ELSEVIER

Exploring the potential of global protein–protein docking: an overview and critical assessment of current programs for automatic *ab initio* docking

Sheng-You Huang[1,2,3]

[1] Research Support Computing, University of Missouri, Columbia, MO 65211, United States
[2] University of Missouri Bioinformatics Consortium, University of Missouri, Columbia, MO 65211, United States
[3] Department of Computer Science, University of Missouri, Columbia, MO 65211, United States

Protein–protein docking is an important computational tool for studying protein–protein interactions. A variety of docking programs with different sampling algorithms and scoring functions as well as computational efficiencies have been subsequently developed over the last decades. Here, we have reviewed the trend and performance of current global docking programs through a comprehensive assessment of the 18 docking/scoring protocols of 14 global docking programs on the latest protein docking benchmark 4.0. The effects of docking algorithms, interaction types, and conformational changes on the docking performance were investigated and discussed. The findings are expected to provide a general guideline for the choice of an appropriate docking protocol and offer insights into the optimization and development of docking and scoring algorithms.

Introduction

Targeting protein–protein interactions for drug design by designing small-molecule modulators has received a growing interest in pharmaceutical science [1–5], in which determination of the complex structure between interacting proteins is valuable [6–12]. Given the technical difficulties and high cost in experimentally determining the three-dimensional structures of complexes, protein–protein docking has been an important computational tool to complement experimental methods for the determination of complex structures since the 1970s [13–19]. Due to less information about the binding site in protein–protein docking than in protein–ligand docking, a global search over the six degrees of freedom (i.e. three translational plus three rotational) is often necessary in protein–protein docking to find possible binding solutions, even though a local search focused on a specific site may be implemented if information is known about the binding site [16–19]. For years, a number of protein–protein docking programs with global search have been developed with various sampling algorithms and scoring functions [20–44]. According to the search strategies, current docking programs can be grouped into three broad categories: fast Fourier transform (FFT) or spherical Fourier transform (SFT)-based correlation algorithms, local shape feature matching approaches, and randomized search algorithms [12,16,18].

All of existing global docking programs have been validated on certain test sets and shown successes as a direct docking tool for binding structure prediction and/or an initial docking program for post-docking approaches [12,16,18]. These docking programs offer a wide range of options to users for their docking applications. However, the large number of available docking programs and the differences in the results given by these programs on different benchmarks also make it hard for a non-specialist to choose an appropriate docking protocol for a given purpose. As a community-wide effort, the CAPRI (Critical Assessment of PRedicted Interactions) experiment [45–49] has been launched as a platform for the blind prediction of developed protein–protein docking and scoring algorithms, though its results may involve significant human intervention and consideration for additional biological information [12,18]. In addition, current docking programs vary significantly in their search strategies, scoring functions, and/or molecular representation of the proteins [12]. Since they were first developed, many of the existing docking algorithms have evolved

E-mail address: huangshe@missouri.edu.

http://dx.doi.org/10.1016/j.drudis.2015.03.007

significantly to accommodate the emerging challenges in real applications [45–49]. With more and more experimental complex structures becoming available, the docking benchmark has also considerably changed in both the number and the diversity of complexes [50–55]. Therefore, it would be valuable to know the status of current global docking programs from the performance perspective, especially for post-docking approaches that strongly depend on an automated initial docking program [56–81], which call for an comprehensive evaluation of existing docking programs on the same ground.

Here, we have performed a critical assessment of the 18 different docking/scoring protocols of 14 commonly used global protein–protein docking programs on the 176 targets of the protein–protein docking benchmark 4.0 developed by Weng and colleagues [53]. The effects of docking algorithms, interaction types, and conformational changes on the docking performance were evaluated and discussed. The findings are expected to provide not only a general guideline for the choice of an appropriate docking protocol but also insightful information for the optimization and development of docking/scoring algorithms.

An overview of global docking programs

Global protein–protein docking, which refers to *ab initio* docking here, samples the relative binding orientations or poses between proteins over all six degrees (three translational + three rotational) of freedom without requiring *a priori* information about the binding site. Some state-of-art docking programs like HADDOCK [82] and RosettaDock [83] are also capable of performing global docking at the expensive computational cost, though they may in general be better suited for refinement or site-specific docking with available information about the binding site. Given the high number of degrees of freedom in backbone and side chains, proteins are normally treated as rigid bodies to make a global search feasible within a practical time [18,19]. Nevertheless, the computational cost is still a challenging issue because of the huge number of translations and rotations in six-dimensional space.

