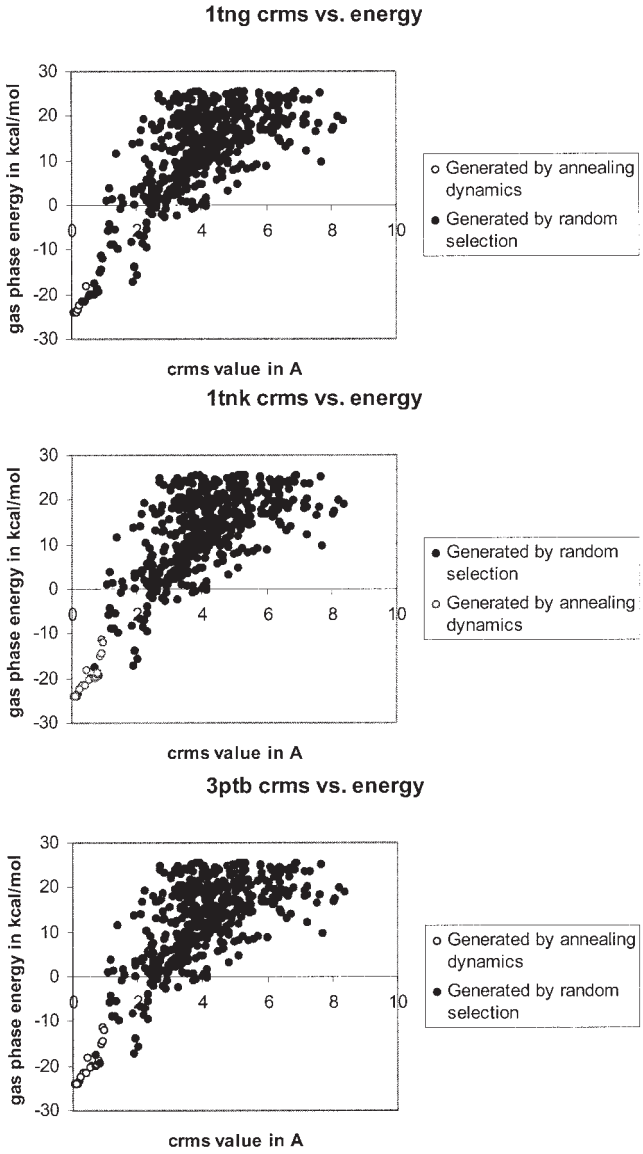


Figure 3. Energy values near crystal structure with respect to the CRMS values. (a) Case of 1tng cocrystal. (b) Case of 1tnk cocrystal. (c) Case of 3ptb cocrystal.

structure as the cocrystal. Thus, each of the eight ligands has been matched to each of the eight protein conformations from its cocrystal structure.

To assess the accuracy for predicting the structure of off-diagonal cases, we need the structure for the "exact" ligand–protein complex where the protein conformation is optimum for a different ligand. To prepare this exact off-diagonal ligand–protein target structure, we minimized each of the eight ligands in each of the eight protein conformations (64 cases total) with protein atoms fixed. Binding energies of each ligand in every other target protein conformation were calculated again using eq. (1), and reported in Table 2.

Table 2. Binding Energies (kcal/mol) of the Eight Ligands to the Eight Trypsin Cocrystals (After Minimization).

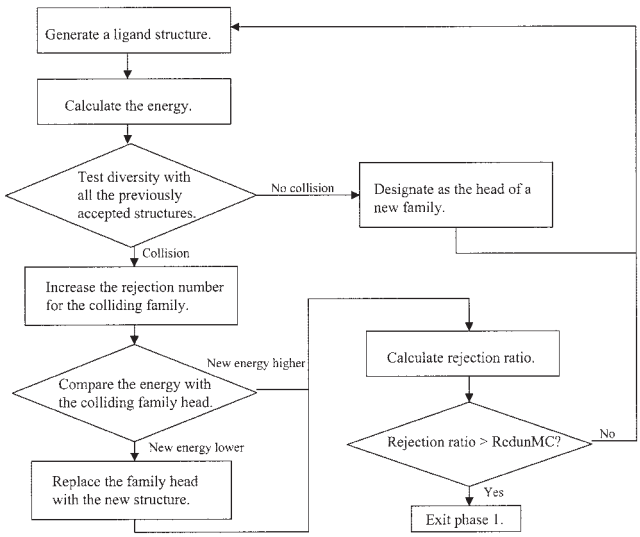
rec/lig	1tng	1tni	1tnj	1tnk	1tnl	1tpp	1pph	3ptb
1tng	42	43	40	41	31	29	61	31
1tni	22	23	26	22	21	42	69	36
1tnj	26	28	29	30	27	26	64	36
1tnk	25	22	23	23	25	41	69	35
1tnl	30	31	33	33	31	39	71	39
1tpp	36	39	32	36	32	54	76	37
1pph	40	43	40	44	39	25	75	39
3ptb	28	28	23	16	27	40	72	39

In these calculations only conserved waters were included in the active site.

MPSim-Dock

The six main steps in MPSim-Dock are (see the flowchart in Scheme 1):

1. *Mapping the potential ligand binding sites:* for each of the eight cocrystal structures, the program Sphgen[31] of DOCK4.0 was used to construct an inverse sphere representation of the protein surface. The set of spheres in the binding region around the cocrystal ligand was chosen for docking. The final set of spheres was constructed by eliminating spheres that are too close (less than 1 Å) to each other. The number of spheres used in the cluster varied between 57 and 87, as listed in Table 11. A probe of 1.4 Å radius was used to generate a molecular (Connolly) surface with 5 dots/Å molecular surface.[32] These spheres serve as the basis for the docking method.
2. *Conformational search:* we generate conformations (that is, orientations and locations) of the ligand in the binding site we match the intersphere distances to the interatom distances in the ligand as in DOCK4.0. We reject any conformation that has more than "MaxBump" bad contacts. Here, MaxBump = Max (2, 0.1 *N*) where *N* is the number of atoms in the ligand. A bad contact is defined as in DOCK4.0.
3. *Diversity:* the orientations generated in DOCK4.0 are statistically random but need not sample every possible topology. Thus, to ensure that the search over the conformation space is comprehensive, the conformations generated are categorized into families. Two ligand conformations are defined to be in the same family if their CRMS difference is within DivPar (default = 0.6 Å). To select this criterion for diversity (DivPar) we note that starting with the X-ray structure of the ligand and minimizing the energy leads to CRMS differences of ~0.4 to 0.8 Å (see Table 1). Also, Figure 2 shows that the energy for Dock4.0 generated conformations is reasonably monotonic with CRMS up to 0.75 Å. This suggests that one could identify a good conformation based on the energy if it is within 0.75 Å of the exact structure. Thus, we consider all conformations within a CRMS of DivPar to belong to the same diversity family (DF) and that any one of these will minimize to the same structure. The family head for each different DF differ from every other family head by DivPar. Thus, if all DF covering the active site are generated and properly selected using all-atom FF energy, the best conformation should be within DivPar = 0.6 Å from the exact answer. The diversity of various conformations is checked as follows. For each conformation generated by Dock4.0 that passes the bump check, the CRMS with the family head for all previously retained DFs is calculated. If the new conformation is farther than DivPar (default = 0.6 Å) in CRMS from the family heads of all previously generated DFs, then this conformation is chosen as the family head for a new DF and retained. The MPSim energy of this ligand conformation is calculated (using the all-atom Dreiding FF with the protein atoms fixed). If the new orientation is within DivPar of any previously generated orientations, then the energy of the new conformation is calculated and compared to that of the current family head. If this new conformation is lower in energy, then it replaces the older one to become the family head of this family. In either case the counter for the number of conformations generated for this family is incremented.
4. *Completeness:* the next question is how many DFs should be generated to ensure that the binding site is sufficiently spanned.



Scheme 1. Schematic flow of the first phase of MPSim-Dock.

To determine when the set is complete, new conformations are generated until most new conformations are rejected because they overlap with any previous DF. To determine how to find the point when we have a complete set of DF, consider Figure 4, which plots the number of diversity families vs. number of generated conformations for the 1tnk–1tnk complex. This shows that the number of accepted structures saturates at ~2800 generated structures, which corresponds to a redundancy ratio = 2800/1209 = 2.2. To avoid statistically improbable events that would terminate this too early, the minimum number of DFs is taken to be DivMin = 500. We took the maximum number to be DivMax = 10,000. The number of diversity families is denoted as NDivFam. For each family the energy of the family head (which includes the internal energy for bonds, angles, torsions, and nonbond interactions) is calculated. We found that redundancy ratio of 2.2 guarantees a close match to the correct site for all 64 systems discussed herein. Next, the NdivFam is sorted by the energy of the family head (all atom FF energy of the ligand in fixed protein) and then the top NEnrich (default = 50) family heads are selected by energy, for the enrichment step to be discussed next.

5. *Enrichment of high-ranking families:* it has been observed that sometimes a family head generated has a good match to the exact structure (CRMS <0.75 Å) but with an energy that is high due to a slightly bad contact. This family might be rejected for subsequent optimization because of the high energy. To avoid this situation it is prudent to sample a sufficiently large number of conformations *within* each family, so that selection on the basis of energy would always find a family with low CRMS to the correct family. The strategy in the enrichment step is to have a sufficiently rich description of the ensemble of conformations for each DF, that we can select the best DF by examining energy. To reasonably guarantee this we find that when there are four or more members in the DF corresponding to the exact structure, we nearly always get a low energy for this DF. This motivates the enrichment step. Thus, after identifying the NEnrich = 50 best families from the Completeness step, the conformations are continued to be generated but this time, after checking the bumps, the conformations are accepted and energy calculated only if it belongs to any of the top NEnrich families (i.e., within DivPar of the family head). This enrichment cycle is stopped when there is an average of AveEnrich (default 4) members per family. An alternative is to designate a minimum number of members to be generated for each and every family. After completion of the enrichment step the NEnrich * (AveEnrich + 1) = 50 * 5 = 250 conformations is sorted by energy and the lowest NoptDF (default 5) DFs is selected for the optimization step.
6. *Optimization of ligand in fixed protein site:* the best energy from the enrichment step is not always the structure closest to the exact answer. However, we find that it is always in the best scoring five conformations. Thus, we select the NoptDF (default 5) best DFs (lowest energy) and minimize the ligand in the fixed protein using conjugate gradient minimization for NminDF steps (default = 10). This minimization is done to resolve the bad contacts that might arise in the conformation generation, but not to find a global minimum (the convergence criterion for a fully converged state is to achieve an RMS force less than 0.1 kcal/mol/Å). Our goal here is to reliably pick the right ligand conformation for the later (much more expensive) full ligand–protein optimization step.

