Journal of Computer-Aided Molecular Design, 10 (1996) 123–132
ESCOM

J-CAMD 330

Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal

Daniel A. Gschwend and Irwin D. Kuntz*

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.

Received 4 August 1995
Accepted 20 December 1995

Keywords: Molecular recognition; Configurational sampling; Ligand docking; Structure-based drug design

Summary

Strategies for computational association of molecular components entail a compromise between configurational exploration and accurate evaluation. Following the work of Meng et al. [Proteins, 17 (1993) 266], we investigate issues related to sampling and optimization in molecular docking within the context of the DOCK program. An extensive analysis of diverse sampling conditions for six receptor–ligand complexes has enabled us to evaluate the tractability and utility of on-the-fly force-field score minimization, as well as the method for configurational exploration. We find that the sampling scheme in DOCK is extremely robust in its ability to produce configurations near to those experimentally observed. Furthermore, despite the heavy resource demands of refinement, the incorporation of a rigid-body, grid-based simplex minimizer directly into the docking process results in a docking strategy that is more efficient at retrieving experimentally observed configurations than docking in the absence of optimization. We investigate the capacity for further performance enhancement by implementing a degeneracy checking protocol aimed at circumventing redundant optimizations of geometrically similar orientations. Finally, we present methods that assist in the selection of sampling levels appropriate to desired result quality and available computational resources.

Introduction

Molecular recognition is a problem fundamental to structural biology. The interaction of molecules, be they macromolecules or small ligands, is a prerequisite for nearly all biological events. Specific modulation of these interactions has been the ambition of medicinal chemists for over a century. To gain more rapid access to therapeutic agents, we must not only understand, but be able to predict, the structural details of recognition events. The prediction of the observed orientations of two interacting components is known as the 'docking problem'.

There exist many computational approaches to the docking problem [1,2], but each must accomplish two principal tasks: sampling and evaluation. The task of sampling relates to the exploration of the large number of configurations varying in the relative geometry of the components. The task of evaluation refers to the ranking of each configuration by some metric. These seemingly independent phases of docking are in fact closely linked. Without an accurate evaluation scheme, the native configuration cannot be recognized, even when it has been sampled. Conversely, without adequate sampling, even the most accurate evaluation scheme cannot recognize the native configuration if it has not been generated. The molecular docking problem is further complicated by the thousands of degrees of freedom available to interacting atomic assemblies. Even when constraining the components to only six translational and rotational degrees of freedom, the docking problem is a difficult one because there are still myriads of possible configurations. Heuristics must be invoked to direct sampling and ensure computational tractability.

We previously have reported a descriptor-based rigid-body method (DOCK) to address the molecular docking problem [3–5]. More recently, Meng et al. [6] showed that modest orientational sampling coupled with post-docking refinement is more effective at retrieving known binding

*To whom correspondence should be addressed.

TABLE 1
TEST SYSTEMS

PDB entry	Resolution (Å)	Receptor	Docked ligand	Ligand atoms[a]	Receptor spheres
1gst	2.2	Glutathione S-transferase	Glutathione	20	114
2gbp	1.9	D-Galactose/D-glucose binding protein	β-D-Glucose	12	75
3cpa	2.0	Carboxypeptidase A	Glycyl-L-tyrosine	17	44
3dfr	1.7	*L. casei* dihydrofolate reductase	Methotrexate	33	72
4dfr	1.7	*E. coli* dihydrofolate reductase	2,4-Diamino-6-methylpteridine	13	86
6rsa	2.0	Ribonuclease A	Uridine 3'-phosphate	21	47

[a] Number of nonhydrogen ligand atoms.

modes than intensive sampling alone. The favorable effects of rigid-body minimization as a post-docking tool were clearly evident – steric clashes were resolved, scores were improved significantly, and experimentally observed geometries were reproduced more accurately. Unfortunately, the implementation was impractically slow. In this paper, we describe an enhancement to the minimization method, achieving nearly a 50-fold increase in speed. This accelerated rate now permits incorporation of the refinement directly into the docking process. Every configuration generated can be optimized in the context of the receptor, thus capturing the power of minimization as a post-docking scoring tool in the evaluation phase of docking. We shall also show that on-the-fly minimization improves sampling, further supporting the close relationship between sampling and scoring.

Despite advances in computational resources, which make features such as on-the-fly optimization more palatable, the time spent in the refinement is still large when compared with the time spent sampling. If one could judiciously reduce the number of orientations actually optimized, however, the refinement bottleneck might be dissipated. The large number of spatially distributed descriptors involved in molecular docking give rise to many geometrically similar orientations. By removing so-called 'degenerate' configurations, many noninformative minimizations are avoided. We describe progress toward this goal with a technique we refer to as 'degeneracy checking'.

Following the work of Meng et al. [6], this paper delves further into issues related to sampling and refinement in molecular docking. We investigate the tractability and utility of on-the-fly optimization, with and without coupling to a degeneracy checking protocol. The current sampling scheme used in DOCK is evaluated in light of these data.

Methods

Test systems

Six well-determined structures of ligand–receptor complexes available in the Brookhaven Protein Data Bank [7] were selected for analysis (Table 1): 1gst (glutathione S-transferase–glutathione [8]), 2gbp (D-galactose/D-glucose binding protein–β-D-glucose [9]), 3cpa (carboxypeptidase A–glycyl-L-tyrosine [10]), 3dfr (*L. casei* dihydrofolate reductase–methotrexate [11]), 4dfr (*E. coli* dihydrofolate reductase–methotrexate [11]) and 6rsa (ribonuclease A–uridine vanadate [12]). The 2gbp, 3cpa, 4dfr, and 6rsa systems have been used in previous investigations of sampling [6] and scoring issues [5], as has the 3dfr system [4,13]. For reasons noted in earlier work [5], the docked ligands for the 4dfr and 6rsa systems differ from the complexed ligands; they are 2,4-diamino-6-methylpteridine and uridine 3'-phosphate, respectively. The 1gst complex has proven a difficult one to reproduce with the current site characterization, so we introduce it as a stringent test of the methods applied here. Preparation for docking for all systems was carried out as described previously [5].