Due to the expensive computational cost, the efficiency of the search strategy is a critical factor for developing a global protein–protein docking program. Among different search strategies, FFT (or SFT) correlation method may achieve a good balance between the computational efficiency and the exhaustive global search, in which sampling covers every position of the grid of six-dimensional space [25]. This is especially appreciated when computational resources are limited. Therefore, most of current global docking programs are based on FFT correlation search algorithms, examples of which include FTDock [23], GRAMM [24], MolFit [25–27], DOT [28,29], ZDOCK1.3 [30], ZDOCK2.1 [31], ZDOCK2.3.2 [32,33], ZDOCK3.0.2 [33,34], PIPER [36], and so on, [12], though a few global docking programs like ATTRACT [20], PatchDock [22], and SwarmDock [43] are based on other types of search strategies like randomized search or local shape matching over the whole protein.

For the same reason of computational efficiency, early global protein–protein docking programs like GRAMM and MolFit only employed simple scoring functions like shape complementarity and/or hydrophobic interactions for fast evaluation of binding modes [24–26], and some programs like PatchDock took advantage of geometric hashing for fast surface patch matching [22]. A more sophisticated pairwise method has also been presented in ZDOCK2.1 to make full use of shape complementarity [31]. Electrostatic interactions were first considered in FTDock [23]. Since then, electrostatics have become a routine part in addition to shape complementarity in later global docking programs like MolFit [26], DOT [28], HEX [38], and ATTRACT [20]. Meanwhile, the desolvation effects have also been included in global docking programs like ZDOCK1.3 [30], ZDOCK2.3 [32], FRODOCK [40], and SDOCK [37]. Then, pairwise knowledge-based potentials that take advantage of information of experimental structures were successfully included in the FFT-based docking programs like PIPER [36] and ZDOCK3.0 [34].

As a whole, compared to the search strategies, the scoring functions have more significantly evolved for global protein–protein docking. Examining recent global docking programs reveals two basic trends. One is to use a more sophisticated scoring function consisting of different energy terms while including additional energy terms like knowledge-based potentials [34,36] or modifying existing scoring functions like electrostatic interactions [37] to enable/optimize the interplay between search strategies and scoring functions [12]. The other is to take advantage of advanced software packages such as the 3D convolution library [33] or new hardware technologies such as the graphics processing unit (GPU) [84] to accelerate the search/sampling process. It is expected that current global docking programs will continue to evolve towards involving more advanced search strategies and more sophisticated scoring functions with the development of computer technology and docking algorithms.

Assessment of 14 global docking programs

Fourteen global protein–protein docking programs that are freely available to the academic community have been selected for the present assessment. Table 1 lists the 14 docking programs evaluated in this study, which include ATTRACT [20,21], PatchDock [22], FTDock [23], GRAMM [24], MolFit [25–27], DOT [28,29], ZDOCK1.3 [30], ZDOCK2.1 [31], ZDOCK2.3.2 [32,33], ZDOCK3.0.2 [33,34], PIPER [36], SDOCK [37], HEX [38,39], and FRODOCK [40]. If a program has more than one scoring method, we have evaluated them separately with the naming convention as follows. If X stands for the name of a docking program, X stands for the docking method with the default scoring function, X/G for the method with the geometric or shape complementarity scoring function, X/GH for the method with the geometric + hydrophobic scoring method, and X:LJ for the docking results of X ranked by the Lennard–Jones (LJ) energy component. Thus, the 14 programs result in a total of 18 different docking/scoring protocols (Fig. 1). During the docking calculations, we have tried to use the default docking parameters to keep the original performance of the docking program, unless the docking parameters need a manual input in which case we have tried to adopt the recommended values by the program or the commonly used values. For example, we have manually set STEP = 1.2 Å for MolFit, nrots = 4500 for PIPER (~an Euler angle interval of 15°), docking_r12_range = 60 Å and max_docking_solutions = 5000 for HEX, respectively. In addition, we have used the results corresponding to the first coefficient index for PIPER, given its relatively better performance.

The protein–protein docking benchmark 4.0 by the Weng group [53] was used to evaluate the docking performance. The

TABLE 1

Fourteen global protein–protein docking programs that cover different categories of search algorithms and scoring functions. Some programs like FTDock and MolFit include more than one scoring scheme, which will be assessed separately, leading to a total of 18 different docking/scoring protocols