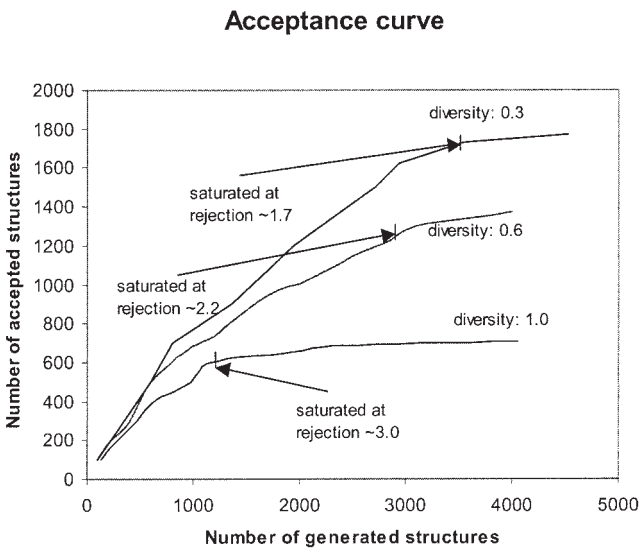


Figure 4. Number of accepted structures as a function of generated structures for 1tnk–1tnk for various diversity parameter.

Results and Discussion

Results on Using Diversity and Enrichment in MPSim-Dock of Top Scoring Families

Choice of Diversity Parameter

Figure 4 shows the number of conformations accepted as a function of the total number of conformations generated (after passing bump tests) for the 1tnk diagonal case using three distinct DivPars. Each curve eventually reaches a plateau defining the proper rejection ratio for saturation of the binding site. Beyond this saturation value (for a given DivPar) few new families are found, so that it is of little use to generate conformations beyond this point. The total number of accepted structures is dependent on the size of the search space and the number of spheres used in MPSim-Dock. However, we are only concerned with the completeness of conformational search, and therefore, we use the rejection ratio as a guideline for whether or not the conformational space has been searched thoroughly. We chose the DivPar to be 0.6 Å, based on the monotonicity of our scoring function and the results of our computational tests. Based on Figure 4, we see that the rejection ratio should be redundancy ratio of 2.2 for DivPar = 0.6 Å.

Need for Enrichment of Top Scoring Families

Table 3 shows the top 50 families and their energies for the case of docking the ligand of 3ptb into the target 3ptb trypsin protein structure. The criterion used to assess a good conformation is that it should be within 1.0 Å of the crystal structure. Indeed, the top

Table 3. The Energy and CRMS to the Exact Structure for the Family Head of the Top NEnrich = 50 Families of 3ptb–3ptb Selected After the Family Generation Step d But Before the Enrichment Step of MPSim-Dock.

Number	Energy	CRMS value
1	−37.18	0.57
2	−27.15	0.35
3	−22.99	1.87
4	−22.33	2.03
5	−22.23	3.18
6	−22.20	5.67
7	−20.49	0.86
8	−20.42	4.10
9	−20.28	6.87
10	−19.37	2.91
11	−18.78	3.94
12	−18.56	2.87
13	−18.43	1.30
14	−18.32	6.85
15	−18.29	0.73
16	−18.05	1.07
17	−17.91	3.66
18	−17.90	4.63
19	−17.78	3.83
20	−17.50	3.72
21	−16.95	3.73
22	−16.87	6.29
23	−16.05	5.80
24	−15.78	4.21
25	−15.62	1.57
26	−15.43	5.30
27	−15.39	3.11
28	−15.20	6.26
29	−15.12	5.69
30	−15.09	3.02
31	−14.72	3.64
32	−14.68	6.82
33	−14.45	3.60
34	−14.26	3.71
35	−14.21	3.62
36	−14.03	6.62
37	−14.02	1.31
38	−13.85	6.21
39	−13.33	6.38
40	−13.27	5.27
41	−13.15	3.62
42	−12.68	2.22
43	−12.56	5.56
44	−12.44	2.04
45	−12.13	4.07
46	−12.05	3.66
47	−11.98	1.73
48	−11.96	4.10
49	−11.87	3.31
50	−11.08	4.14

scoring orientations generally have low CRMS to the crystal structure. But there are exceptions. We found some cases for which conformation within 1.0 Å of the crystal do not have the best energy score. To prevent this situation, which would lead to poor selections for subsequent optimizations, we enrich the NEnrich = 50 top scoring diversity families by generating more orientations for each family (within DivPar = 0.6 Å of the each family head). This procedure is extremely fast, because one only has to evaluate the CRMS value of the newly generated conformation to the NEnrich = 50 family heads. Only if these new structures fall within DivPar of any of the 50 family heads does one calculate the energy. Table 4 shows the case of 1tpp in which the enrichment process finds child conformations of the second family to be the top scoring structures. By having an ensemble of AveEnrich = 4 conformations for each family, we greatly increase the probability of selecting the best family on the basis of a single point energy (no optimization). Indeed, we can confidently choose just the 10% (NoptDF = 5 out of the 50) of the families by energy for subsequent energy minimization. Figure 5 shows the energy and CRMS change of these top five structures after minimization in 1tnk, 1tng, and 3ptb diagonal cases. The result is that MPSim-Dock guarantees the structural diversity of the good ligand conformations and provides a more comprehensive coverage of the conformational search space, while the scoring with a full one-energy calculation (optionally including solvation) of each candidate orientation improves the accuracy of choosing the best orientation.

Prediction of Cocrystal Structures: MPSim-Dock Results for 8 × 8 Trypsin Cases

Table 5 shows the CRMS and rank of the best ligand conformations obtained from MPSim-Dock and from DOCK4.0 for the 8 × 8 = 64 trypsin cases. For DOCK4.0, we report two different results, one with no minimization (rigid docking) and the other with torsion-minimization option on (flexible docking). MPSim-Dock at this point employs no minimization, but enrichment process does give lower energy values for most of the new family head structures, so it would be enlightening to compare all of these results. Each column corresponds to a particular one of the protein receptors with various ligands and each row corresponds to a particular one of the eight ligands for all eight target receptors. The first entry is CRMS of the lowest (best) ranking structure whose CRMS is below 1.0 Å. Below this entry is the ranking of that structure.

Using only DOCK4.0, and generating 2000 structures while ranking 500 without minimizations, we obtained conformations within 1.0 Å of the crystal among 500 ranked structures for 36 out of 64 cases (56%). The entry "none" indicates that there is no structure with CRMS below 1.0 Å under rank 500.

If torsion-minimization is included, DOCK4.0 performs significantly better giving structures below 1 Å among 500 ranked structures in 59 out of 64 cases (92%). However, some of the structures that were found quite high in the ranking by no-minimization DOCK4.0 were ranked lower by DOCK4.0 with minimizations. Moreover, for only two more cases the top ranked structures had CRMS value smaller than 1.0 Å (four for no minimization and six for minimization).

Using MPSim-Dock we obtained predicted cocrystal structures within 1 Å for 100% of the 64 cases, as shown in the second column under each protein name in Table 5. Indeed, for 59 out of

the 64 cases the top ranked structure is within 1.0 Å, while for two cases it was second, for two other cases it was third, and for one case it was eighth. This makes it clear that MC searching to saturate the search site using diversity restrictions combined with MC saturation of the best families in the enrichment step dramatically improves the docking procedure by automatically ensuring a sufficient sampling of the site. This guarantees comprehensive conformational search while minimizing of the number of structures retained for scoring.

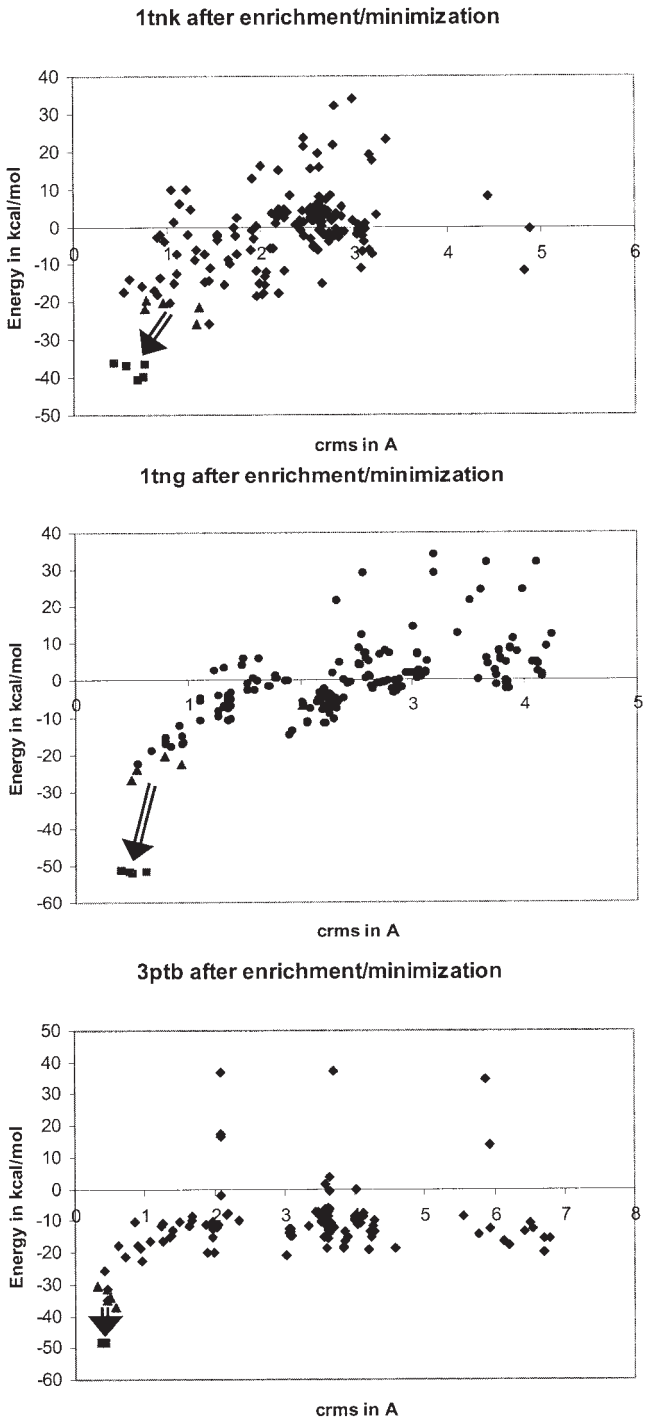


Figure 5. Energy vs. CRMS after enrichment and minimizations. The triangular dots are the top five structures after enrichment, which are minimized to give the square dots. (a) 1tnk (diagonal). (b) 1tng (diagonal). (c) 3ptb (diagonal).

