Development and Testing of an Automated Approach to Protein Docking

Andrei Tovchigrechko[†] and Ilya A. Vakser[*†]
Bioinformatics Laboratory, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York

ABSTRACT **A new version of GRAMM was applied to Targets 14, 18, and 19 in CAPRI Round 5. The predictions were generated without manual intervention. Ten top-ranked matches for each target were submitted. The docking was performed by a rigid-body procedure with a smoothed potential function to accommodate conformational changes. The first stage was a global search on a fine grid with a projection of a smoothed Lennard–Jones potential. The top predictions from the first stage were subjected to the conjugate gradient minimization with the same smoothed potential. The resulting local minima were reranked according to the weighted sum of Lennard–Jones potential, pairwise residue–residue statistical preferences, cluster occupancy, and the degree of the evolutionary conservation of the predicted interface. For Targets 14 and 18, the conformation of the complex was predicted with root-mean-square deviation (RMSD) of the ligand interface atoms 0.68 Å and 1.88 Å correspondingly. For Target 19, the interface areas on both proteins were correctly predicted. The performance of the procedure was also analyzed on the benchmark of bound–unbound protein complexes. The results show that, on average, conformations of only 3 side-chains need to be optimized during docking of unbound structures before the backbone changes become a limiting factor. The GRAMM-X docking server is available for public use at http://www.bioinformatics.ku.edu. Proteins 2005;60:296–301.**

Key words: CAPRI, GRAMM; protein modeling; molecular recognition; energy landscape

INTRODUCTION

The growing needs of experimental and computational biology require reliable computational procedures for modeling of protein interactions. Recent progress in docking algorithms and computer hardware makes it possible to implement such procedures as automated Web servers, which greatly improves the utility of the docking approaches in the biological community. Such servers also allow an objective test of the underlying docking methodologies, unbiased by expert human intervention. In preparation for the launch of our docking Web server GRAMM-X (http://www.bioinformatics.ku.edu), we applied the procedure to the Round 5 targets of the CAPRI challenge. GRAMM-X is a development of the Fast Fourier Transformation (FFT) GRAMM methodology.[1–3]

The field of protein docking is currently undergoing a major expansion.[4–6] A traditional paradigm in many existing docking procedures, both FFT- and non-FFT-based, is to adopt a multiscale approach in which proteins are initially docked as rigid bodies using simple force fields (scan stage), followed by a reevaluation of the predicted matches by computationally more expensive scoring functions and a possible structure refinement (refinement stage). The scan stage targeted precision varies among the approaches, from relatively "realistic" high-resolution representation and force fields to low-resolution protocols[7–9] capable of tolerating significant structural discrepancies (with the loss of precision). The refinement stage sometimes involves, along with the minimization in the external coordinates (rigid-body translations and rotations), a limited search through the internal coordinates of the individual protein structures.

The advances in computer power make possible more detailed force fields and the explicit treatment of an increasing number of the internal coordinates. Interestingly, however, recent studies[10,11] suggest that a conformational search in only a small number of selected side-chains may account for most of the internal conformational flexibility needed for an accurate docking prediction. The results of our study confirm this notion.

GRAMM-X procedure implements a fine-grid FFT search with a smoothed Lennard–Jones potential, followed by local minimization and rescoring. The performance of the procedure is evaluated on the Round 5 targets, including the released bound conformations, and on the benchmark set of protein–protein complexes.

METHOD

GRAMM-X is a development of the correlation by FFT GRAMM docking methodology.[1–3] In the original GRAMM method, the intermolecular energy potential is a step function approximating Lennard–Jones potential, based

Grant sponsor: National Institutes of Health; Grant number: R01 GM61889.

*Correspondence to: Ilya A. Vakser, Center for Bioinformatics, University of Kansas, 1460 Jayhawk Boulevard, Lawrence, KS 66045-7523. E-mail: vakser@ku.edu

[†]Current address: Center for Bioinformatics, University of Kansas, 1460 Jayhawk Boulevard, Lawrence, KS 66045-7523

Received 16 January 2005; Accepted 1 February 2005

DOI: 10.1002/prot.20573

on the grid representation of the molecules.[12] The smoothing of the intermolecular energy landscape is achieved by lowering the value of the repulsion part component and the increase of the range. Since the range is the step of the grid, the step becomes larger and the structural representation of the molecules is reduced to lower resolution. Such an approach has an important advantage of implementation simplicity and computational speed. It also allows the study of fundamental molecular recognition characteristics focusing on underlying simplicity of the basic principles.[13–15] However, in practical docking applications aimed at maximizing the chances of the correct prediction, the association of the potential range with the grid step often becomes a disadvantage. The grid step–potential range association is not suitable for more sophisticated forms of the potential. At lower resolution it is also sensitive to the positioning of the molecules for the grid digitization, introducing a significant degree of random noise. Disassociation of the interatomic energy potential from the grid, implemented in GRAMM-X, provides a possibility to address these negative factors.