Program	Version[a]	Search algorithm	Scoring function	Refs
ATTRACT	–	Randomized search	LJ-type effective potentials and electrostatics	[20]
PatchDock	β 1.3	Local shape match	Geometric shape complementarity	[22]
FTDock	–	FFT-based correlation	Shape complementarity and electrostatic interactions	[23]
GRAMM	1.03	FFT-based correlation	Shape and hydrophobic match	[24]
MolFit	2	FFT-based correlation	Geometric complementarity, hydrophobic complementarity, and electrostatic interactions	[25–27]
DOT	2.01	FFT-based correlation	van der Waals and electrostatic energies	[28,29]
ZDOCK 1.3	–	FFT-based correlation	Shape Complementarity, desolvation, and electrostatics	[30]
ZDOCK 2.1	–	FFT-based correlation	Pairwise shape complementarity	[31]
ZDOCK 2.3.2	–	FFT-based correlation	Pairwise shape complementarity, desolvation, and electrostatics	[32,33]
ZDOCK 3.0.2	–	FFT-based correlation	Shape complementarity, electrostatics, and knowledge-based pair potentials	[33,34]
PIPER	–	FFT-based correlation	Shape complementarity, electrostatic interactions, and knowledge-based pair potentials	[36]
SDOCK[b]	1p0	FFT-based correlation	van der Waals attractive potential, geometric collision, electrostatic potential, and desolvation energy	[37]
HEX	6.3	SFT-based correlation	Surface complementarity and electrostatics	[38]
FRODOCK	1.04	SFT-based correlation	van der Waals, electrostatics, and desolvation	[40]

[a] The version numbers are listed if available, or they are the latest packages obtained from their official web sites or their authors as of August, 2013 or later.
[b] The weighting coefficients for the scoring function of SDOCK have been optimized on the protein docking benchmark 4.0.

benchmark has a total of 176 diverse targets including 52 enzyme-inhibitor, 25 antibody–antigen, and 99 cases with other function. Considering that many protein–protein docking programs are not able to handle the interactions with ligands, we have removed the cofactors from the structures. Residues on the terminus of the unbound structure were also removed if they do not present in the corresponding bound conformation to make the bound and unbound structures comparable in terms of sequence. To minimize