Table 4. Before and After the Enrichment of 1tpp–1tpp (Steps d and e).

Number	Energy	CRMS value
1	−47.09	2.03
2	−46.70	1.05
3	−45.70	4.61
4	−43.92	5.61
5	−43.91	6.38
6	−43.60	5.35
7	−43.42	6.67
8	−43.37	6.43
9	−42.62	6.32
10	−42.46	6.18
11	−42.42	1.41
12	−42.40	6.48
13	−41.56	5.94
14	−41.15	6.54
15	−40.83	6.61
16	−40.09	6.59
17	−39.81	7.33
18	−39.74	6.46
19	−39.19	5.89
20	−39.05	1.35
21	−38.28	1.28
22	−37.20	6.67
23	−36.74	6.93
24	−36.64	6.05
25	−36.01	7.04
26	−35.53	1.01
27	−35.50	7.16
28	−35.45	1.61
29	−35.01	6.43
30	−34.96	7.84
31	−34.96	1.07
32	−34.93	6.62
33	−34.56	1.22
34	−34.04	6.26
35	−34.03	7.13
36	−33.82	6.73
37	−33.80	6.12
38	−33.55	0.96
39	−33.32	5.76
40	−33.31	6.66
41	−33.30	6.09
42	−32.84	1.65
43	−32.74	9.44
44	−32.57	6.20
45	−32.08	6.76
46	−31.67	4.80
47	−31.29	6.90
48	−31.02	8.92
49	−30.74	1.36
50	−30.64	5.33

Table 4. (*Continued*)

Family	Children	CRMS value	Energy	Family	Children	CRMS value	Energy
2	3	0.66	−59.91	31	12	1.09	−35.01
2	2	0.84	−52.01	35	1	6.95	−34.52
1	1	2.03	−47.09	21	7	1.20	−34.28
3	4	4.61	−45.70	45	2	6.85	−34.24
5	1	6.44	−45.65	31	4	0.96	−33.94
3	3	4.66	−45.05	31	10	1.00	−33.90
2	4	0.97	−44.98	5	8	6.38	−33.89
5	4	6.41	−44.25	40	1	6.64	−33.89
5	5	6.40	−44.18	36	3	6.93	−33.71
5	7	6.37	−43.87	47	3	6.94	−33.28
16	1	6.59	−43.73	37	1	6.05	−33.25
6	1	5.40	−43.64	43	1	9.44	−33.02
20	10	1.17	−43.36	22	6	6.65	−32.98
5	2	6.45	−43.31	36	2	6.65	−32.96
20	13	1.18	−42.81	36	1	7.12	−32.82
7	1	6.63	−42.78	25	1	7.14	−32.79
20	4	1.19	−42.35	47	7	6.95	−32.71
21	14	1.02	−41.59	40	3	6.91	−32.37
20	15	1.21	−41.38	40	2	6.79	−32.27
14	1	6.25	−41.21	13	2	6.12	−32.18
38	5	1.06	−41.12	48	2	8.93	−32.13
13	7	5.95	−40.98	47	4	6.98	−32.02
13	4	5.97	−40.73	15	1	6.72	−31.89
20	9	1.22	−40.69	15	3	6.72	−31.61
5	6	6.43	−40.54	15	2	6.67	−31.60
28	2	1.61	−39.94	7	2	6.53	−31.53
31	1	0.79	−39.88	22	3	6.74	−31.51
7	3	6.66	−39.28	47	8	6.93	−31.48
31	8	1.04	−39.21	13	3	6.20	−31.34
26	6	0.97	−39.14	48	1	8.79	−31.09
1	3	2.01	−39.08	45	1	6.75	−31.07
18	3	6.48	−38.97	46	1	4.89	−31.03
33	1	0.99	−38.61	21	2	1.35	−30.98
22	1	6.62	−38.60	47	6	7.06	−30.77
8	2	6.44	−38.39	34	1	6.40	−30.73
18	1	6.53	−38.39	38	4	1.10	−30.72
9	1	6.60	−38.33	24	1	6.01	−30.63
17	1	7.06	−38.00	5	9	6.36	−30.54
5	3	6.52	−37.80	24	2	6.07	−30.35
1	4	2.01	−37.62	7	4	6.55	−30.07
28	3	1.45	−37.61	31	15	1.19	−30.01
22	5	6.67	−37.18	47	9	7.00	−29.99
2	6	1.16	−37.01	31	5	1.21	−29.76
5	1	6.54	−36.64	47	5	6.98	−29.67
33	2	1.48	−36.64	20	5	1.35	−29.59
21	13	1.09	−36.56	22	4	6.75	−29.22
22	2	6.55	−36.54	21	18	1.30	−29.06
9	2	6.62	−36.34	18	6	6.43	−28.96
20	2	1.26	−36.07	13	5	6.12	−28.89
9	3	6.46	−35.94	15	5	6.61	−28.49
22	7	6.59	−35.69	4	1	5.67	−28.45
9	4	6.55	−35.68	13	6	6.20	−28.28
15	4	6.62	−35.61	31	9	1.01	−28.18
18	4	6.51	−35.59	21	17	1.42	−28.14
47	1	6.98	−35.53	29	6	6.41	−27.96
28	1	1.62	−35.30	3	2	4.77	−27.42
27	1	6.82	−35.07	29	4	6.48	−27.15

(*continued*)

Table 4 (*Continued*)

Family	Children	CRMS value	Energy	Family	Children	CRMS value	Energy
29	2	6.48	−26.85	20	6	1.27	10.43
27	2	6.96	−26.48	20	16	1.26	15.57
42	2	1.64	−26.46	26	7	0.97	17.13
29	1	6.61	−25.44	21	5	1.60	18.69
42	10	1.45	−25.43	21	8	1.22	24.89
29	5	6.46	−25.16	18	5	6.55	24.91
1	1	1.94	−25.08	1	5	2.01	32.61
29	3	6.51	−24.57	21	11	1.15	40.95
21	6	1.38	−24.37	21	9	1.11	43.91
13	1	6.21	−24.23	20	3	1.37	46.73
20	17	1.20	−23.77	18	2	6.63	60.56
8	5	6.44	−22.96	21	15	1.22	71.24
3	1	4.64	−22.62	21	12	1.11	72.71
21	16	1.46	−22.45	38	3	0.96	79.96
20	21	1.36	−21.74	20	12	1.23	83.86
47	2	6.93	−21.74	44	2	6.20	85.24
20	20	1.33	−21.43	20	8	1.16	91.87
50	2	5.25	−20.53	42	9	1.31	93.50
42	3	1.64	−20.20	47	11	6.92	94.13
42	7	1.53	−18.42	20	11	1.43	158.80
42	5	1.73	−18.35	42	1	1.47	177.22
31	6	1.03	−18.09	42	6	1.42	180.11
50	1	5.49	−17.79	1	10	2.27	181.12
31	13	1.04	−17.58	44	1	5.99	197.06
1	8	1.99	−16.67	20	1	1.37	215.50
1	9	2.08	−16.26	31	14	0.97	249.76
31	11	1.04	−14.74	1	12	2.05	252.23
2	5	1.07	−14.42	44	3	6.20	302.23
31	2	1.35	−13.37	26	5	1.25	335.93
47	10	6.93	−13.27	50	3	5.22	391.70
1	2	2.05	−13.14	1	11	2.21	392.48
21	3	1.40	−12.34	42	4	1.62	396.68
20	14	1.21	−11.75	44	4	6.29	412.89
31	7	1.01	−11.13	42	8	1.41	459.04
21	4	1.54	−8.45	31	3	0.86	574.10
1	7	2.02	−5.80	26	2	1.28	637.43
20	19	1.26	−4.55	26	3	1.25	1034.78
47	12	6.94	−0.89	26	1	1.25	1091.50
21	10	1.60	−0.53	8	3	6.43	1356.57
20	18	1.25	0.47	26	4	1.32	2946.61
7	5	6.83	5.34	38	2	1.10	3526.30
20	7	1.29	5.40	8	1	6.54	3600.43
1	6	1.96	5.42	38	1	1.31	12770092.00

Comparison of computational times between MPSim-Dock and DOCK4.0 in Table 6 shows that MPSim-Dock takes about thrice as long as DOCK4.0 for a given system. One reason is that the conformations are still generated by DOCK4.0 before they are checked for diversity or enrichment. Because the conformational generation is not targeted towards diversity or enrichment, we do not have control over the time it takes to saturate the conformational space for a given system. For example in Table 6, 1tng ligand and 1tnj receptor entry is 949 while 1tng ligand and 3ptb receptor entry is 324, showing a difference of almost three times. This is because MPSim-Dock will continue to sample the conformation space, which can be different from receptor to receptor, until it is saturated for a given rejection ratio. The second reason is that MPSim-Dock evaluates all-atom FF energy for each ligand conformation, whereas in DOCK4.0, this task is done with pregenerated energy grids. It is conceivable to implement the same method for MPSim-Dock by generating energy grids with MPSim function, which will speed up MPSim-Dock calculations. In many cases, just like 1tpp–1tpp case illustrated in Table 4, enrichment brings a low ranked family to the top allowing the best families to be recognized with single-point energies. This dramatically improves performance. For example, Table 8 shows the top 10 families and 10 children for docking 3ptb ligand on 1tni receptor. The right side shows the parents, which are representatives of

Table 5. Results After the Enrichment Step e of MPSim-Dock (Denoted MPS-D) and from DOCK4.0 Without Minimizations (DOCK-NM) and With Torsion-Minimizations (DOCK-M) for Each of the Eight Ligands (Rows) and Each of the Eight. MPSim-Dock results are also prior to minimization (step f). Entries shaded in yellow are the cases in which the structure with the best energy score has a CRMS < 1 Å while those in red did not find a configuration within CRMS < 1 Å out of 500 ranked structures.