The procedure uses a fine-grid projection of a softened Lennard–Jones potential function[16] calculated for a probe atom:

$$V_{ij}(r) = \frac{1}{\alpha\sigma_{ij}^6 + r^6}\left(\frac{4\varepsilon_{ij}\sigma_{jj}^{12}}{\alpha\sigma_{ij}^6 + r^6} - 4\varepsilon_{ij}\sigma_{ij}^6\right).$$

The benchmarking docking showed that the optimal values of the parameters are $\alpha = 0.4$, $\sigma = 0.33$ nm, and $\epsilon = 0.5$. These uniform values of σ and ϵ applied to all nonhydrogen atoms yielded better results than the values taken from the AMBER atom types. The docking runs also showed that the results do not improve for translation grid steps < 1.5 Å and rotation steps $< 10°$, which can be explained by the subsequent minimization of the grid predictions in continuous space.

The top 2000 grid-based predictions were subjected to a conjugate gradient minimization in continuous 6-dimensional (6D) (rigid body) space with the same soft potential. The minimization accumulates many points, initially located on the grid, in a fewer local minima. Thus, only 1 representative prediction for each minimum was stored, and the number of initial predictions which fell into this minimum was marked as the "volume of the minimum." The minimized predictions were reranked by a scoring function that is a weighted sum of the soft Lennard–Jones term, the evolutionary conservation of predicted interface, the statistical residue–residue preference term, and the volume of the minimum.

Our assumption was that interface areas are functionally important and thus must be more conserved among homologs. Thus, we used the evolutionary conservation term as the sum of individual residue conservation scores at the predicted interfaces of both proteins, weighted by the sequence length. The individual scores were taken from http://consurf.tau.ac.il.[17]

The statistical residue preferences term was calculated on the basis of a 20×20 table of normalized contact frequencies for all residue types. The table was obtained by analyzing the statistical data on protein complexes from the Protein Data Bank (PDB).[18] We use the average value for contact pairs formed at the predicted interface as the target function component.

GRAMM-X is implemented in Python and C++, thus combining fast prototyping power of Python and numerical performance of C++ modules. The Message Passing Interface (MPI) library is used for parallelization. On the benchmark of unbound–bound complexes,[19] the full docking protocol (FFT + refinement) for a single complex, on average, completes in 6 min on the cluster of 20 1Ghz PIII processors. The GRAMM-X Web interface has been made public at http://www.bioinformatics.ku.edu and allows users to run docking jobs on our computer cluster. Upon the completion of the user's job, the reference to the generated Web page with results is sent to the submitter via e-mail. The procedure is currently ready for participation in future CAPRI rounds as an automated server.

RESULTS AND DISCUSSION

Since our research is focused on the development of the docking methods, we were interested in the blind trial of the docking program. The docking was performed without any use of external data to restrict the search space or manual selection of the results. The top 10 ranked predictions from the refinement stage were submitted to the CAPRI server. For the antibody–antigen complex (Target 19) the search was not restricted to complementarity determining regions (CDRs), although only the variable domain was used in docking.

The accuracy of Target 14 (prediction 4) and Target 18 (prediction 5) submissions was 0.68 Å and 1.88 Å root-mean-square deviation (RMSD) of the predicted ligand interface, correspondingly. For Target 19, the interface areas on both receptor and ligand were correctly predicted (the ligand was predicted in a rotated position, thus failing to correctly identify pairwise contacts or produce small RMSD).