Force-field score optimization

A rigid-body minimizer, affecting only the six intermolecular rotational and translational degrees of freedom, was incorporated directly into the DOCK scoring scheme. The simplex technique of Nelder and Mead [14] was employed, with slight modifications in the convergence treatment. Because the simplex method requires no derivatives, it lends itself to optimization on a jagged potential surface. The function that is minimized is the grid-based force-field score of Meng et al. [5]. Polar hydrogens were given a small (0.6 Å) non-zero radius to prevent the minimizer from taking advantage of the large electrostatic attraction that would result from a charged, volumeless hydrogen approaching an oppositely charged nucleus. Construction of the initial simplex allowed up to 1.0 Å translation and 0.5° of rotation. Minimization convergence is treated in a two-stage fashion. Convergence within a simplex occurs when upper and lower bounds concur within 0.2 kcal/mol. Completion of a simplex signals a restart, initiating a new simplex. The minimization is deemed complete when a restarted simplex fails to reduce the force-field score by more than 1.0 kcal/mol. Other parameter values for simplex construction or convergence criteria resulted in slower and/or premature convergence (data not shown).

Explicit comparisons between the simplex minimizer

and the quasi-Newton method published previously [6] were carried out using the stand-alone programs DOCKMIN_SIM and DOCKMIN_DFP (distributed with DOCK 3.5). For each system, output from one DOCK run at an intermediate sampling level (400–600 orientations saved) was subject to stand-alone minimization. Performance was assessed for both minimization techniques in each of two modes: continuum (using exact interatomic distance calculations) and grid-based (using precalculated interaction scores and trilinear interpolation [5]). Stand-alone minimization was performed with default parameters.

Degeneracy checking

Basic algorithm Degeneracy checking aims to remove geometrically similar orientations of the ligand to reduce the number of time-consuming, on-the-fly minimizations. Because geometrically similar orientations usually converge to the same local minimum upon refinement, bypassing the optimization of these configurations will increase efficiency. We choose to remove degenerate orientations before the time-intensive 'orienting' phase [2,15], which places the ligand into the context of the receptor. As atomic coordinates are therefore not available, the difficulty lies in deciphering where in the active site an orientation lies based solely on the sphere–atom pairings involved in the match. The degeneracy checking algorithm must be able to perceive when the same geometry has been produced with different sphere–atom pairings (Fig. 1). When a unique orientation is found (e.g. the very first match), the new procedure records the nearest sphere to every atom in the ligand. Every subsequent orientation must be checked for degeneracy prior to the orienting phase. A simple check to see if all pairings occurred simultaneously in a previous unique match imparts the answer. To reduce the memory requirements of a list of matches for each possible sphere–atom pairing, we implement hashing, using open addressing with double hashing as described by Knuth [16]. The hash code enables rapid retrieval of matches containing a particular sphere–atom pairing.

Additional features To increase the level of observed degeneracy, we use a reduced set of 'virtual spheres' to define the active site. Virtual spheres result from single-linkage clustering (neighbor distance vsph) and averaging over the matching sphere set, providing an even distribution of points throughout the site. The nearest virtual sphere to each point on a cubic lattice is stored for rapid access during degeneracy assessment, analogous to the utilization of a force-field scoring grid for interaction evaluation. Additionally, we permit mismatch, termed wobble, in comparing sphere–atom pairings. Introducing mistakes into the degeneracy check (nonzero wobble) increases the number of degenerate orientations because binding modes are smeared out over a larger volume. Finally, we reduce the sensitivity, which results from representing orientational families by their first member, by affording popular binding modes renewed chances at locating a superior representative. The parameter degenerate_save_interval dictates how often a degenerate orientation must be found in a given family before orienting and minimizing another member. This feature has the desirable effect of smoothing sampling over all binding modes.

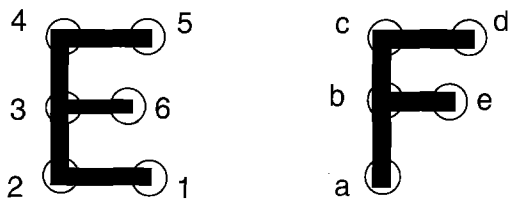

Fig. 1. Hypothetical, two-dimensional degeneracy checking example. **E** represents the receptor, **F** the ligand; spheres are numbered and atoms are indicated with a letter. Using a three-node match, one can superimpose **F** onto **E** by the pairings b3, c4, d5 or a2, e6, d5. The algorithm must recognize that these pairings will produce an identical geometric orientation.

Several degeneracy parameters may be varied, but their effects have not been examined systematically here. In preliminary exploration, we find that a vsph of 1.5–2.0 Å for creating virtual spheres, a wobble of 2, and a degenerate_save_interval of 10–25 offer a reasonable compromise between speed and accuracy. Despite the use of hashing to reduce storage requirements, memory demands are nonetheless considerable. Performance degrades as the hash table fills, so we advocate the use of degeneracy checking for low to medium sampling levels only.