	1pph			1tng			1tni			1tnj		
Ligand	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D
1pph	0.53	0.67	0.91	0.97	0.92	0.14	None	0.93	0.45	None	0.96	0.21
	22	98	1	109	236	1		408	1		122	1
1tng	0.79	0.58	0.49	0.70	0.80	0.94	0.60	0.86	0.82	0.72	0.65	0.36
	20	1	1	1	1	1	1	6	1	5	2	1
1tni	0.96	0.91	0.38	None	0.87	0.57	None	0.94	0.54	0.74	0.87	0.83
	6	6	1		20	1		2	2	4	1	1
1tnj	None	0.67	0.38	None	0.75	0.28	None	0.60	0.32	0.98	0.62	0.57
		81	1		51	1		118	1	16	12	1
1tnk	0.51	0.63	0.79	0.92	0.83	0.95	0.95	0.86	0.75	0.97	0.90	0.29
	31	2	1	3	71	1	23	36	1	2	42	1
1tnl	0.92	0.87	0.79	None	0.99	0.96	None	0.97	0.40	0.52	0.94	0.45
	500	44	1		9	1		1	1	4	4	1
1tpp	0.68	0.67	0.55	None	0.88	0.50	None	None	0.27	None	None	0.24
	11	494	1		354	1			1			1
3ptb	0.85	0.87	0.80	0.89	0.83	0.36	0.71	0.95	0.50	None	0.89	0.74
	28	124	1	335	14	1	13	53	1		297	3

	1tnk			1tnl			1tpp			3ptb		
Ligand	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D
1pph	0.76	None	0.24	0.97	None	0.59	None	None	0.49	None	0.93	0.65
	40		1	41		1			1		158	1
1tng	0.98	0.77	0.83	0.76	0.89	0.64	0.96	0.74	0.91	0.97	0.70	0.85
	8	2	1	29	4	1	101	1	2	1	12	1
1tni	0.88	0.49	0.35	None	0.88	0.47	None	0.98	0.97	None	0.83	0.50
	1	13	1		8	1		1	1		3	1
1tnj	0.94	0.84	0.97	0.96	0.86	0.52	None	0.96	0.40	None	0.78	0.22
	6	85	1	2	391	1		118	1		68	1
1tnk	0.78	0.92	0.76	0.96	0.98	0.15	None	0.73	0.23	None	0.91	0.98
	44	35	2	11	27	1		31	1		7	1
1tnl	0.99	0.99	0.67	0.90	0.95	0.72	None	0.92	0.79	None	0.93	0.29
	8	2	1	96	13	8		7	1		21	1
1tpp	None	0.91	0.59	None	0.84	0.37	None	0.89	0.83	None	0.88	0.29
		228	1		194	1		291	1		255	1
3ptb	None	0.96	0.89	None	0.88	0.96	0.77	0.98	0.79	0.46	0.81	0.57
		25	1		365	3	227	75	1	1	23	1

families before enrichment. Parent 2 has a low CRMS yet its scoring is second to parent 1, which has high CRMS value. After enrichment, however, the children structures of parent 2 are the three top structures. The energy improvement of the children of family 2 is dramatic. In many other cases among our examples, the first stage gives the top structure to be below 1 Å, and the enrichment only improves the energy value of that top structure.

Discussions of the Results for 8 × 8 Trypsin Cases

Table 7 shows the result of top five structures predicted by MPSim-Dock protocol for the eight trypsin ligands docked to the eight cocrystal structures of the trypsin receptor (denoted as 8×8 trypsin). Using the criterion the one of the structures within 2 kcal/mol of the best calculated energy must be within 1 Å CRMS of the exact answer, MPSim-Dock success is 100%. With the criterion that the lowest energy structure must be within 1 Å CRMS of the exact answer, MPSim-Dock succeeds in 57 out of 64 cases (91%).

Although including AVGB solvation was necessary to obtain the high correlation in binding constants for the FF validation studies, it acts negatively for the docking studies. Thus, the failure rate of the best energy having a CRMS <1 Å

Table 6. Comparison of the CPU Times in Seconds Taken for the Cases Shown in Table 5.

		1pph			1tng			1tni			1tnj		
Ligands		DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D
1pph	cpu time	1135	3314	4173	1156	3320	4231	1157	3331	4089	1150	3317	3965
	ratio	1.0	2.9	3.7	1.0	2.9	3.7	1.0	2.9	3.5	1.0	2.9	3.4
1tng	cpu time	162	426	613	163	430	706	162	427	547	165	431	949
	ratio	1.0	2.6	3.8	1.0	2.6	4.3	1.0	2.6	3.4	1.0	2.6	5.8
1tni	cpu time	130	339	1251	131	342	1230	130	338	1110	131	345	1520
	ratio	1.0	2.6	9.6	1.0	2.6	9.4	1.0	2.6	8.5	1.0	2.6	11.6
1tnj	cpu time	336	722	760	335	730	892	332	729	818	333	731	1080
	ratio	1.0	2.1	2.3	1.0	2.2	2.7	1.0	2.2	2.5	1.0	2.2	3.2
1tnk	cpu time	739	2011	1134	741	2111	1118	753	2208	1331	745	2189	1533
	ratio	1.0	2.7	1.5	1.0	2.8	1.5	1.0	2.9	1.8	1.0	2.9	2.1
1tnl	cpu time	160	338	920	158	342	1270	157	345	1073	152	362	1533
	ratio	1.0	2.1	5.8	1.0	2.2	8.0	1.0	2.2	6.8	1.0	2.4	10.1
1tpp	cpu time	998	2318	2405	1007	2370	2043	1013	2389	2674	1101	2378	3819
	ratio	1.0	2.3	2.4	1.0	2.4	2.0	1.0	2.4	2.6	1.0	2.2	3.5
3ptb	cpu time	298	512	540	287	523	872	276	514	802	275	498	1092
	ratio	1.0	1.7	1.8	1.0	1.8	3.0	1.0	1.9	2.9	1.0	1.8	4.0

		1tnk			1tnl			1tpp			3ptb		
Ligands		DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D	DOCK-NM	DOCK-M	MPS-D
1pph	cpu time	1134	3312	3875	1123	3210	4021	1123	3211	3792	1110	3115	3632
	ratio	1.0	2.9	3.4	1.0	2.9	3.6	1.0	2.9	3.4	1.0	2.8	3.3
1tng	cpu time	165	432	362	162	429	440	162	428	641	159	417	324
	ratio	1.0	2.6	2.2	1.0	2.6	2.7	1.0	2.6	4.0	1.0	2.6	2.0
1tni	cpu time	130	339	961	130	338	1131	130	340	1125	130	332	664
	ratio	1.0	2.6	7.4	1.0	2.6	8.7	1.0	2.6	8.7	1.0	2.6	5.1
1tnj	cpu time	332	730	546	331	731	670	334	740	665	327	719	425
	ratio	1.0	2.2	1.6	1.0	2.2	2.0	1.0	2.2	2.0	1.0	2.2	1.3
1tnk	cpu time	742	2194	633	738	2007	1056	739	2012	1023	737	1997	598
	ratio	1.0	3.0	0.9	1.0	2.7	1.4	1.0	2.7	1.4	1.0	2.7	0.8
1tnl	cpu time	157	341	794	155	339	999	151	321	839	147	313	473
	ratio	1.0	2.2	5.1	1.0	2.2	6.4	1.0	2.1	5.6	1.0	2.1	3.2
1tpp	cpu time	1003	2365	1793	1108	2399	2898	1112	2403	2541	992	2101	1797
	ratio	1.0	2.4	1.8	1.0	2.2	2.6	1.0	2.2	2.3	1.0	2.1	1.8
3ptb	cpu time	294	511	343	298	513	790	295	516	781	283	504	358
	ratio	1.0	1.7	1.2	1.0	1.7	2.7	1.0	1.7	2.6	1.0	1.8	1.3

The second row for each entry shows the ratio of the computational time to that of DOCK without minimization.

increases from 9 to 30% when including AVGB solvation in the scoring. This is probably because we did not include crystal waters so that solvation altered the energy values for the loose space within binding sites. Moreover, solvation sometimes give better energy values to bad structures that have part of the ligand hanging out of the binding site (which gets electrostatic energy contribution from implicit solvation calculations). Solvation calculations are often essential in effectively distinguishing good binders from bad ones among a set of different ligands, but can act against distinguishing good conformations from bad ones for a single ligand.

Choice of the Target Protein Structure

We have eight different choices of target receptor structures that can be used for Virtual Ligand Screening (VLS). The choice of the target protein structure is important for the VLS procedure. Table 9 shows the CRMS differences between the eight target receptor structures for the trypsin cocrystals. The CRMS values have been calculated only for the 23 residues within 5 Å of the ligand in each FF minimized cocrystal structure. Table 9 shows that the structural changes in different conformations of the trypsin are small, and hence should not affect the choice of the target protein structure. However, Table 8 shows that different receptors lead to differences in producing docked structures that are below 1 Å. The target structure 1pph yields the largest number of ligand-docked conformations below 1 Å CRMS. Whereas using 3ptb as the target structure we find only 78% of the ligand conformations are within 1 Å. A possible explanation for this result is that 1pph ligand is the largest in size, and hence, samples all the features of the trypsin surface used by all other inhibitor ligands. In fact, as seen in Table 11, 1pph receptor produces the largest number of spheres. This

Table 7. Top Five Structures for Each of the Eight Ligands to Each of the Eight Protein Structures from MPSim-Dock Ordered by the Energy (E0) After Enrichment (step e) But Including the Results After Minimization (step f), for 10 Steps E10 min or Full Minimization (50 Step Max) Emin.