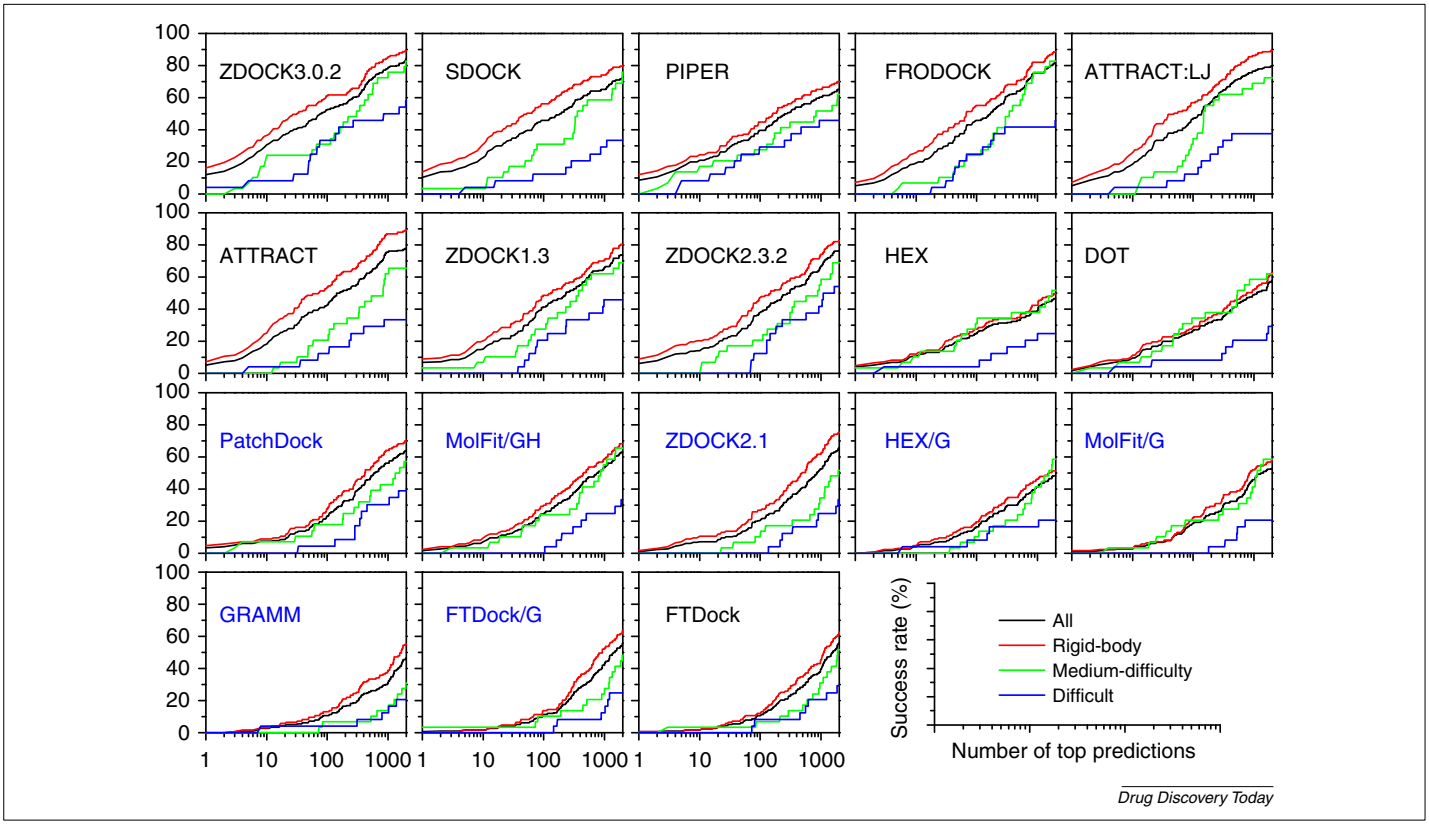

FIGURE 1

The success rates as a function of the number of top predictions for unbound docking on all 176 targets, 123 rigid-body, 29 medium-difficulty, and 24 difficult cases in the benchmark 4.0. The docking methods without electrostatics are labeled in blue font. The docking methods are ordered by top 10 success rate. The legend applies to all the panels of the figure.

the effect of those conformational changes far from the interface on docking results, all the unbound structures were superimposed onto the corresponding bound structures based on the interface residues before docking calculations. The ligand protein was also randomly rotated and translated before docking to remove the possible effect of initial binding interface on the docking results.

In the assessment, a binding mode is defined as a successful prediction, that is, a hit, if its ligand RMSD from the bound structure is less than 10 Å [12,36]. Because the present benchmark includes some complexes from the training sets of PIPER and ZRANK, we have examined the possible bias of the docking performance towards PIPER and ZDOCK3.0.2 to ensure an impartial assessment (Fig. S1 in the supplementary material). Given the realistic feature of unbound structures, the assessment in this study is based on the results of unbound docking, though the results of bound docking may also be discussed in the text for reference.

The overall performance

Figure 1 shows the success rates as a function of the number of the top predictions for the 18 docking/scoring protocols of 14 protein–protein docking programs using the unbound structures of the benchmark. The corresponding results for several numbers of top predictions are shown in Table 2 and Fig. 2. It can be seen from Table 2 and Fig. 2 that overall ZDOCK3.0.2 gave the best performance with the success rates of 11.9%, 30.7%, and 52.3% when the top 1, 10, and 100 predictions were considered, followed by SDOCK, PIPER, FRODOCK, and ATTRACT:LJ, and FTDock yielded the lowest performance. If more than 10 predictions can be inspected/evaluated, which is often the case for CAPRI experiments and post-docking approaches, ZDOCK3.0.2, ATTRACT, and FRODOCK are all satisfactory candidates as an initial-stage docking program, given their relative high success rates for top 100 and 1000 predictions (Table 2). The success rates may reach as high as 84.1% for ZDOCK3.0.2, 79.6% for ATTRACT:LJ, and 81.8% for FRODOCK, compared to 46.6% for HEX and GRAMM when the top 2000 predictions were considered. Therefore, all the docking programs would be able to obtain correct predictions for ~50% or higher of the targets in realistic applications like CAPRI if information about the binding site or an ideal post-docking approach was available to process a few thousand (e.g. 2000) predictions. For reference, we have also performed docking calculations using the bound structures of the benchmark, in which ATTRACT:LJ, SDOCK, and, ZDOCK3.0.2 obtained a higher success rate than the other 15 docking/scoring methods (Fig. S2 and Table S1 in the supplementary material). One common feature in the unbound and bound docking results is that the docking performances of different docking programs tend to be more comparable when more predictions (e.g. 2000) are considered, indicating that the differences of the success rates for top predictions (e.g. 10) may come from the scoring function more than from the sampling algorithm.

TABLE 2

The success rates (%) on the 176 targets of the benchmark when the top 1, 10, 100, 1000, and 2000 predictions were considered. The docking methods are ordered by top 10 success rate

Method	Top 1	Top 10	Top 100	Top 1000	Top 2000
ZDOCK3.0.2	11.93	30.68	52.27	78.98	84.09
SDOCK	10.23	22.73	46.02	65.34	73.30
PIPER	8.52	21.02	39.77	60.23	65.34
FRODOCK	5.11	19.32	46.02	75.57	81.82
ATTRACT:LJ	5.11	18.75	47.16	76.70	79.55
ATTRACT	5.11	18.18	42.61	75.57	77.84
ZDOCK1.3	6.82	15.34	41.48	65.91	73.86
ZDOCK2.3.2	6.25	14.21	38.07	67.05	76.70
HEX	3.98	10.79	25.00	39.20	46.59
DOT	1.71	9.66	26.70	48.86	57.95
PatchDock[a]	3.45	7.47	22.99	56.32	63.79
MolFit/GH[a]	1.71	7.39	25.00	53.41	63.07
ZDOCK2.1[a]	1.14	7.39	20.45	52.27	65.91
HEX/G[a]	0.00	3.98	16.48	41.48	48.30
MolFit/G[a]	1.14	2.84	18.75	46.02	52.84
GRAMM[a]	0.00	2.84	10.79	30.68	46.59
FTDock/G[a]	0.57	1.71	11.36	42.61	56.25
FTDock	0.57	1.71	10.79	39.77	56.25

[a] These methods do not include electrostatics in their scoring functions for docking.