Configurational sampling

DOCK, v. 3.5, was run in single mode for all docking studies. The matching algorithm for generating ligand orientations remains unchanged from that in DOCK 2.0 [4]. The number of configurations (matches) generated and thus the level of sampling performed is under user control through five parameters (all in units of Ångstroms). In addition to the matching tolerance, the user controls the ligand bin size, receptor bin size, ligand bin overlap, and receptor bin overlap. Enlarging bin sizes results in a greater number of atoms or spheres per bin, and a corresponding combinatorial expansion in possible matches. The overlap parameters smooth the discrete nature of the bin architecture and increase sampling by merging portions of neighboring bins. All orientations producing negative force-field scores were examined. To insure that timing results were unbiased by slow I/O routines, coordinates for acceptable matches were never written to disk.

Performance evaluation

A clear picture of the impact of new features related to sampling results from examining performance over a diverse array of sampling parameters. We vary two sampling parameters, bin size and bin overlap, independently

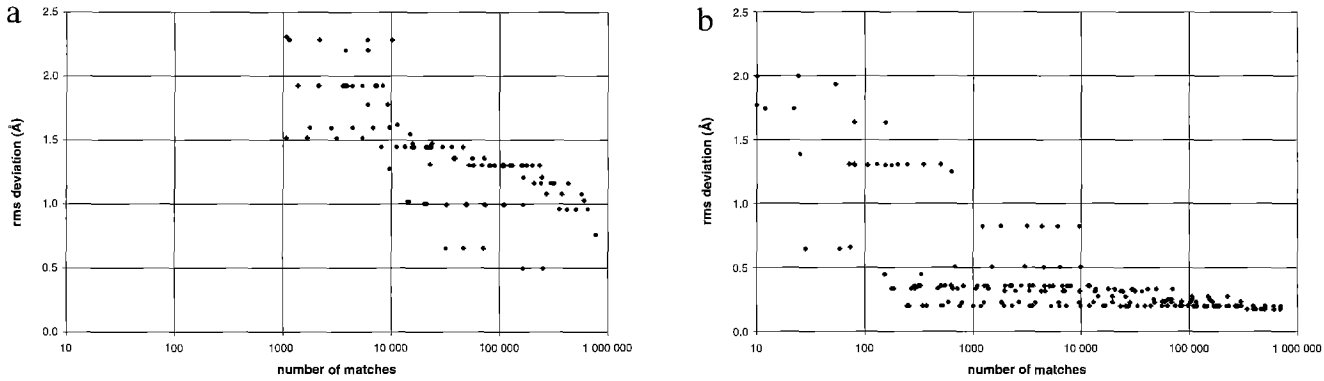

Fig. 2. Best rms deviation to the experimentally observed configuration seen, regardless of force-field score, plotted as a function of the number of matches attempted for (a) 1gst, and (b) 2gbp. Each point represents a single DOCK run with distinct sampling parameters. The behavior for the 3cpa, 3dfr, 4dfr, and 6rsa systems is like that of 2gbp.

in discrete increments over a large range. We set both the ligand bin size and receptor bin size equal to the variable bin size. Similarly, we set both the ligand bin overlap and receptor bin overlap equal to the variable bin overlap. Finally, we set the matching distance tolerance to be equal to the sum of the bin size and the bin overlap. The dependence of the distance tolerance on the bin parameters insures that all distance compatibility assessments for growing cliques are made with similar stringency. Bin sizes and bin overlaps ranged in increments of 0.05 Å from 0.05 to between 0.40 and 1.00 Å. In general, bin parameters were no longer incremented when run times began to exceed several minutes. This protocol led to a few hundred individual single mode DOCK runs per system, enabling a statistically significant analysis of result quality versus CPU time. All acceptable matches were formed from exactly four nodes and tolerated no more than two bad contacts.

For evaluation of new technology, four sets of runs as described above were performed for each system: native DOCK, with no new features; native DOCK with post-docking, lattice-based simplex optimization using DOCK-MIN_SIM; DOCK with on-the-fly force-field score minimization; and DOCK with on-the-fly force-field score minimization coupled to degeneracy removal. Data were transformed into a success-versus-effort format as follows. Effort was quantified in two ways: by the number of matches attempted, and by the amount of CPU time required. Success was also measured in two ways: by whether the rms deviation of the best force-field-scoring orientation was within 1.0 Å of the observed mode, and by whether the best force-field score obtained was within 5 kcal/mol about the global minimum. The global minimum force-field score was taken as the best force-field score seen by any of the DOCK runs for that system. Thus, this extremum represents the best among no fewer than several million configurations. The 5 kcal/mol threshold for success about the global minimum was selected based on an examination of the effect on the success-versus-effort plots of varying this threshold over the range 2.5–10 kcal/mol (data not shown). Effort is binned on a logarithmic scale: within each effort bin, a probability of success was computed by dividing the number of successful DOCK runs in the bin into the total number of DOCK runs falling in the bin. A seven-point moving average was used to smooth plots.