A. 1pph receptor 1pph–1pph Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
2	1873	−6.03	0.43	−21.76	−69.90	0.52	−3288.17
3	576	−5.56	0.76	−21.64	−70.78	0.57	−3288.09
1	2038	−4.52	0.97	−26.55	−70.86	0.62	−3288.14
3	576	−3.52	0.80	−20.40	−70.70	0.56	−3288.05
3	576	−3.19	0.60	−19.58	−71.27	0.53	−3288.44
1tng–1pph							
1	587	−24.12	0.77	−34.27	−38.69	0.61	−3335.81
3	276	−23.56	1.34	−32.97	−39.06	1.08	−3335.47
2	87	−23.16	0.94	−31.31	−38.66	0.61	−3335.77
2	87	−22.39	0.98	−34.75	−38.59	0.83	−3335.76
3	276	−19.32	1.58	−32.39	−40.67	1.47	−3336.43
1tni–1pph							
2	847	−8.24	0.54	−14.28	−38.30	0.46	−3334.08
1	542	−7.87	0.45	−15.14	−39.91	0.43	−3334.28
1	542	−7.18	0.97	−9.81	−35.77	0.78	−3332.03
3	1002	−7.07	1.02	−17.32	−36.92	0.96	−3325.72
3	1002	−7.03	1.23	−18.15	−39.71	0.85	−3334.24
1tnj–1pph							
1	973	−18.00	1.68	−24.25	−37.56	1.09	−3328.15
1	973	−17.04	1.45	−23.00	−37.47	0.85	−3334.82
1	973	−16.84	0.78	−18.20	−36.18	0.54	−3326.15
1	973	−16.55	1.51	−21.05	−37.49	0.94	−3334.87
1	973	−15.94	0.71	−19.03	−36.43	0.45	−3326.62
1tnk–1pph							
1	912	−17.00	0.33	−36.16	−24.18	0.22	−3326.10
3	542	−16.98	1.24	−34.19	−24.16	0.53	−3326.19
3	542	−16.88	1.57	−35.46	−24.17	0.43	−3326.13
1	912	−16.83	0.57	−36.10	−24.14	0.32	−3326.14
3	542	−16.83	1.91	−34.75	−24.19	0.15	−3326.13
1tnl–1pph							
1	587	107.48	0.37	101.95	92.71	0.24	−3335.15
3	237	113.56	0.86	101.49	92.42	0.80	−3335.80
3	237	113.88	1.99	108.88	97.96	1.80	−3335.84
2	78	114.16	1.33	107.47	98.74	1.30	−3338.82
2	78	114.41	1.50	111.24	96.75	1.42	−3326.36
1tpp–1pph							
1	205	−48.93	0.56	−60.77	−87.80	0.45	−3312.30
1	205	−48.71	0.68	−78.38	−86.90	0.44	−3312.12
2	903	−38.89	0.96	−62.82	−87.17	0.45	−3311.64
2	903	−38.24	0.99	−77.68	−94.30	0.84	−3297.63
2	903	−38.01	0.96	−63.51	−87.16	0.45	−3311.67
3ptb–1pph							
1	321	−40.01	0.67	−47.27	−57.30	0.25	−3279.63
2	203	−29.93	1.64	−35.67	−47.18	1.23	−3283.86
2	203	−29.87	1.97	−34.16	−47.78	1.93	−3282.42
2	203	−27.90	1.74	−31.33	−47.63	1.66	−3282.77
1	321	−27.44	0.95	−32.53	−47.64	0.87	−3282.78

(*continued*)

Table 7. (*Continued*)

B. 1tng receptor 1pph–1tng Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
2	372	−26.74	0.66	−26.21	−47.07	0.56	−3240.69
2	372	−22.55	1.18	−13.86	−46.91	1.15	−3240.42
4	188	−21.55	1.02	−9.35	−46.51	0.98	−3240.60
9	771	−20.50	1.09	−7.22	−46.66	1.01	−3240.49
3	162	−20.49	1.13	−8.92	−46.18	1.05	−3240.55
1tng–1tng							
1	772	−26.74	0.49	−34.51	−51.49	0.45	−3275.92
1	772	−24.16	0.55	−31.99	−51.89	0.53	−3276.06
1	772	−22.55	0.94	−30.24	−51.68	0.74	−3275.85
3	245	−21.55	1.99	−27.54	−34.67	1.95	−3282.67
1	772	−20.50	0.79	−30.34	−52.11	0.78	−3276.03
1tni–1tng							
1	1083	−8.01	0.87	−31.11	−42.02	0.78	−3273.73
2	1428	−7.48	1.54	−28.93	−41.88	1.33	−3274.80
2	1428	−7.41	1.51	−28.47	−41.96	1.34	−3274.59
9	876	−7.23	1.66	−14.16	−47.58	0.56	−3274.94
9	876	−5.95	1.72	−14.30	−47.41	0.66	−3274.81
1tnj–1tng							
1	625	−31.92	0.92	−36.26	−41.90	0.91	−3226.27
1	625	−30.32	1.12	−37.07	−41.91	1.03	−3231.87
1	625	−28.98	0.90	−35.63	−41.90	0.89	−3229.61
11	398	−24.83	0.73	−31.38	−42.67	0.71	−3227.15
11	398	−23.71	0.93	−29.96	−44.15	0.65	−3235.19
1tnk–1tng							
1	384	−28.72	0.83	−36.16	−40.32	0.82	−3278.88
1	384	−28.21	1.05	−34.19	−39.69	1.03	−3278.90
2	297	−26.61	0.83	−35.46	−38.73	0.74	−3276.91
2	297	−26.40	0.81	−36.10	−40.42	0.62	−3279.29
2	297	−26.04	0.65	−34.75	−38.88	0.64	−3276.92
1tnl–1tng							
1	89	99.84	0.64	93.32	89.29	0.61	−3275.98
1	89	100.05	0.62	93.64	89.39	0.77	−3275.97
2	892	104.03	1.56	97.82	82.87	0.32	−3276.53
10	1026	105.22	3.13	93.70	83.41	3.14	−3274.59
15	721	106.47	1.30	102.38	98.94	1.30	−3275.47
1tpp–1tng							
1	673	−31.90	5.02	−50.33	−55.99	5.15	−3252.23
5	932	−31.25	0.91	−47.00	−82.32	0.73	−3249.24
5	932	−30.34	1.30	−41.28	−82.30	0.74	−3249.66
2	1033	−29.29	0.65	−48.76	−82.62	0.74	−3248.88
5	932	−29.14	1.05	−42.08	−65.19	1.06	−3249.00
3ptb–1tng							
47	388	−21.32	1.91	−29.17	−43.43	1.79	−3238.63
12	271	−20.67	1.23	−25.95	−37.61	1.34	−3238.32
1	591	−20.53	0.78	−26.44	−45.54	0.69	−3232.23
1	591	−20.51	0.92	−30.56	−42.37	0.93	−3239.45
12	271	−20.38	1.32	−29.95	−43.63	1.34	−3239.52

(*continued*)

Table 7. (*Continued*)

C. 1tni receptor 1pph–1tni Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
1	292	6.30	1.02	−13.55	−49.68	0.96	−3191.41
1	292	6.39	1.84	−15.70	−36.78	1.60	−3191.56
1	292	6.82	1.05	−12.42	−49.63	0.97	−3191.41
2	1101	10.62	1.03	−8.21	−50.56	0.94	−3191.04
3	521	14.42	0.75	−15.11	−50.04	0.85	−3191.34
1tng–1tni							
1	24	−35.95	0.51	−42.95	−46.26	0.49	−3218.83
1	24	−34.18	0.69	−43.07	−46.39	0.65	−3219.45
1	24	−30.78	0.57	−38.33	−46.55	0.52	−3219.46
2	399	−26.49	1.96	−33.76	−42.36	1.93	−3225.69
3	832	−24.83	0.91	−35.66	−46.34	0.80	−3219.94
1tni–1tni							
8	74	−18.03	1.87	−33.82	−45.06	0.65	−3228.94
2	182	−16.21	1.60	−22.39	−43.82	0.98	−3223.32
2	182	−15.69	1.52	−22.60	−43.65	0.85	−3223.34
1	304	−15.36	0.54	−22.39	−44.81	0.52	−3222.88
2	182	−15.18	1.80	−20.80	−43.53	1.09	−3223.34
1tnj–1tni							
3	934	−23.14	1.73	−28.72	−41.51	0.89	−3226.27
3	934	−23.06	1.67	−25.33	−40.19	1.61	−3231.87
1	288	−22.10	1.23	−26.58	−38.71	1.34	−3229.61
7	504	−21.20	0.72	−27.66	−42.27	0.66	−3227.15
7	504	−20.59	0.71	−25.60	−33.29	0.97	−3235.19
1tnk–1tni							
1	89	−23.73	0.94	−28.94	−35.57	0.87	−3221.87
2	192	−23.71	1.56	−26.61	−32.13	1.34	−3225.91
2	192	−20.04	1.57	−25.19	−32.06	1.32	−3225.92
2	192	−19.99	1.54	−26.31	−33.73	1.43	−3226.08
2	192	−19.69	1.11	−24.18	−45.11	0.67	−3219.22
1tnl–1tni							
2	828	93.65	1.14	90.70	85.76	0.98	−3227.42
2	828	93.93	1.08	90.91	85.82	0.87	−3227.42
1	392	94.05	0.98	90.48	85.82	0.76	−3227.94
1	392	94.11	0.95	90.47	85.81	0.67	−3227.41
2	828	95.47	1.02	92.00	85.75	0.85	−3227.46
1tpp–1tni							
1	113	−44.70	0.97	−57.97	−65.12	1.11	−3214.77
1	113	−40.40	0.90	−58.44	−64.96	1.05	−3213.91
1	113	−40.23	0.84	−54.82	−75.77	0.62	−3206.08
2	923	−39.41	1.12	−54.54	−65.05	1.10	−3214.74
1	113	−36.50	0.95	−51.49	−65.19	1.13	−3215.13
3ptb–1tni							
1	78	−31.03	2.02	−38.39	−44.52	0.98	−3185.62
2	553	−28.03	0.75	−38.87	−43.46	0.59	−3186.03
2	553	−27.53	0.71	−36.65	−43.99	0.60	−3185.82
1	78	−26.13	2.03	−39.60	−44.44	0.88	−3185.63
3	253	−25.89	4.33	−34.19	−41.81	4.24	−3188.85

(*continued*)

Table 7. (*Continued*)