To investigate the effect of interaction types, Fig. 2a also shows the success rates for 52 Enzyme/Inhibitor (EI), 25 Antibody/Antigen (AA), and 99 other-type (OT) targets of the benchmark. It can be seen from the figure that overall the EI targets yield the best performance and perform well for all the docking/scoring methods because of their relatively rigid complementary interface formed through main-chain-main-chain interactions, as expected. The AA targets give the lowest success rates because they do not necessarily form a good binding interface and may experience significant conformational changes upon binding given their side-chain-main-chain interaction mechanism [15]. The effects of interaction types are docking method-dependent. Some docking/scoring protocols like PIPER, HEX, PatchDock, HEX/G, MolFit/G, GRAMM, and FTDock perform especially worse on AA targets. Both FTDock/G and FTDock fail to predict a hit in the top 10 predictions on OT targets. Some methods like ZDOCK3.0.2, FRODOCK, ATTRACT:LJ, ATTRACT, and ZDOCK2.3.2 seem to be more robust and perform relatively well on all three interaction-types of targets (Fig. 2a). Similar trends can also be observed in the corresponding success rates for bound docking (Fig. S2 in the supplementary material).

Effect of scoring functions

From Table 2 and Fig. 2, one can find that overall the docking/scoring methods including electrostatics like ZDOCK3.0.2, SDOCK, PIPER, and FRODOCK perform better than those without electrostatics like PatchDock, MolFit, and ZDOCK2.1, demonstrating the importance of electrostatics in protein–protein interactions. For the docking/scoring protocols without considering electrostatics, overall PatchDock obtained a slightly higher performance with the success rates of 3.5%, 7.5%, and 23.0%, which tie with 1.7%, 7.4%, and 25.0% for MolFit/GH and are closely followed by 1.1%, 7.4%, and 20.5% for ZDOCK2.1, compared to 0.6%, 1.7%, and 10.8% for FTDock when the top 1, 10, and 100 predictions were considered (Table 2). Further examining those docking/scoring methods including electrostatics reveals that the docking/scoring programs including desolvation effects like ZDOCK3.0.2, SDOCK, and FRODOCK tend to perform better than those without considering desolvation like FTDock, HEX, and DOT, suggesting the role of desolvation in protein–protein interactions. Therefore, future scoring functions are expected to include more types of energy terms for a better performance.