Sampling robustness

To assess whether failure by DOCK to reproduce experimentally observed geometries generally results from

TABLE 2
PERFORMANCE COMPARISON OF MINIMIZATION METHODS

System	CPU time per ligand (s)		Rms deviation[a] (Å)		Correlation[b]
	Continuum DFP	Grid simplex	Continuum DFP	Grid simplex	
1gst	3.10	0.070	1.04 ± 0.63	1.09 ± 0.63	$y = 0.88x - 2.30$; $r^2 = 0.77$
2gbp	1.47	0.039	0.76 ± 0.46	0.81 ± 0.44	$y = 0.97x - 0.83$; $r^2 = 0.80$
3cpa	2.93	0.062	0.91 ± 0.67	0.95 ± 0.72	$y = 1.01x - 0.76$; $r^2 = 0.86$
3dfr	3.92	0.115	1.14 ± 0.76	1.09 ± 0.65	$y = 1.01x - 1.03$; $r^2 = 0.95$
4dfr	2.72	0.037	0.50 ± 0.29	0.49 ± 0.28	$y = 1.00x - 1.30$; $r^2 = 0.87$
6rsa	2.02	0.064	1.37 ± 0.98	1.39 ± 0.96	$y = 0.99x - 1.28$; $r^2 = 0.88$

[a] The rms deviation from the starting position is given as average ± standard deviation; hydrogens were not included in the calculations. Values represent minimization of approximately 500 DOCK output orientations for each system.
[b] Correlations of continuum DFP force-field scores (y) versus grid simplex force-field scores (x).

TABLE 3
SAMPLING CONDITIONS EXPLORED IN METHODOLOGY EVALUATION

System	DOCK features[a]	Bin size range[b] (Å)	Bin overlap range[b] (Å)	No. of DOCK runs[c]	Total no. of matches	Matches per second	Global minimum[d]
1gst	Native	0.05–0.50	0.05–1.00	200	18 833 108	2647	–49.037
	Min	0.05–0.40	0.05–0.80	128	2421 887	83	
	Min+deg[e]	0.05–0.50	0.05–1.00	128	1783 514	626	
2gbp	Native	0.05–0.50	0.05–1.00	200	14 270 188	4187	–24.538
	Min	0.05–0.50	0.05–0.50	100	306 144	52	
	Min+deg[e]	0.05–0.50	0.05–1.00	126	956 378	281	
3cpa	Native	0.05–1.00	0.05–1.00	300[f]	28 582 470	2566	–47.188
	Min	0.05–0.50	0.05–0.50	100	117 713	78	
	Min+deg[e]	0.05–0.50	0.05–1.00	162	1597 427	626	
3dfr	Native	0.05–0.40	0.05–0.80	128	7136 487	2863	–70.945
	Min	0.05–0.40	0.05–0.40	64	125 421	326	
	Min+deg[g]	0.05–0.40	0.05–0.80	111	2616 774	1882	
4dfr	Native	0.05–0.50	0.05–1.00	200	6207 365	2354	–33.916
	Min	0.05–0.50	0.05–0.50	100	180 282	36	
	Min+deg[e]	0.05–0.50	0.05–1.00	152	1951 826	293	
6rsa	Native	0.05–0.50	0.05–1.00	200	2834 980	1731	–66.003
	Min	0.05–0.50	0.05–0.50	100	68 953	68	
	Min+deg[e]	0.05–0.50	0.05–1.00	171	1493 590	596	

[a] 'Native' refers to DOCK runs in which neither force-field score minimization nor degeneracy checking was used. 'Min' refers to DOCK runs in which force-field score minimization was used without degeneracy checking. 'Min+deg' refers to DOCK runs in which force-field score minimization was used in conjunction with degeneracy checking.
[b] Increments of 0.05 Å were used within these ranges.
[c] The number of DOCK runs examined is in some cases less than the bin ranges would indicate, for three possible reasons: run times began to exceed several minutes, convergence at 100% in the success-versus-effort plots had been reached, or the maximum number of allowable unique matches for degeneracy checking had been exceeded.
[d] Minimum force-field score (kcal/mol) observed over all DOCK runs for each system.
[e] Degeneracy parameters: wobble = 2, vsph = 1.5, degenerate_save_interval = 10.
[f] For bin sizes of 0.55–1.00 in the 3cpa native DOCK runs, bin overlaps ranged only from 0.55 to 1.00, hence only 300 runs resulted. This was an effort to obtain more high-sampling runs.
[g] Degeneracy parameters: wobble = 2, vsph = 2.0, degenerate_save_interval = 25.

deficiencies in sampling or in scoring, we isolated the effects from sampling. By removing scoring restrictions and analyzing only agreement in Cartesian space between docked orientations and the observed binding mode, the precision of the sampling algorithm is revealed. A set of DOCK runs with sampling level varied as described above was thus performed in which all orientations within 2.5 Å rms deviation from the experimentally observed configuration were written out, regardless of force-field score.