Statistically Significant Testing Set

For an objective analysis of the algorithm performance, it is essential to use a statistically significant testing set. Thus, although we describe how different aspects of our approach affected the CAPRI results, we also support our conclusions with the data obtained on the bigger benchmark set of bound–unbound protein structures.[19] Because our study involved testing the performance of the combined scoring function that included the evolutionary conservation of residues, we had to exclude the antibody–antigen benchmark complexes where this component does not have a role. We also excluded complexes marked by the benchmark authors as "difficult" (those with large backbone changes at the interface), because the rigid-body methods are not suited for such cases by design.[5] That left us with 31 benchmark complexes. In this set, the bound–unbound all-atoms RMSD for the receptors and the ligands interfaces was 1.09 Å and 1.12 Å correspondingly, and C^α RMSD = 0.57 Å and 0.45 Å, correspondingly. The average number of the interface residues on the receptors

TABLE I. Significance of Different Refinement Stages in the Benchmark Docking

No.	Target function	Minimization	Clustering	Rescoring	No. of complexes with good rank ≤ 10	Avg. good rank
1	VW only[a]				9	142
2	VW only	+			9	146
3	VW only	+		+	6	243
4	VW only	+	+		9	102
5	VW only	+	+	+	11	79
6	Full[b]			+	12	89
7	Full	+	+	+	15	51

[a]The rescoring (when performed as indicated) is based on the volume of the minimum (see text) alone.
[b]The rescoring is based on the volume of the minimum plus evolutionary conservation and statistical potential.

and ligands was 21 and 15, correspondingly. An interface residue was defined as the residue with at least 1 atom within 4 Å from any atom in the other protein.

Evaluation Criteria

An important aspect of benchmarking is the criterion by which to evaluate the success of docking for a particular complex. Because the benchmarking is used to assess the improvement usually brought by incremental changes in the algorithm, it is important to have some integral scalar value that can be easily compared in different runs of the benchmarking. The number or the ratio of predicted native residue contacts is poorly suited for this purpose, because it must always be considered in tandem with the number (ratio) of predicted non-native contacts (otherwise, the comparison will be biased toward the predictions with a lot of receptor–ligand interpenetration). Thus, the contact ratio results cannot be always meaningfully ordered to reflect the relative quality of docking. Our preference was to use the RMSD of the ligand interface C^{α} atoms after superimposing the interface C^{α} atoms of the receptor (similarly to the formal CAPRI assessment). We define a particular prediction as "good" if its RMSD is < 5 Å. Then we find the rank of the first "good" prediction in our final output. We consider that the complex was docked "successfully" if the rank of such prediction is ≤ 10. The rationale is that a prediction with 5 Å RMSD provides useful biological information, and that 10 putative conformations is a number small enough for experimental verification, giving such a prediction a practical value. The number of complexes docked "successfully" according to the above criterion is the basic output of our benchmarking protocol. These results are presented in the rest of this article. We must note that the ranks of the first good match for CAPRI targets are slightly different from those reported by the CAPRI assessors for our original submission (rank 8 for Target 14, and rank 3 for Target 18 instead of ranks 4 and 5, correspondingly). The reason must be that at the time of preparing this report, we had introduced some small changes to the docking program. Also, our definition of *contact residue* differs from the one used by CAPRI assessors. However, these differences are not important in the context the presented analysis.

Bound–Unbound Docking Comparison

In the set of 31 benchmark complexes, 15 unbound complexes were successfully predicted (compared to 27 bound ones), according to the above criterion. In 6 of the unbound complexes, the good match had rank 1 (in 24 of the bound complexes). For CAPRI Target 14, the good match had rank 8 (rank 1 for bound structures); for Target 18, rank 3 (rank 1 for bound structures); and for Target 19, rank 138 (rank 1 for bound structures). The results indicate that the problems encountered in docking of the unbound structures in most cases may be attributed to the conformational changes upon binding, rather than to the inaccurate description of the energy landscape.

Refinement Significance

Refinement stage, in the benchmark set, increased the number of good predictions in the first 10 predictions from 9 to 15, and in the first 100 predictions, from 20 to 22. For the CAPRI Target 14, the first good prediction moved up from rank 18 to rank 8, and for Target 18, from 17 to 3. For Target 19, it actually moved down, from rank 27 to rank 138. As these results show, the refinement significantly improved the number of complexes where a good prediction was found with rank < 10 (from 9 to 15 complexes), but not significantly for ranks < 100. In the case of the CAPRI targets, which were limited to 10 predictions per complex, the refinement stage turned out to be crucial for the success.