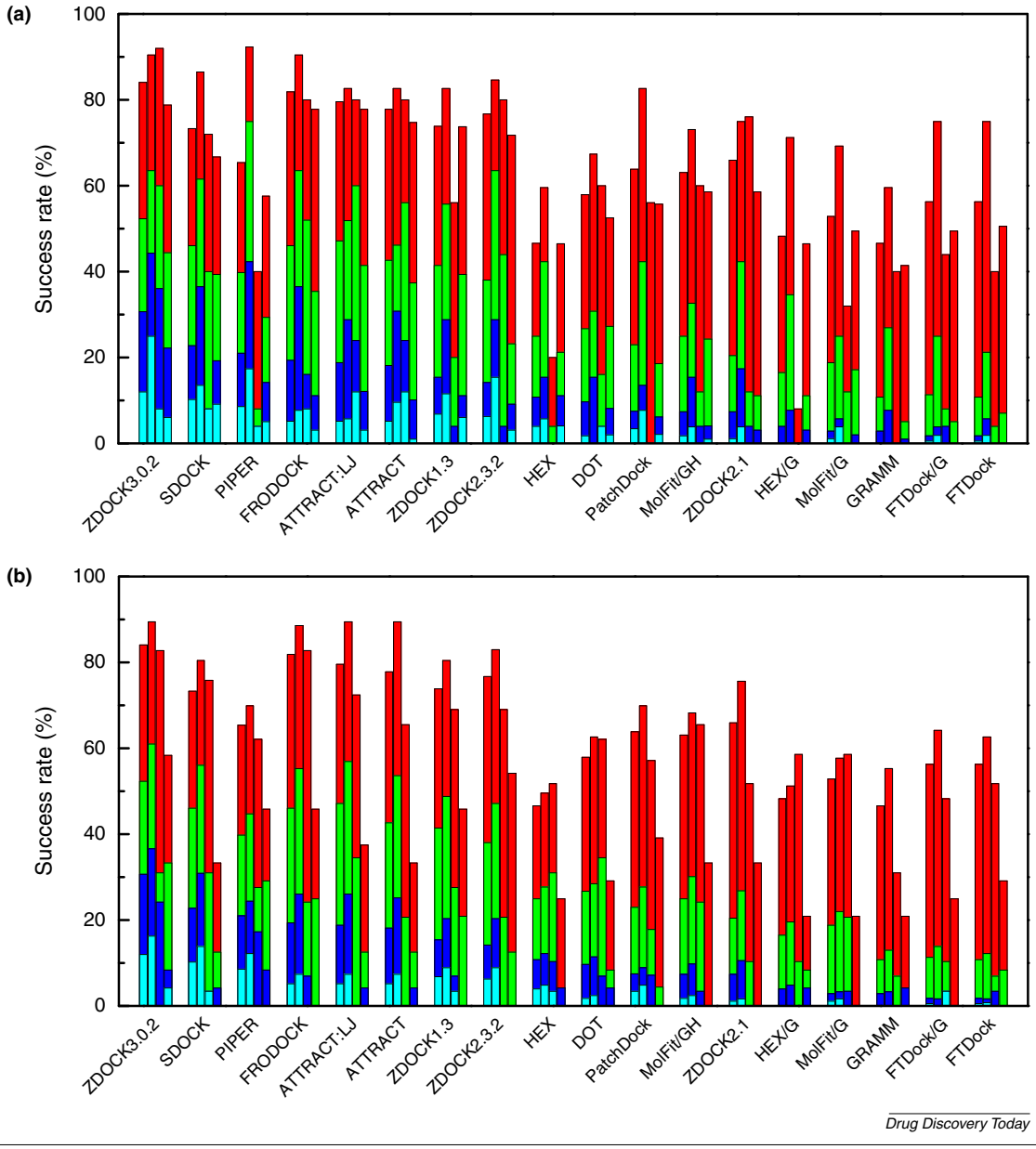

FIGURE 2

The histograms of the success rates for unbound docking on the benchmark 4.0 when the top 1 (cyan), 10 (blue), 100 (green), and 2000 (red) predictions were considered. The docking methods are ordered by top 10 success rate. (a) From left to right for each docking method are for all 176, 52 enzyme/inhibitor, 25 antibody/antigen, and 99 other-type targets, respectively. (b) From left to right for each docking method are for all 176, 123 rigid-body, 29 medium-difficulty, and 24 difficult targets, respectively.

Several additional notable features can also be observed from Fig. 2 and Table 2. One interesting feature is that the results ranked by the LJ-type energies only perform significantly better than those ranked by the LJ + eletrostatics scores in ATTRACT (Fig. 2), even though the binding modes are sampled based on the default LJ + eletrostatics scoring function [20], suggesting a possible scoring strategy for protein docking. In addition, although the geometry + eletrostatics scoring method is normally expected to outperform the geometry scoring function in the same docking program, FTDock is the only one that yielded the opposite performance among tested docking programs (Fig. 2), indicating that the sampling/scoring protocol in FTDock may need further improvement. It may also partly explain the low success rates of FTDock when a small number of top predictions were considered, even though it gave a fair performance with a success rate of 56.25% for top 2000 predictions that may be due to the robustness of FFT-based search algorithms (Table 2). One more notable feature is that two of the three top-performed programs in unbound docking, ZDOCK3.0.2 and PIPER, both include knowledge-based pair potentials in their scoring functions, indicating the efficacy of knowledge-based potentials.

Impact of conformational changes

To investigate the effects of conformational changes, we have calculated the success rates of 18 docking/scoring protocols on three categories of 123 rigid-body, 29 medium difficulty and 24

difficult cases in the benchmark, as shown in Fig. 1. It can be seen from the figure that overall the docking programs perform the best on the rigid-body targets and the worst on the difficult targets, as expected (Fig. 2b). Figure 1 also shows that the top-performed docking programs like ZDOCK3.0.2 and SDOCK obtained much more improvement than the less-satisfactory methods like HEX and DOT on the rigid-body category, compared to the medium-difficulty and difficult categories (Fig. 1). For example, ZDOCK3.0.2 yielded the success rates of 16.3% and 36.6% for the rigid-body category, which are significantly higher than 0.0% and 24.1% for the medium-difficulty category and 4.2% and 8.3% for the difficult category when the top 1 and 10 predictions were considered. In contrast, HEX gave the similar performances for both the rigid-body and the medium-difficulty categories with the success rates of 12.2% vs. 10.4% and 4.9% vs. 3.5% when the top 1 and 10 predictions were considered. These suggest that a well-optimized rigid docking algorithm has a high potential to consider small conformational changes in rigid-body and some medium-difficulty cases, although how to incorporate large conformational changes in the difficult cases is still a challenge for all the docking programs.