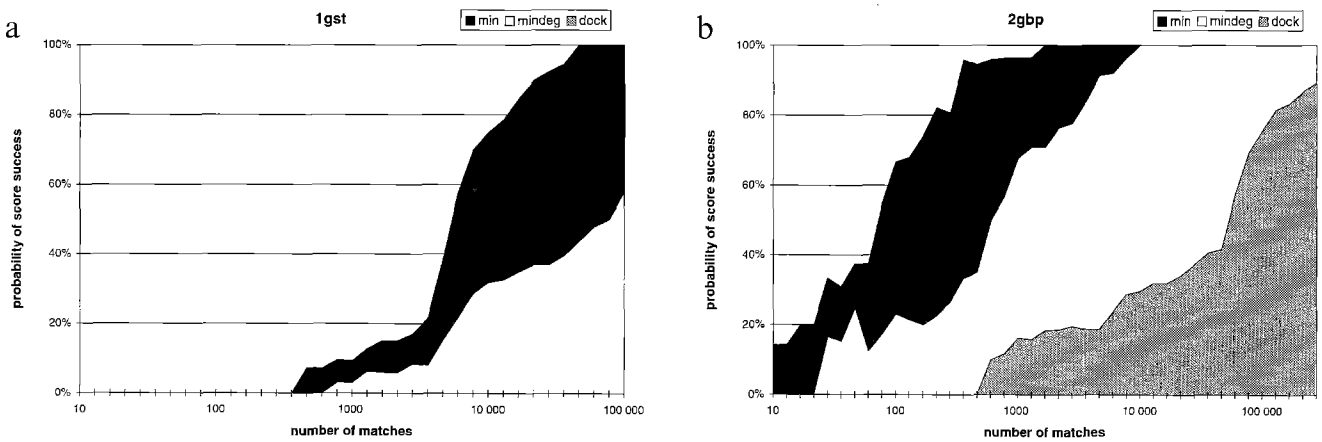

Fig. 3. The probability of locating an orientation having a force-field score within 5 kcal/mol of the global minimum, plotted as a function of number of matches attempted for (a) 1gst, and (b) 2gbp. 'dock' represents native DOCK, 'min' represents DOCK with on-the-fly minimization, and 'mindeg' represents DOCK with on-the-fly minimization and degeneracy checking. The absence of a curve for native DOCK indicates that no successful run ever occurred. The 1gst and 2gbp systems represent extrema; the four other systems show similar features at intermediate points along the abscissa.

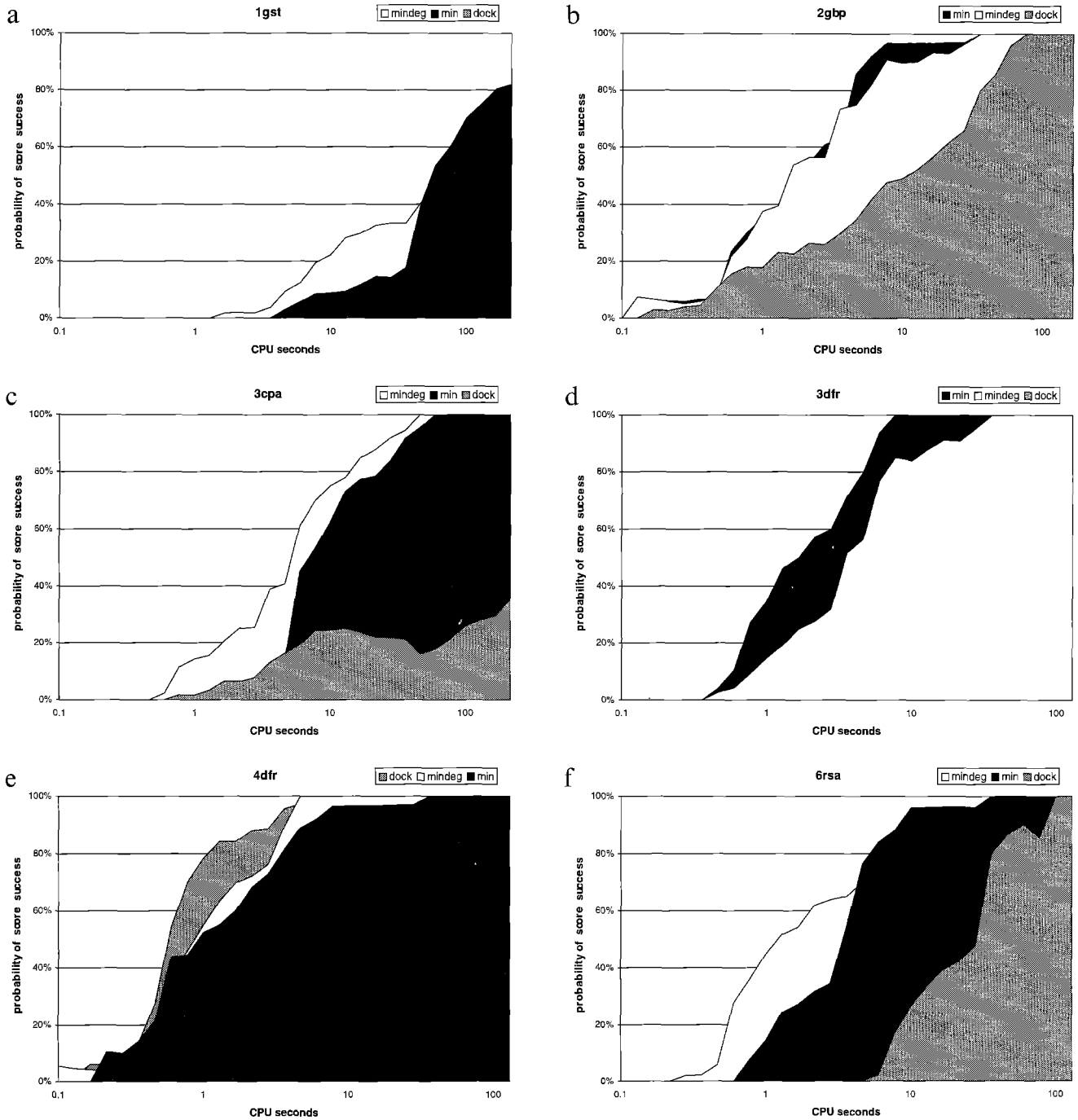

Fig. 4. Probability of locating an orientation having a force-field score within 5 kcal/mol of the global minimum, plotted as a function of CPU seconds required for (a) 1gst; (b) 2gbp; (c) 3cpa; (d) 3dfr; (e) 4dfr; and (f) 6rsa. The key is as given in the legend for Fig. 3.

Hardware

All calculations were carried out on a 200 MHz R4400 INDIGO^2 workstation (Silicon Graphics, Inc., Mountain View, CA) with 128 Mb of physical memory.