Role of Refinement Factors

As discussed in the Method section, the refinement itself consists of 3 stages: local minimization with soft van der Waals (VW) interaction, clustering of predictions within the same local minima, and rescoring with the target function combining Lennard–Jones and other terms. To understand the role of the refinement factors, we performed the benchmarking switching on and off its different stages (Table I). Row 1 corresponds to the results of the FFT stage alone, without any additional processing, and row 2 corresponds to minimization of the FFT predictions without changing their ranking. The fact that the results are practically identical suggests that the minimization does not significantly change RMSD of the FFT predictions. In other words, minimization will not create good

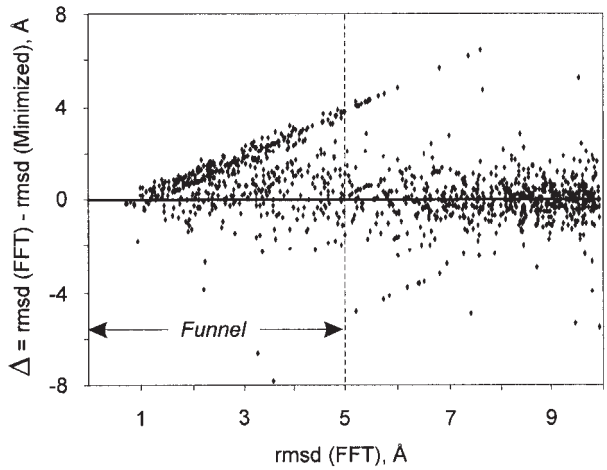

Fig. 1. RMSD change after the refinement minimization as a function of RMSD before the refinement in 31 benchmark complexes. RMSD is calculated for the ligand C^{α} atoms at the predicted interface relative to their positions in the crystallographic complex. RMSD of ~5 Å indicates characteristic size of the intermolecular energy funnel (see text).

predictions out of poor ones. To investigate this aspect further, we built a scatter plot of initial RMSD of prediction versus the change of its RMSD after the local minimization (a positive value means that the final RMSD is lower). Figure 1 shows the plot, up to RMSD 10 Å (beyond that, RMSD change is random around zero, because predictions are located in minima completely unrelated to the correct crystallographic structure). The data show that the 5 Å cutoff value has a special meaning: Minimization generally improves the predictions that already have the initial RMSD below that value, and decreases the accuracy of those beyond it. We can think of 5 Å RMSD as the characteristic size of the energy funnel on our smoothed landscape. The average improvement of RMSD for predictions within this funnel is 0.6 Å. Row 3 of Table I demonstrates that when the predictions are reranked according to the minimized VW energy without clustering them first, the docking results get worse. It happens because minimization separates predictions into groups with identical energy and position within each group. Thus, any wrong prediction that has the energy below some correct prediction can occupy many consecutive ranks after minimization (equal to the number of other predictions minimized to the same position) and push the correct one beyond the first 10 limit. If we apply clustering after the minimization (i.e., take just 1 representative match for each local minimum) without subsequent rescoring (Table I, row 4), we still have the prediction quality similar to that in rows 1 and 2, although the average rank of correct prediction improves. Either minimization with clustering and rescoring, or application of full-target function can independently improve the performance to 11 or 12 correctly docked complexes correspondingly (rows 5 and 6). Finally, application of the full procedure increases the number of successfully predicted complexes to 15 (row 7). These results show that all stages of the refinement protocol are important. One can also conclude that the refinement actually works as a filter: It improves the ensemble of the initial FFT predictions by excluding duplicate minima and rescoring with the better estimate of the energy function, without significant changes in geometry.