D. 1tnj receptor 1pph–1tnj Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
1	915	22.05	0.92	5.44	−37.60	0.71	−3198.38
1	915	22.29	0.97	1.12	−38.43	0.69	−3198.36
1	915	23.94	0.99	1.59	−38.75	0.69	−3198.34
2	497	25.90	3.43	8.43	−32.27	3.03	−3196.30
5	57	26.21	1.17	−0.22	−30.56	0.78	−3198.88
1tng–1tnj							
1	384	−23.78	0.32	−34.15	−35.46	0.27	−3239.56
1	384	−23.51	0.28	−33.20	−35.41	0.14	−3239.45
4	283	−23.14	1.87	−30.01	−32.99	0.89	−3236.04
4	283	−23.02	1.59	−30.28	−36.25	0.94	−3238.19
1	384	−21.08	0.30	−35.43	−35.49	0.87	−3238.81
1tni–1tnj							
1	527	−17.09	1.12	−26.90	−36.83	0.98	−3235.74
2	187	−16.41	0.74	−27.09	−38.21	0.69	−3237.62
5	276	−14.81	3.05	−31.40	−37.91	3.00	−3238.32
2	187	−14.04	0.79	−25.36	−38.59	0.78	−3237.91
3	1092	−13.66	1.99	−20.60	−32.16	1.89	−3236.75
1tnj–1tnj							
1	822	−28.26	0.93	−35.54	−37.70	0.93	−3238.42
1	822	−25.64	0.88	−30.59	−34.87	0.82	−3239.05
1	822	−25.27	0.94	−34.10	−37.75	0.95	−3238.34
1	822	−24.34	0.75	−32.74	−36.33	0.78	−3238.99
2	932	−22.64	1.16	−26.97	−39.22	1.16	−3223.75
1tnk–1tnj							
1	287	−21.31	0.97	−28.05	−34.28	0.83	−3237.74
4	173	−18.56	1.78	−28.05	−33.22	1.40	−3236.74
1	287	−17.96	0.82	−25.71	−33.86	0.66	−3237.66
5	98	−17.04	1.51	−23.63	−30.64	1.31	−3230.75
2	1021	−16.84	1.37	−26.51	−34.54	1.70	−3233.76
1tnl–1tnj							
1	113	102.54	0.65	93.14	87.83	0.52	−3233.67
3	83	103.13	2.21	93.85	91.53	2.09	−3236.56
11	102	103.24	1.02	97.52	86.25	0.78	−3224.35
11	102	104.61	1.11	90.64	87.80	0.88	−3233.37
1	113	104.94	0.56	91.34	87.55	0.46	−3233.10
1tpp–1tnj							
1	293	−45.33	0.59	−59.75	−78.78	0.48	−3203.09
1	293	−42.88	0.57	−67.38	−84.16	0.67	−3202.41
30	821	−38.49	5.40	−50.47	−62.98	4.76	−3206.93
30	821	−37.84	5.57	−48.74	−55.75	6.16	−3209.98
30	821	−37.61	5.43	−49.41	−56.16	5.06	−3209.81
3ptb–1tnj							
1	532	−32.68	0.23	−39.23	−44.98	0.22	−3196.24
1	532	−31.76	0.29	−38.54	−44.91	0.25	−3196.25
2	187	−28.29	1.95	−35.05	−48.32	0.85	−3190.94
2	187	−28.00	1.97	−32.68	−47.91	0.87	−3190.98
3	273	−25.70	3.18	−30.99	−37.98	3.01	−3193.09

(*continued*)

Table 7. (*Continued*)

E. 1tnk receptor 1pph–1tnk Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
2	598	6.04	0.79	−12.22	−53.00	0.84	−3186.94
3	288	8.24	0.94	−7.37	−50.40	0.85	−3186.42
1	907	8.90	0.78	−15.26	−51.35	0.63	−3186.89
2	598	10.68	0.76	−7.81	−53.15	0.86	−3186.92
3	288	10.90	0.79	−6.23	−52.98	0.84	−3186.34
1tng–1tnk							
1	234	−30.11	0.95	−36.94	−42.80	0.78	−3218.20
1	234	−27.52	1.03	−32.05	−42.66	0.81	−3222.25
14	82	−26.06	1.92	−31.60	−33.47	1.80	−3218.24
1	234	−21.37	1.19	−31.18	−42.67	0.82	−3222.27
14	82	−20.78	1.94	−24.27	−33.43	1.81	−3227.25
1tni–1tnk							
1	1033	−21.20	0.75	−30.39	−39.48	0.73	−3229.29
23	233	−18.93	1.08	−25.49	−43.96	1.15	−3216.26
23	233	−16.96	1.05	−22.91	−41.13	1.11	−3223.67
1	1033	−16.06	0.87	−20.43	−40.82	0.90	−3223.93
1	1033	−15.75	0.78	−19.88	−43.57	0.75	−3215.79
1tnj–1tnk							
1	788	−21.44	0.29	−24.20	−35.49	0.23	−3219.68
2	362	−21.38	2.66	−25.80	−34.82	2.63	−3224.92
17	57	−20.98	1.77	−24.66	−43.22	0.76	−3223.35
17	57	−20.42	1.75	−23.32	−31.88	1.38	−3219.60
3	221	−20.02	2.76	−23.43	−32.02	1.42	−3218.86
1tnk–1tnk							
2	237	−25.94	1.45	−34.54	−39.98	0.98	−3215.21
1	787	−21.88	0.76	−26.57	−37.15	0.73	−3229.12
2	237	−21.27	1.42	−27.89	−40.52	0.89	−3215.39
1	787	−20.22	1.03	−25.76	−36.55	0.99	−3228.45
1	787	−19.37	0.87	−25.00	−36.08	0.64	−3228.88
1tnl–1tnk							
1	362	103.18	0.83	100.09	90.58	0.81	−3223.56
1	362	104.11	0.87	99.52	90.57	0.86	−3223.71
1	362	104.26	0.87	101.91	91.45	0.87	−3220.44
3	884	104.41	1.86	96.69	86.82	0.97	−3233.83
1	362	107.19	0.85	102.36	91.25	0.91	−3220.31
1tpp–1tnk							
2	824	−52.30	0.92	−62.08	−87.30	0.71	−3199.47
1	1023	−48.34	0.83	−55.92	−87.26	0.72	−3199.57
1	1023	−46.34	0.80	−59.77	−87.33	0.66	−3198.88
1	1023	−45.97	0.80	−59.75	−87.29	0.67	−3198.90
5	342	−41.08	6.56	−51.30	−61.35	6.35	−3201.83
3ptb–1tnk							
7	329	−26.00	0.98	−31.56	−39.23	0.95	−3184.59
7	329	−25.99	0.97	−31.72	−39.22	0.94	−3184.59
1	118	−23.96	0.35	−27.86	−41.36	0.33	−3184.56
1	118	−22.23	0.51	−26.52	−42.41	0.43	−3184.60
1	118	−20.78	0.53	−25.90	−42.40	0.45	−3184.60

(*continued*)

Table 7. (*Continued*)

F. 1tnl receptor 1pph–1tnl Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
2	284	6.04	0.79	1.38	−34.82	0.72	−3255.20
3	455	8.24	0.94	2.34	−41.98	0.88	−3252.16
1	473	8.90	0.78	3.47	−31.70	0.76	−3253.98
2	284	10.68	0.76	4.56	−30.13	0.75	−3254.12
3	455	10.90	0.79	4.92	−31.42	0.69	−3254.37
1tng–1tnl							
2	325	−16.30	0.97	−28.21	−32.02	0.96	−3295.85
2	325	−15.61	1.21	−28.02	−32.81	0.89	−3299.12
1	891	−14.80	0.26	−26.83	−30.71	0.24	−3283.69
2	325	−14.71	1.18	−27.97	−32.81	0.86	−3299.07
2	325	−14.50	0.91	−23.00	−32.10	0.90	−3296.08
1tni–1tnl							
1	722	−16.57	0.32	−28.16	−41.26	0.31	−3282.51
1	722	−13.82	0.40	−28.23	−41.51	0.38	−3282.63
38	321	−13.15	1.76	−31.68	−36.11	1.78	−3297.30
38	321	−12.99	1.75	−27.00	−41.52	1.67	−3282.75
1	722	−11.89	0.62	−22.77	−41.71	0.60	−3282.68
1tnj–1tnl							
1	83	−19.88	0.45	−26.02	−34.05	0.43	−3298.58
5	193	−18.71	1.19	−23.41	−24.47	1.21	−3300.48
4	162	−16.66	1.04	−22.54	−30.81	0.88	−3283.74
27	563	−16.53	1.12	−26.55	−31.54	1.01	−3301.81
27	563	−16.45	1.15	−25.75	−31.50	0.97	−3284.04
1tnk–1tnl							
7	172	−22.08	0.67	−29.28	−34.48	0.71	−3296.78
4	266	−19.24	1.78	−30.68	−39.38	0.99	−3282.92
4	266	−18.37	1.88	−25.38	−32.31	1.92	−3296.54
7	172	−16.92	0.80	−25.53	−34.29	0.79	−3296.76
8	81	−16.91	2.33	−27.60	−38.14	1.09	−3282.44
1tnl–1tnl							
1	823	105.57	0.72	96.61	92.69	0.65	−3292.75
2	123	107.04	2.55	102.70	98.03	2.56	−3297.44
3	321	108.76	1.88	101.60	94.15	1.56	−3299.46
2	123	109.34	2.34	98.01	93.99	2.12	−3300.06
5	933	110.24	1.29	104.91	92.46	0.91	−3292.46
1tpp–1tnl							
1	663	−40.26	0.79	−64.17	−89.62	0.69	−3260.50
1	663	−39.59	0.81	−62.68	−89.14	0.66	−3259.70
1	663	−38.99	0.84	−62.27	−89.05	0.67	−3259.41
1	663	−38.79	0.85	−61.45	−89.59	0.70	−3260.50
1	663	−37.71	0.89	−60.39	−89.05	0.67	−3259.40
3ptb–1tnl							
1	521	−24.96	0.46	−32.51	−36.10	0.43	−3252.83
1	521	−24.52	0.29	−31.11	−36.09	0.21	−3253.04
1	521	−24.37	0.23	−31.52	−36.07	0.22	−3252.83
2	188	−23.17	3.37	−28.70	−34.76	3.33	−3249.96
17	293	−22.53	5.54	−30.11	−33.02	5.87	−3253.69

(*continued*)

Table 7. (*Continued*)