Results

Sampling robustness

The ability of DOCK's sampling algorithm to locate the experimentally observed binding mode is illustrated in Fig. 2. For each system, all sampling levels that produced an orientation within 2.5 Å rms deviation are plotted. For all six receptor–ligand complexes explored here, the matching algorithm is robust enough to find the native configuration. With the exception of the 1gst system (Fig. 2a), a few hundred to a thousand matches are sufficient to locate an orientation within 1.0 Å rms deviation. This point highlights the robust nature of the sphere description and matching algorithm used in DOCK. As the sampling method is adequate, it thus becomes a task

for scoring schemes to recover the native mode as the optimal configuration.

Minimizer performance

A fast rigid-body optimization, suitable for incorporation into DOCK, operates as effectively as the more resource-intensive method explored by Meng et al. [6]. By implementing a simplex method using a precalculated interaction lattice, between 30- and 75-fold faster operation is achieved over the continuum-mode quasi-Newton Davidon–Fletcher–Powell (DFP) [17] method described previously [6]. The near-unit slopes and high correlation between optimized scores indicate that the result quality is comparable. The offset favoring the continuum DFP by 1–2 kcal/mol is attributable to the use of exact interatomic distances rather than trilinear interpolation among precalculated grid scores. We take as a measure of convergence radius, or the capacity to pull distant structures into a local minimum, the rms deviation occurring during minimization. Convergence radii for the two minimization techniques are nearly identical. The simplex operating in continuum mode and the grid-based DFP demonstrated performance intermediate to the two methods presented in Table 2 (data not shown).

On-the-fly optimization and degeneracy checking

The performance impact of on-the-fly force-field score optimization and of degeneracy checking was gauged via success-versus-effort analyses. The range of sampling parameters, number of DOCK runs, and total configurations generated for each set of runs are enumerated in Table 3. More than 2500 DOCK runs, covering a wide range of sampling conditions, have allowed a comprehensive analysis of tradeoffs between configurational exploration and rigid-body optimization.

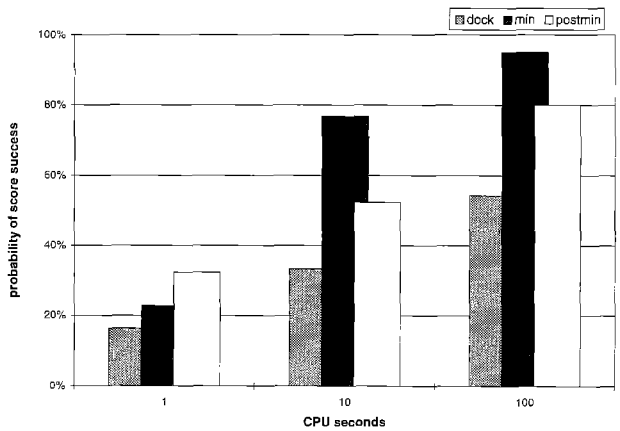

Fig. 5. Probability of locating an orientation having a force-field score within 5 kcal/mol of the global minimum, as a function of CPU seconds required. 'dock' represents native DOCK, 'min' represents DOCK with on-the-fly minimization, and 'postmin' represents native DOCK with stand-alone grid-based simplex minimization performed on the output. Data represent an average over the systems explored for each method.

Figure 3 illustrates the ability of new technology to locate the global minimum in force-field score as a function of sampling level. The use of score optimization consistently outperforms native DOCK in this respect. This was to be expected: both methods generate identical orientations, but the former is afforded a refinement of intermolecular interactions, an operation that can only improve results. Coupling to a degeneracy checking protocol would ideally show identical behavior to minimization alone, when effort is measured by the number of matches. In actuality, the degeneracy checking method generally falls intermediate to DOCK with and without minimization. In two systems examined here (1gst and 3dfr, not shown), native DOCK is completely unable to locate an orientation close to the global minimum in the absence of refinement, even when sampling on the order of one million configurations. Plots of success in placing the best force-field-scoring orientation within 1 Å rms deviation of the experimentally observed configuration as a function of number of matches tried parallel the force-field score success plots of Fig. 3 (data not shown).

In practice, however, the primary concern for molecular docking is not how many configurations are examined, but rather how much computer time is required. Because each optimization takes on average one hundred times longer to carry out than a single score evaluation (data not shown), DOCK runs employing force-field score minimization are likely to become intractable unless sampling is reduced. But can sampling be reduced sufficiently to counteract this great disadvantage while maintaining high-quality solutions? Figure 4 depicts the transformation from effort measured in numbers of configurations to effort gauged by computational demands.

Excepting only the 4dfr system (Fig. 4e), we see that using on-the-fly optimization is dramatically more efficient than native DOCK at arriving at near-global-minimum solutions, despite the much higher per-match resource requirements (Table 3). The implementation of the degeneracy checking protocol, while equally superior to native DOCK, does not display as dramatic improvements when compared with minimization alone. In one case (6rsa) we see significant gains, in two cases (1gst, 3cpa) slight improvements, in two cases (2gbp, 4dfr) no difference, and in one case (3dfr) slightly worse behavior. Degeneracy checking generally manifests its advantages at lower sampling levels, as evidenced by the early successes seen in the 1gst, 3cpa, and 6rsa complexes.

A simple alternative to introducing force-field score optimization into the docking process would be to perform stand-alone minimization on the output of a native DOCK run. Given the negligible cost of a single grid-based simplex refinement (Table 2), this could conceivably be an efficient method for improving results. We have entertained this possibility in four of the test systems, and compare post-DOCK minimization to native DOCK and

DOCK with on-the-fly optimization in Fig. 5. At very short run times, post-DOCK minimization is the most effective method, but its usefulness becomes limited as run times lengthen. In one system (6rsa, data not shown), post-DOCK minimization showed almost no improvement over native DOCK. Possible explanations for why this behavior is likely to be common are given below.