Lessons may be learned from the significantly better performance on the rigid-body category achieved by those top-performed docking approaches. By convention, rigid-body cases are those targets that have an interface RMSD of <1.4 Å and a fraction of non-native residue of <0.4 between the unbound and the bound structures, superposed onto each other [53]. The satisfactory performance on such rigid-body cases indicates that some docking algorithms like ZDOCK3.0.2 are robust to incorporate small conformational changes at the binding interface. Given the variety of docking and scoring algorithms in the top-performed docking programs, it may be learned that in addition to using an appropriate grid representation or reduced model for the protein, effort may also be made on scoring functions to better handle small conformational changes upon binding by including energy terms like knowledge-based potentials in ZDOCK3.0.2 [33,34] and PIPER [36], stepwise potentials in SDOCK, and effective soft pair potentials in ATTRACT [20]. Despite the encouraging performance on rigid-body targets, many protocols like FRODOCK, ZDOCK2.3.2, PatchDock, MolFit, ZDOCK2.1, and FTDock fail to give any hit on difficulty targets in the top 10 predictions and some methods like ATTRACT, ZDOCK2.3.2, ZDOCK2.1, HEX/G, GRAMM also have the similar failure on medium-difficulty targets (Fig. 2). In addition, the overall docking performance of current docking algorithms is still not high for unbound docking with the maximum success rate of 30.7% for top 10 predictions (Table 2). Therefore, there is much room to improve the performance of current docking algorithms by developing more sophisticated docking algorithm and scoring functions as well as protein flexibility consideration.

Target-dependent performance

In the above, we have assessed the overall performance of 18 docking/scoring protocols on the complete benchmark, which is an important measurement of the robustness for a docking/scoring method. However, the docking performance on single targets may be a more practical reference when it comes to a real application that normally involves one or a few cases. Therefore, the target-dependent performance has similar importance weight to the overall performance. Tables S2 and S3 in the supplementary material list the ranks of first-predicted hits by the 18 docking/scoring methods for unbound and bound docking on the 176 targets in the benchmark 4.0, respectively. Here, the rank may be regarded as a rough measure of how difficult a target is in terms of docking. The higher rank indicates a more difficult target.

Several notable features can be observed from Table S2. First, the docking performance is significantly target-dependent and no docking method can perform well on all the cases. The overall top-performed programs do not necessarily perform better than those approaches with an overall lower success rate on a given target (Table S2). For example, the top docking method ZDOCK3.0.2 failed on some targets like 1DR6, 1QA9, 1SBB, and 2HQS, while the overall worst-performed FTDock obtained successful hits within the considered predictions on these cases. Therefore, although those overall top-performed methods appear to be a safe general choice for protein docking, several other programs may be also needed to provide alternative predictions due to the significantly target-dependent performances of existing docking approaches. In addition, some difficult targets like 2I9B and 2IDO may be not as challenging as those medium-difficulty cases like 1BGX and 1M10 and even some rigid-body cases like 1FC2 and 1GHQ (Table S2 in the supplementary material), suggesting the target-dependent effect of conformational changes.

Examining Tables S2 and S3 also shows that some targets tend to be difficult cases in terms of docking with high ranks for both bound and unbound docking for many of the docking methods. Given no conformational changes in the input structures for bound docking, there must be something other than flexibility that determines the difficulty of a target in terms of docking. Closely inspecting the accessible surface areas (ASAs) of the targets reveals that those generally difficult cases often have a small relative change of accessible surface areas upon binding (rΔASA), which is defined as

$$\mathrm{r}\Delta\mathrm{ASA} = \frac{\Delta\mathrm{ASA}}{\mathrm{ASA_R} + \mathrm{ASA_L}} \times 100\%$$

where ΔASA is the change of the ASAs for the receptor and ligand proteins upon binding, and $\mathrm{ASA_R}$ and $\mathrm{ASA_L}$ are the ASAs of the individual receptor and ligand proteins before binding, respectively. That means that the relative ΔASA (rΔASA) may be a common factor that limits the docking performance of a target for both bound and unbound docking. Figure 3 shows the relationship between the average ranks of all the 18 docking/scoring protocols for bound and unbound docking on the 176 targets. It can be seen from the figure that there is a significant correlation between the average ranks of bound and unbound docking with a Pearson correlation coefficient of 0.586. The points at the upper-right corner that represent difficult cases indeed have a small rΔASA as shown by their size and color (Fig. 3). This confirms that the rΔASA is an important factor affecting all the docking programs. As shown in Fig. 4, the ranks for all the docking protocols except ATTRACT for bound docking, show a negative correlation with the rΔASAs on the benchmark, although an ideal docking algorithm should have a non-significant correlation on this relationship. This common limitation for all the docking programs may be understood because the smaller rΔASA will lead to more possibilities of binding