G. 1tpp receptor 1pph–1tpp Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
1	732	8.01	0.59	−10.89	−52.25	0.45	−3347.83
1	732	8.91	0.60	−11.31	−52.61	0.42	−3347.80
1	732	10.43	0.62	−10.06	−52.44	0.55	−3347.83
3	188	10.50	4.95	−10.90	−51.69	4.88	−3347.78
1	732	10.80	0.63	−11.59	−52.71	0.59	−3347.79
1tng–1tpp							
2	68	−12.08	1.09	−19.30	−22.40	0.97	−3401.22
27	107	−10.55	1.65	−15.58	−19.67	1.58	−3399.58
27	107	−10.51	2.18	−15.24	−19.68	2.20	−3399.52
7	435	−10.42	0.56	−21.43	−36.21	0.45	−3389.56
2	68	−10.41	0.78	−15.86	−27.84	0.76	−3382.44
1tni–1tpp							
1	921	−16.56	0.27	−25.54	−39.64	0.25	−3387.82
1	921	−16.31	0.28	−25.22	−39.74	0.26	−3387.79
1	921	−15.69	0.56	−23.50	−39.69	0.27	−3387.54
2	113	−13.48	1.84	−20.53	−39.43	1.06	−3387.50
2	113	−12.86	1.53	−18.32	−28.37	1.56	−3393.44
1tnj–1tpp							
1	554	−17.55	1.97	−26.02	−35.92	1.50	−3389.61
1	554	−15.35	2.10	−23.41	−34.75	1.48	−3389.58
4	132	−14.71	1.86	−22.54	−29.94	1.89	−3383.94
8	288	−14.60	0.73	−26.55	−31.92	0.72	−3383.39
4	132	−14.53	1.83	−25.75	−36.30	0.98	−3392.11
1tnk–1tpp							
18	172	−15.92	0.59	−20.24	−25.99	0.87	−3389.31
38	92	−12.83	0.91	−19.07	−32.58	0.53	−3390.74
1	823	−12.59	1.65	−22.85	−29.80	1.62	−3383.97
2	324	−12.35	1.94	−16.82	−34.46	1.82	−3391.48
38	92	−11.80	0.95	−18.53	−32.08	0.77	−3389.96
1tnl–1tpp							
3	366	104.52	0.80	96.02	92.00	0.81	−3391.15
1	1123	112.04	1.52	107.85	99.14	1.58	−3383.09
5	42	114.38	1.10	106.83	102.27	1.00	−3390.74
2	283	114.88	2.38	108.85	103.24	2.35	−3397.95
1	1123	116.01	1.67	110.71	99.61	1.47	−3383.03
1tpp–1tpp							
2	372	−59.91	0.66	−69.58	−100.68	0.47	−3363.02
2	372	−52.01	0.84	−64.52	−100.76	0.48	−3363.17
1	192	−47.09	2.03	−61.80	−82.76	2.19	−3363.43
3	573	−45.70	4.61	−59.45	−63.44	4.62	−3358.70
5	98	−45.65	6.44	−53.85	−52.57	6.27	−3367.57
3ptb–1tpp							
1	192	−33.42	1.04	−37.46	−42.06	1.02	−3342.89
1	192	−32.67	2.18	−36.97	−41.00	2.23	−3342.87
2	673	−31.96	0.30	−38.55	−45.00	0.29	−3343.93
2	673	−31.95	0.38	−38.92	−45.82	0.28	−3344.62
1	192	−31.78	1.21	−37.06	−41.66	1.22	−3343.00

(*continued*)

Table 7. (*Continued*)

H. 3tpb receptor 1pph–3ptb Family	Generated E0		crms0	E10step	Emin	crms-min	AVGB energy
1	2994	−9.98	0.80	−29.02	−59.28	0.82	−3133.17
1	2994	−9.66	0.92	−34.80	−59.86	0.86	−3133.28
1	2994	−6.59	1.15	−29.97	−58.75	0.89	−3133.34
1	2994	−5.92	0.99	−26.67	−58.79	0.85	−3133.23
1	2994	−4.71	0.81	−30.99	−59.88	0.83	−3133.21
1tng–3ptb							
5	27	−14.88	0.97	−24.25	−40.10	0.88	−3160.92
5	27	−14.69	1.07	−24.08	−39.46	1.04	−3161.26
2	191	−13.90	1.93	−27.92	−39.00	1.94	−3162.12
2	191	−12.16	1.92	−26.15	−38.99	1.94	−3162.11
2	191	−12.16	1.98	−19.45	−38.78	1.95	−3162.01
1tni–3ptb							
2	773	−16.58	0.50	−26.00	−42.38	0.49	−3157.76
2	773	−15.44	0.57	−23.78	−42.35	0.57	−3157.73
1	1022	−14.92	2.13	−25.55	−42.28	2.13	−3157.88
1	1022	−14.67	2.61	−21.55	−26.38	2.81	−3179.17
2	773	−13.45	0.40	−22.84	−42.17	0.38	−3157.86
1tnj–3ptb							
1	371	−21.28	1.76	−29.50	−35.02	1.75	−3161.49
2	92	−19.82	2.07	−27.83	−39.69	1.97	−3163.11
4	211	−19.22	0.74	−28.08	−39.75	0.54	−3163.05
2	92	−19.19	2.20	−27.90	−39.58	1.98	−3163.12
2	92	−17.62	2.15	−26.53	−40.06	1.02	−3162.88
1tnk–3ptb							
1	117	−22.07	0.96	−28.22	−40.66	0.97	−3160.44
1	117	−21.13	0.98	−27.34	−40.65	0.97	−3160.37
3	756	−12.94	1.56	−21.70	−45.35	0.89	−3162.63
2	85	−8.77	1.97	−23.05	−45.36	0.85	−3162.62
2	85	−5.80	2.12	−10.65	−12.70	2.20	−3194.93
1tnl–3ptb							
15	231	117.00	2.03	108.80	99.24	2.21	−3171.57
1	643	117.49	0.96	110.53	97.47	0.95	−3173.60
2	422	117.96	0.46	111.31	95.60	0.45	−3176.98
3	97	118.03	1.87	107.92	105.28	1.88	−3173.23
15	231	118.89	2.22	112.58	99.60	2.01	−3177.05
1tpp–3ptb							
3	445	−64.74	0.83	−76.40	−94.10	0.64	−3142.12
3	445	−64.15	0.70	−82.17	−94.01	0.62	−3143.57
3	445	−63.13	0.79	−73.66	−94.03	0.63	−3142.27
1	722	−62.94	0.36	−74.45	−93.84	0.65	−3141.94
1	722	−62.51	0.23	−72.78	−93.00	0.63	−3142.26
3ptb–3ptb							
2	135	−37.18	0.57	−42.50	−48.53	0.37	−3120.42
1	203	−34.82	0.45	−41.64	−48.54	0.38	−3120.41
2	135	−33.93	0.49	−39.72	−48.53	0.43	−3120.63
1	203	−31.37	0.46	−37.26	−48.55	0.41	−3120.58
1	203	−30.68	0.31	−38.74	−48.69	0.37	−3120.30

Shaded in each column is the lowest energy or best crms. The column "Family" shows the rank of the clusters and "Generated" shows the generation order of the structure. −82.17 0.64

Table 8. Top Structures After Enrichment (step e) of MPSim-Dock for 3ptb Ligand on 1tni Receptor.

Family	Children	crms	E0	Parent	crms	E0	Rejection
2	3	0.50	−32.98	1	2.02	−18.68	47
2	4	0.61	−28.82	2	0.71	−18.56	0
2	1	0.75	−28.03	3	4.33	−17.66	1
1	1	2.03	−26.13	4	2.94	−17.21	5
5	2	1.19	−25.19	5	1.07	−17.14	12
9	4	4.07	−24.74	6	3.82	−16.94	2
2	2	0.87	−24.41	7	3.56	−16.91	1
9	3	4.02	−23.69	8	3.25	−16.44	7
6	2	3.81	−22.10	9	4.11	−16.25	1
5	4	0.91	−22.00	10	0.73	−16.11	3

"Family" column show the rank of the cluster. "Children" column is the generation order of a structure within the cluster. "Parent" are the structures which are the heads of the clusters, ranked by the energy.

target structure has a larger binding pocket compared to the other seven target structures. This means that there are more ways for MPSim-Dock to find good conformations and that ligands have more room to move around within the binding pocket without overlapping protein atoms. Thus, we consider that the target structure should be chosen on the basis of geometric shape and size of the ligand. By the same token, if there is more than one choice of receptors but their structures are very different, it might be necessary to group the ligands to be screened according to their geometric features and then perform VLS to the appropriate target receptor structures.

Table 10 shows the experimental affinities of the ligands along with our predicted affinities both with and without waters.

Discussion and Summary

The most widely used docking algorithm is DOCK, developed by Kuntz and his coworkers.[18,33] In DOCK4.0, the potential ligand docking regions are represented by a set of overlapping spheres generated using the molecular surface of the target protein. The conformational search method uses either rigid or flexible ligand(s) for the fixed receptor conformation. The scoring in DOCK4.0 includes electrostatics and van der Waals interactions to estimate the binding energy of a docked conformation, but the energy function does not include hydrogen bonding or off-diagonal van der Waals interactions. DOCK4.0 has an efficient mapping of conformational space to be searched, but the conformational search is not guaranteed to be exhaustive. Other docking methods involving evolutionary and heuristic algorithms[34] and internal coordinate modeling[35] have also been used.

In this article we develop a more comprehensive conformational search of the conformation space, while using procedures based on DOCK4.0 for generating orientations of ligands into the mapped potential ligand binding site on the protein (receptor). We also combine this conformational search method with generalized scoring using a flexible all-atom FF including continuum solvation techniques (using the MPSim molecular dynamics software that allows rigid body molecular dynamics and NEIMO torsional MD to further enrich the conformational search.). We refer to this method as "MPSim-Dock."

In developing the MPSim-Dock methods described herein, our eventual target is virtual ligand screening (VLS) of a large data base to select a small subset of candidates that bind strongly to a protein with known structure. Subsequent experiment or theory would then be done on just this small subset to optimize the lead into a useful drug. A second important application is to identify function for an orphan receptor or enzyme. Here there may be a large library of potential endogenous ligands that one would scan to find those that bind strongly to the protein. This application is

Table 9. CRMS Differences Between the Eight Receptor Structures Considering Only the 23 Residues in the Active Site.