Discussion

Perspective

Molecular docking has become an increasingly popular tool for drug discovery in recent years [18]. To be truly useful, docking methods must successfully integrate effective site description techniques, robust configurational sampling algorithms, and accurate evaluation schemes in an efficient manner. Our focus here is on a feature that ties together sampling and evaluation, i.e., interaction optimization. This feature is designed to improve how two components fit together, but the physical movement involved in the refinement impinges directly upon the apparent performance of the sampling algorithm. Thus, our investigation into the utility of rigid-body refinement in DOCK necessarily probes configurational search methods.

Interaction optimization is not new to automated molecular docking methods [19–22]. However, to our knowledge, this article represents the first published systematic exploration of sampling space for a docking method. We analyze in excess of 2500 docking runs, not simply an arbitrary slice of the vast configurational universe. This study enables an objective analysis of the tradeoff between computationally inexpensive, discrete optimization in the form of configurational sampling and the considerably more expensive, continuous optimization in the form of rigid-body refinement.

Our assessment of the results is colored by our standpoint on molecular docking as a tool for database searching toward lead discovery. This perspective carries two biases associated with it: (i) we prefer the amount of CPU time spent per ligand to be on the order of seconds, not minutes; and (ii) we rank binding modes and ligands by interaction scores, not rms deviations to observed configurations. Experimental orientations are unavailable for nearly all database ligands, so a geometrical success criterion cannot be utilized. The latter point implies that efforts should be directed toward locating the global minimum in a scoring function, not necessarily toward identifying a known binding mode. We make the assumption that the experimentally observed orientation is at the global minimum. It is therefore the task of scoring function developers to insure coincidence between the global optimum of the evaluation scheme and the observed mode. For all six systems studied here, the global minimum of the force-field score developed by Meng et al. [5] does indeed correspond to the crystallographic solution to within 0.5 Å rms deviation.

Robustness in sampling and optimization

The coupling of on-the-fly interaction optimization with an effective configurational sampling algorithm results in a robust docking strategy. The speed of the accurate grid-based simplex minimizer introduced here enables the incorporation of refinement into the docking process, albeit still at considerable computational expense when compared with the speed of matching or force-field scoring alone (Table 3). Nevertheless, on-the-fly optimization is able to not only counteract this handicap, but significantly surpass native DOCK in efficiently locating low-energy solutions. Because the sampling method readily retrieves configurations close to the experimentally observed configuration, failure to identify this mode as optimal lies with scoring and not with sampling. The effect of refinement, then, is to salvage the many orientations generated near the crystallographic mode that would otherwise be thrown out due to steric clashes with the receptor. Optimization allows maximal use to be made of information provided by the matching algorithm. We expect on-the-fly optimization to benefit database searches most by rescuing ligands for which the proper binding mode is sampled but for which no low-energy orientations can be found. Two such examples appear in this work, 1gst and 3dfr (Fig. 4), and their recovery underscores the utility of on-the-fly optimization. The tolerance of a nonzero number of bad contacts within DOCK is imperative to taking full advantage of minimization as a rescue device.

Degeneracy removal

The degeneracy checking protocol described here has met with mixed success. Although typically 90% of orientations are deemed degenerate and are not examined further, this savings under the current implementation does not significantly outweigh the cost of assessing degeneracy. The advantages are manifested primarily at shorter run times, as evidenced in the 1gst, 3cpa, and 6rsa systems (Fig. 4). This capacity will find use in database searching applications when CPU resources are quite limited, as not all ligands are likely to be sampled adequately with the same set of sampling parameters.

The judicious selection of fewer orientations for optimization is a compromise. By refining all orientations, resources are spent insuring each orientation is within a local minimum, not sampling the vast configurational universe (akin to a depth-first search). Conversely, by not refining any orientations, resources are spent exploring configuration space without particular regard to the quality of each orientation (a breadth-first search). Refinement is relatively expensive computationally and configurational exploration is inexpensive, so the optimal tradeoff comes when configuration space is thinly but evenly sam-

pled with refined orientations. The advantages evidenced with the degeneracy removal protocol at short run times are the result of exactly this tradeoff. At longer run times, when inexpensive configurational sampling is more intense, minimization alone generally performs at least as well as when coupled with degeneracy removal.

We believe the largest hurdle in devising a more successful degeneracy removal protocol lies in the selection of a representative for each binding mode. In this work, we choose the first orientation found in a binding mode as that family's 'parent' for assessing degeneracy. If this orientation should be a poor representative, further orientations in that family will nonetheless be thrown out, regardless of how they might have scored. The degenerate_save_interval alleviates this bias to some extent, but functions as a crutch rather than a solution.

Prospects for post-DOCK optimization

The appropriate control experiment for the introduction of on-the-fly minimization entails performing a DOCK run without minimization and subsequently optimizing the output in the same fashion. In this way, the benefits imparted by minimizing all DOCK orientations as opposed to minimizing only the best unoptimized orientation are revealed. The danger in selecting only the lowest energy unoptimized orientation is that other orientations may lie higher in energy but in a deeper well, so that upon optimization these other orientations would have finished lower in energy. This possibility is borne out by the shuffling of pre- and post-optimization force-field scores (data not shown).