Conformations of a Small Number of Side-Chains Are Important for Docking

The assumption that the backbone changes upon binding are small so that the induced surface fit can be achieved by optimizing the surface side-chain conformations is an important docking approximation. This assumption is correct for most of the benchmark complexes. Although this approximation drastically narrows the conformational space that needs to be explored in docking, the flexible treatment of all surface side-chains during the exhaustive global search is still prohibitively expensive computationally. Thus, it is important to investigate whether it is possible to optimize just a small subset of surface side-chains and still solve the unbound docking problem in most of the cases. A recent study[11] focused attention on interface residues with the largest solvent accessible surface loss upon binding, demonstrating that such residues in solvated protein often maintain conformation close to the one in a bound structure. On the other hand, in our study, we concentrated on interface residues with the largest side-chain RMSD between their bound and unbound conformations. The goal was to determine how the replacement of a limited number of such side-chains with their bound conformations affects the performance of our rigid-body docking procedure on the benchmark of the unbound protein structures. The RMSD of a side-chain was calculated after superimposing the N, C, C^{α}, and C^{β} atoms in the same residue taken from bound and unbound protein conformations. All the interface residues were sorted by this side-chain RMSD. N side-chains with the largest RMSD were selected and replaced by their bound conformers. The side-chain replacement was done for both receptor and ligand. The value of N was set to 1, 3, 5, and to all interface residues. We ran the benchmarking protocol for each pair of generated structures. The results are summarized in Figure 2. The number of successfully docked complexes significantly increases when only 3 side-chains are brought into the bound conformation [Fig. 2(a)]. There is no improvement with a larger number of replaced side-chains, and the number of correctly docked complexes still does not reach the one obtained for the true bound structures (27 complexes, see above). Therefore, one can conclude that for $N > 3$, the backbone conformational change (interface receptor and ligand C^{α} RMSD = 0.57 Å and 0.45 Å, correspondingly; see above) becomes the limiting factor for the unbound rigid-body docking. It is worth noting that although the number of correctly docked complexes does not increase for $N > 3$, the average RMSD of the first good prediction continues to improve gradually [Fig. 2(b)].

To further test this approach, we applied it to the CAPRI targets once the bound structures were released. We selected and replaced the coordinates of *3 side-chains* on each receptor and ligand according to the above algorithm

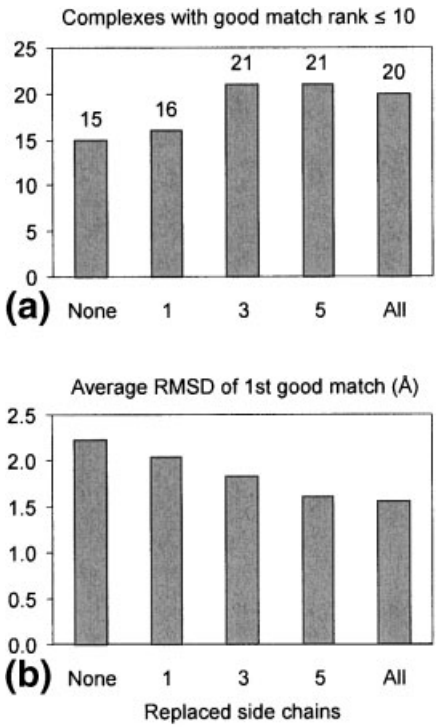

Fig. 2. The significance of side-chains in docking. The total number of complexes is 31. The unbound interface side-chains conformers with the largest bound–unbound RMSD were replaced by the bound ones (see text).