	1tng	1tni	1tnj	1tnk	1tnl	1tpp	1pph	3ptb
1tng	0.00	0.68	0.75	0.71	0.73	0.71	0.78	0.76
1tni	0.68	0.00	0.45	0.40	0.46	0.44	0.95	0.97
1tnj	0.75	0.45	0.00	0.24	0.41	0.35	0.97	1.02
1tnk	0.71	0.40	0.24	0.00	0.30	0.27	0.84	0.98
1tnl	0.73	0.46	0.41	0.30	0.00	0.31	0.94	0.99
1tpp	0.71	0.44	0.35	0.27	0.31	0.00	0.93	0.96
1pph	0.78	0.95	0.97	0.84	0.94	0.93	0.00	0.37
3ptb	0.76	0.97	1.02	0.98	0.99	0.96	0.37	0.00

Table 10. Comparison With Experiment (ln Ki) of the Calculated Binding Energies for Various Options.

Cocrystal	ln Ki	No solv./no w.	No solv./all w.	No solv./con. w.	Avgb/no w.	Avgb/con. w.	MD	UHBD
1tng	−2.94	−45	−46	−42	−33	−44	−45	−51
1tni	−1.70	−56	−25	−23	−31	−27	−22	−28
1tnj	−1.96	−52	−30	−29	−30	−35	−18	−30
1tnk	−1.49	−54	−26	−23	−27	−30	−14	−21
1tnl	−1.88	−47	−29	−31	−27	−33	−15	−32
1tpp		−124	−54	−54	−124	−50		−24
1pph	−6.23	−129	−185	−75	−135	−84	−87	−72
3ptb	−4.74	−54	−47	−39	−40	−54	−55	−42

(No solv. = no implicit solvation, no w. = no explicit water, all w. = all explicit water included, con. w. = only conserved water included, Avgb = analytic volume generalized Born method, MD = molecular dynamics, UHBD = University of Houston Brownian dynamics package energy calculations).

becoming of critical importance as the genomics revolution has provided us with vast collections of proteins whose functions are uncertain or completely unknown. In both applications it is essential that *we not falsely eliminate the best binding ligands* (no false negatives). Thus, we require a method that can reliable identify the strongly binding ligands by finding the correct binding site and conformation.

In addition for this article we focused further on the part of the VLS problem, which is to find the best binding site for a particular conformation of the receptor. In general, the 3D structure of the protein that is available may be predicted from theory or from experiment but without the ligand bound. Here we consider just a subset of this problem. By binding only to cocrystals we do not face the problem that binding the ligands to the protein may cause the active site for some systems to undergo significant changes in the backbone or sidechain conformations. Thus, we consider just the case in which a structure is available with a ligand already bound into the active site. However, by considering the protein structures from several different cocrystals, we do consider the possibility that the protein structure is not quite correct for the new ligand being considered. In particular, we do consider cases in which the ligand is known to utilize water in binding to the active site, but without including this water explicitly in our calculations.

As a criterion for properly identifying the strongly binding ligands, we consider that the docking stage should find the binding conformation sufficiently accurately that the ligand–protein structure matches the exact answer to an accuracy of at least 1 Å CRMS error and that the energy for this predicted bound structure be within 2 kcal/mol of the predicted strongest binding conformation. With this criterion we find that MPSim-Dock does this 100% of the time for the eight ligands to the eight sites considered here. Using a similar criterion for DOCK4.0 leads to 31% success.

A tighter criterion for success is that the *best scoring ligand* match the exact answer to an accuracy of at least 1 Å CRMS error. In this case MPSim-Dock is 91% successful for the 64 cases considered here. In comparison using a similar criterion for DOCK4.0 leads to 10% success. A looser criterion is that the *best scoring ligand* matches the exact answer to an accuracy of at least 2 Å CRMS error. In this case MPSim-Dock is 100% successful for the 64 cases considered here and DOCK4.0 leads to 45% success.

Thus, we consider that the elements of conformation completeness and enrichment incorporated into MPSim-Dock are essential ingredients to successful docking applications. This involves two phases:

1. *Completeness:* the first phase is to search the whole conformation space to find *all* distinct structures. We do this by partitioning the possible structures into diversity families in which the head of each family is at least DivFam = 0.6 Å CRMS away from the other family heads. Here we ensure that we have located all the families by requiring that our algorithm (e.g., MC) for generating new structures within the search space (defined by the spheres) proceed until the successive MC steps stop finding new families (we find that a ratio of RedunMC = 2.2 MC attempts per structure generated is sufficient, leading to at least 500 families). From these families use the interaction energy with the protein (including solvation) to select a subset of best families (we find that NEnrich = 50 is sufficient) for subsequent investigation.
2. *Enrichment:* for the second phase, we enrich the families selected from the completeness step by continuing the MC process of generating structures but keeping only these within DivFam of one of the NEnrich families until we have added an average of at least AveEnrich = 4 members of each family. The

Table 11. Number of Spheres for Each Receptor Conformation of Trypsin.

Receptor	1tng	1tni	1tnj	1tnk	1tnl	1tpp	1pph	3ptb
No. of spheres	69	57	68	62	64	62	87	58

idea here is that with an ensemble of ~5 members per family, we can reliably identify from the MC scores the best small number (NoptDF = 5) of families for subsequent minimization of the ligand in the fixed protein.

Fixed Protein Optimization

Using the NoptDF = 5 structures from the Enrichment step we carry out sufficient optimization of the ligand structure in the fixed protein that the best structure can be selected. We find that OptStep = 50 steps of minimization is sufficient.

The best structure selected from this procedure now can be used for binding energy calculations, which might involve molecular dynamics minimization of the full ligand–protein complex to find the global minimum. However, finding true binding energies is beyond the scope of the present article.

The parameters selected for the steps of MPSim-Dock were chosen to be much looser than actually required for the application here, with the idea that they may be appropriate for reliably doing other cases. MPSim-Dock has been used successfully for complementary studies of the binding of sugars to proteins and for ligands to tRNA synthetases as will be published separately.

Summary of the Main Contributions

1. Use of all atom force-field function calculations combined with protein–ligand matching method of DOCK4.0 to improve scoring.
2. Implementation of the clustering concept and use of rejection ratio to ensure the completeness of search of conformational space.
3. Implementation of enrichment to find and select better conformations.
4. Test of the resulting software, MPSim-Dock, on 8 by 8 combinations of trypsin cocrystals.

Acknowledgments

The facilities of the MSC used in the studies have been supported by ARO-DURIP, ONR-DURIP, NSF-MRI, IBM-SUR, and Beckman Institute. The MSC is also supported by NSF, DOE, General Motors, Chevron-Texaco, Seiko-Epson, Asahi Kasei, and Toray.

References

1. Maggio, E. T.; Ramnarayan, K. Trends Biotechnol 2001, 19, 266.
2. Abagyan, R.; Totrov, M. Curr Opin Chem Biol 2001, 5, 375.
3. Kuntz, I. Science 1992, 257, 1078.
4. Lengauer, T.; Rarey, M. Curr Opin Struct Biol 1996, 6, 402.
5. Joseph–McCarthy, D. J Mol Biol 1997, 267, 727.
6. Gane, P. J.; Dean, P. M. Curr Opin Struct Biol 2000, 10, 401.
7. Carlson, H. A.; McCammon, J. A. Mol Pharmacol 2000, 57, 213.
8. Sotriffer, C. A.; Flader, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varga, J. M. Methods 2000, 20, 281.
9. Floriano, W. B.; Vaidehi, N.; Singer, M.; Shepherd, G.; Goddard, W. A., III. Proc Natl Acad Sci USA 2000, 97, 10712.
10. Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Goddard, W. A., III. Proc Natl Acad Sci USA 2002, 99, 12622.
11. Datta, D.; Vaidehi, N.; Floriano, W. B.; Kim, K. S.; Prasadarao, N. V.; Goddard, W. A., III. Proteins Struct Funct Genet 2003, 50, 213.
12. Datta, D.; Vaidehi, N.; Xu, X.; Goddard, W. A., III. Proc Natl Acad Sci USA 2002, 99, 2636.
13. Datta, D.; Vaidehi, N.; Zhang, D.; Goddard, W. A., III. Protein Sci, to appear.
14. Kekenes–Huskey, P. M.; Vaidehi, N.; Floriano, W. B.; Goddard, W. A., III. J Phys Chem 2003, 107, 11549.
15. Wang, P.; Vaidehi, N.; Tirrell, D. A.; Goddard, W. A. J Am Chem Soc 2002, 124, 14442.
16. Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., III. J Med Chem 2004, 47, 56.
17. Patrick, B. G., Jr.; Stouten, P. F. W. J Comput Aid Mol Des 2000, 14, 38.
18. Ewing, T. A.; Kuntz, I. D. J Comput Chem 1997, 18, 1175.
19. Lim, K.-T.; Brunett, S.; Iotov, M.; McClurg, R. B.; Vaidehi, N.; Dasgupta, S.; Taylor, S.; Goddard, W. A., III. J Comput Chem 1997, 18, 501.
20. Casale, E.; Collyer, C.; Ascenzi, P.; Balliano, G.; Milla, P.; Viola, F.; Fasano, M.; Menegatti, E.; Bolognesi, M. Biophys Chem 1995, 54, 75.
21. Kurinov, I. V.; Harrison, R. W. Struct Biol 1994, 1, 735.
22. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. J Phys Chem 1990, 94, 8897.
23. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph–McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stot, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
24. Gasteiger, J.; Marsili, M. Tetrahedron 1980, 36, 3219.
25. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phy 1983, 79, 926.
26. Ding, H.; Karasawa, N.; Goddard, W. A., III. Chem Phys Lett 1992, 196, 6.
27. Zamanakos, G. Ph.D. thesis 2001, Caltech, Pasadena.
28. Canutescu, A.; Shelenkov, A.; Dunbrack, R. L., Jr. Protein Sci 2003, 12, 2001.
29. Vriend, G. J Mol Graph 1990, 8, 52.
30. Ghosh, A.; Rapp, C. S.; Friesner, R. A. J Phys Chem B 1998, 102, 10983.
31. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. J Mol Biol 1982, 161, 269.
32. Connolly, M. L. J Appl Crystallogr 1983, 16, 548.
33. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. J Com Mol Des 2001, 15, 411.
34. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. J Comp Chem 1998, 19, 1639.
35. Abagyan, R.; Totrov, M.; Kuznetsov, D. J Com Chem 1994, 15, 488.