Although we observe the best performance with post-docking optimization for very short DOCK runs (Fig. 5), this behavior does not extend to more intensive sampling. Finding an orientation in the observed binding mode is a necessary but not a sufficient condition for obtaining a force-field score near the global minimum after optimization. Because on-the-fly optimization refines every orientation, it is afforded the luxury of the chance that any of the orientations near the observed binding mode (Fig. 2) will refine near to the global minimum in force-field score. In contrast, DOCK without on-the-fly optimization has available only one orientation deemed best by an unoptimized force-field score, with the additional constraint that this one orientation must be in the observed binding mode. DOCK without on-the-fly optimization therefore gets at most one chance to refine an orientation into the global minimum if post-docking optimization is performed. Nevertheless, we find that performing a post-DOCK optimization is in all cases superior, and in many cases substantially so, to performing a native DOCK run without any refinement.

Matching algorithm discontinuities

A disconcerting consequence of the bin architecture for ligand-site matching is that results obtained at a low level of sampling are not guaranteed to be a subset of results obtained at a higher level of sampling. This point has been noted previously [6]. Although in general this is not the case, this artifact can lead to strange behavior, particularly when examining arbitrary slices of sampling parameters. The analysis of hundreds of DOCK runs for each system in this study enables us to collect statistically significant success probabilities and bypass much of the problem. One will note, however, that the plots in Figs. 2–4 do not display monotonic functions: the jagged nature of these curves is the result of the discontinuity arising from the bin architecture. Fortunately, the physical convergence of orientations into local minima by on-the-fly minimization mitigates the severity of this artifact.

Sampling guidelines

One of the most instructive findings from the great number of DOCK runs examined is insight into the amount of sampling required to obtain a desired probability of success. The success-versus-effort plots carry a great deal of information, and can be used as guidelines for performing DOCK runs appropriate to available resources. For instance, one might be interested in performing a large database search where each ligand would be allotted the minimum resources to obtain 100% success. In this case, one might calibrate sampling conditions to expend an average of 10 CPU seconds per ligand (or on the order of 1000–3000 matches). In another example, one might be interested in analyzing a small database with the assurance that each ligand was well into the 100% success plateau. For this case, one might calibrate sampling conditions to expend 100 CPU seconds per ligand. It would be reasonable to construct a success-versus-effort plot for a known ligand, if available, for performance gauges customized to the system being studied. In this manner, the success-versus-effort plots provide a valuable mechanism for setting sampling levels in molecular docking.

Conclusions

We have coupled a fast and effective grid-based, rigid-body simplex minimizer with the robust configurational sampling algorithm used in DOCK to allow on-the-fly force-field score optimization in a tractable manner. This coupling, despite the heavy resource demands of refinement, results in a docking strategy that is computationally more efficient at retrieving experimentally observed configurations than docking in the absence of optimization. In some cases, only with the use of on-the-fly optimization could the observed binding mode be identified as the global minimum in the scoring function. On-the-fly optimization salvages poor orientations that would otherwise be discarded, thus making maximal use of informa-

tion afforded by the sampling algorithm. The removal of geometrically similar orientations to circumvent redundant optimizations is a tradeoff between expensive refinement and inexpensive sampling – our implementation shows mixed success, but with greatest potential at short per-ligand run times. Finally, while not as effective as on-the-fly optimization, it is highly beneficial to perform an inexpensive post-docking optimization, particularly at low sampling levels. We find that success-versus-effort plots for gauging docking performance lend valuable insight into the setting of sampling levels for the inevitable compromise between result quality and computational resources.

Acknowledgements

We thank Todd Ewing, Andrew Good, Elaine Meng, and Brian Shoichet for numerous insightful discussions. This work was funded through NIH Grants GM-31497 (I.D.K.) and GM-39552 (G.L. Kenyon, principal investigator). D.A.G. gratefully acknowledges support from an ACS Division of Medicinal Chemistry Fellowship. Forcefield score optimization and degeneracy checking features are available in DOCK v. 3.5. The DOCK suite of programs is implemented in FORTRAN and accessible through I.D.K.

References

1 Blaney, J.M. and Dixon, J.S., Perspect. Drug Discov. Design, 1 (1993) 301.
2 Kuntz, I.D., Meng, E.C. and Shoichet, B.K., Acc. Chem. Res., 27 (1994) 117.
3 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
4 Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
5 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
6 Meng, E.C., Gschwend, D.A., Blaney, J.M. and Kuntz, I.D., Proteins, 17 (1993) 266.
7 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
8 Ji, X., Zhang, P., Armstrong, R.N. and Gilliland, G.L., Biochemistry, 31 (1992) 10169.
9 Vyas, N.K., Vyas, M.N. and Quiocho, F.A., Science, 242 (1988) 1290.
10 Rees, D.C. and Lipscomb, W.N., Proc. Natl. Acad. Sci. USA, 80 (1983) 7151.
11 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
12 Borah, B., Chen, C.-W., Egan, W., Miller, M., Wlodawer, A. and Cohen, J.S., Biochemistry, 24 (1985) 2058.
13 Shoichet, B.K. and Kuntz, I.D., Protein Eng., 7 (1993) 723.
14 Nelder, J.A. and Mead, R., Comput. J., 7 (1965) 308.
15 Ferro, D.R. and Hermans, J., Acta Crystallogr., A33 (1977) 345.
16 Knuth, D.E., The Art of Computer Programming, Vol. 3, Addison–Wesley, Menlo Park, CA, 1973, pp. 506–549.
17 Fletcher, R., Practical Methods of Optimization, Interscience, New York, NY, 1960.
18 Gschwend, D.A., Good, A.C. and Kuntz, I.D., J. Mol. Recogn., in press.
19 Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
20 Hart, T.N. and Read, R.J., Proteins, 13 (1992) 206.
21 Yamada, M. and Itai, A., Chem. Pharm. Bull., 41 (1993) 1200.
22 Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., J. Comput.-Aided Mol. Design, 8 (1994) 153.