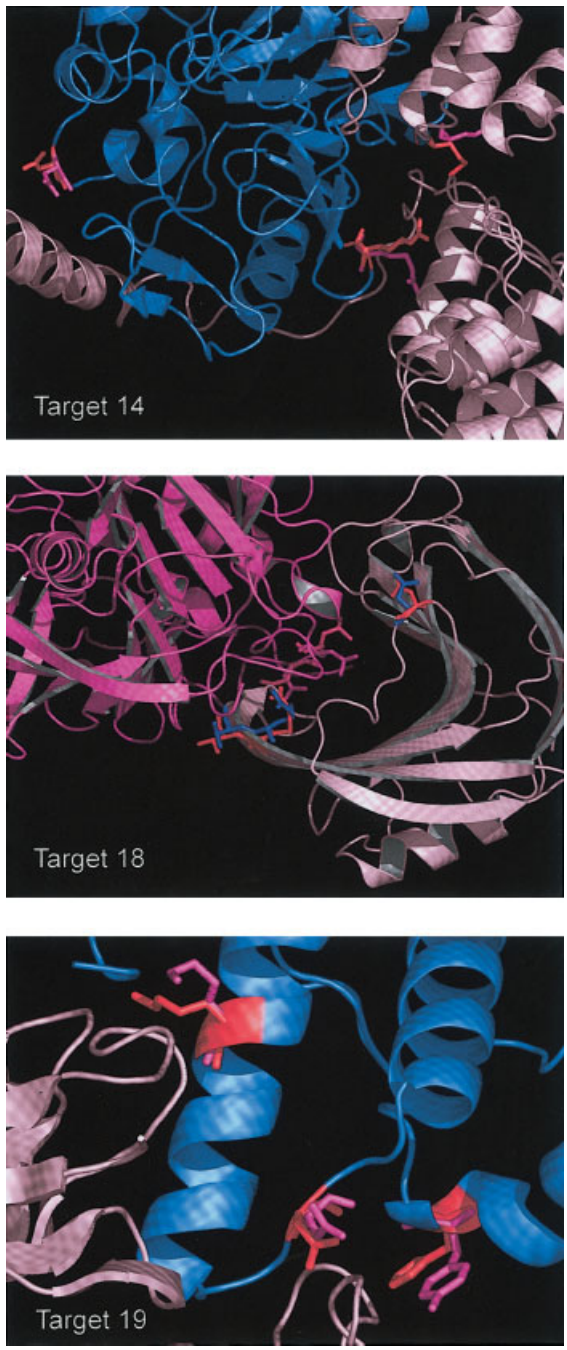

Fig. 3. Side-chains with replaced unbound conformations in CAPRI targets. Bound and unbound side-chains are shown as stick rendering (unbound are in red).

(Fig. 3). Compared with the docking of unbound structures, the docking of the modified structures changed the rank of the first good match: for Target 14, from 8 to 1; for Target 18, from 3 to 54; and for Target 19, from 138 to 5. Although the result for Target 18 deteriorated, the effect on Targets 14 and 19 was encouraging, especially for Target 19, where for the original unbound structure, we failed to make a good prediction with rank below 10.

The results of side-chain replacement simulation on the benchmark set have clear implications for the development of the docking methodology. They suggest that in addition to the rigid-body optimization, it may be enough to optimize the conformations of just 3 interface side-chains. Of course, unlike in our simulated test, in predictive docking experiments we do not know in advance the interface side-chains and the extent of their conformational change upon binding. The conformational flexibility may be estimated by rotamer library analysis or force field simulation (Monte Carlo or molecular dynamics). The number of the side-chains to optimize may be selected in proportion to the size of the protein, so that there would be at least 3 optimized side-chains on an average putative interface. However, if the binding site prediction methods are used prior to the docking, the choice of flexible residues can be restricted to the predicted binding areas.

CONCLUSIONS

Unsupervised rigid-body docking with soft potential and refinement stage was able to solve ~50% of complexes in a benchmark set and 2 out of 3 in CAPRI Round 5 (with the correct interfaces predicted in the remaining target).

Tests on a set of complexes suggest that, on average, only 3 interface side-chains on each protein must be optimized to achieve a significant improvement in unbound docking. Beyond that, the backbone conformational changes become the limiting factor.

REFERENCES

1. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
2. Vakser IA, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
3. Vakser IA. Protein docking for low-resolution structures. Protein Eng 1995;8:371–377.
4. Marshall GR, Vakser IA. Protein–protein docking methods. In: Waksman G, editor. Proteomics and protein–protein interaction: biology, chemistry, bio-informatics, and drug design. Kluwer Academic Press; 2005. Forthcoming.
5. Vajda S, Camacho CJ. Protein–protein docking: is the glass half-full or half-empty? Trends Biotechnol 2004;22:110–116.
6. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
7. Vakser IA. Low-resolution docking: prediction of complexes for underdetermined structures. Biopolymers 1996;39:455–464.
8. Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin–antibody complex. Proteins 1997; Suppl.1:226–230.
9. Tovchigrechko A, Wells CA, Vakser IA. Docking of protein models. Protein Sci 2002;11:1888–1896.
10. Li X, Keskin O, Ma B, Nussinov R, Liang J. Protein–protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J Mol Biol 2004;344:781–795.
11. Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein–protein interactions. Proc Natl Acad Sci USA 2004;101:11287–11292.
12. Vakser IA. Long-distance potentials: an approach to the multiple-minima problem in ligand–receptor interaction. Protein Eng 1996;9:37–41.
13. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.
14. Tovchigrechko A, Vakser IA. How common is the funnel-like energy landscape in protein–protein interactions? Protein Sci 2001;10:1572–1583.
15. Jiang S, Tovchigrechko A, Vakser IA. The role of geometric complementarity in secondary structure packing: a systematic docking study. Protein Sci 2003;12:1646–1651.
16. Schafer H, Van Gunsteren WF, Mark AE. Estimating relative free energies from a single ensemble: hydration free energies. J Comput Chem 1999;20:1604–1617.
17. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003;19:163–164.
18. Glaser F, Steinberg D, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins 2001;43:89–102.
19. Chen R, Mintseris J, Janin J, Weng Z. A protein–protein docking benchmark. Proteins 2003;52:88–91.