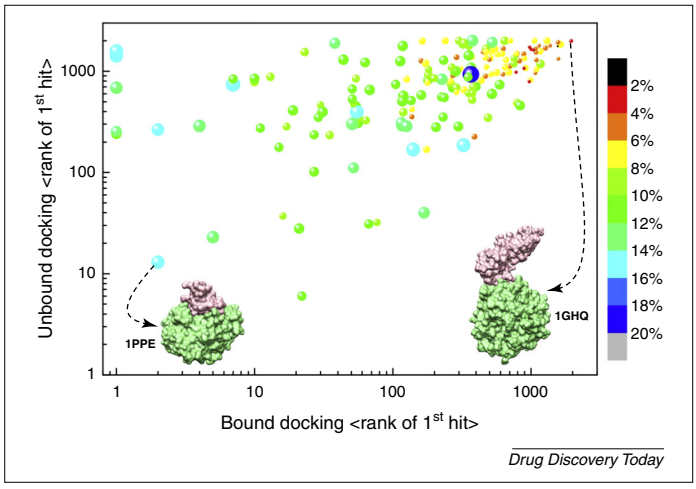

FIGURE 3

The relationship between the average ranks of the first hits for unbound and bound docking, in which the points representing the 176 targets are colored and sized according to their relative changes of accessible surface areas (rΔASAs) upon binding. Two rigid-body cases (1PPE and 1GHQ), which show very different ranks/difficulties in terms of docking, are shown as examples. The values of the average ranks and rΔASAs are listed in Tables S2 and S3 of the supplementary material.

interfaces. In such cases, an accurate scoring function is crucially needed to distinguish the large number of false positives from true binding modes.

Computational efficiency

In addition to the docking performance, another notable feature of a docking algorithm is its computational efficiency. That is, how fast a program can finish a docking run against a pair of protein structures. This is an important factor especially when the computational resources are limited or docking runs are needed for a large number of proteins. Figure 5 shows the average running times of the 18 docking/scoring methods for docking a pair of protein structures on the benchmark of 176 targets. All the calculations were done on a single Intel(R) Xeon(R) CPU X5650 @ 2.67 GHz core on a Linux x86_64 cluster.

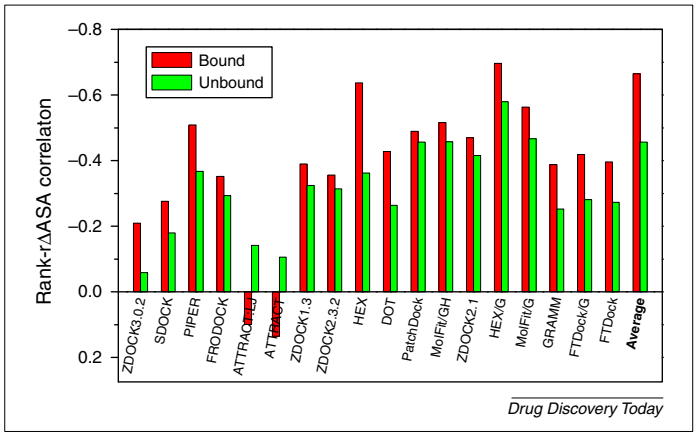

FIGURE 4

The correlations between the ranks of the first hits and the relative changes of accessible surface areas (rΔASAs) upon binding for bound and unbound docking for the 18 docking/scoring protocols on the 176 targets of the benchmark.

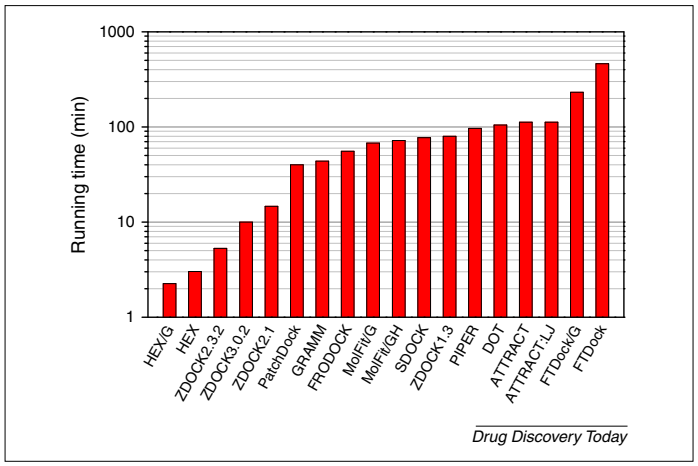

FIGURE 5

The average running time for a protein–protein docking run averaging over the 176 targets of the benchmark 4.0.

It can be seen from Fig. 5 that HEX has the highest computational efficiency and only took an average of 2.3 and 3.0 min for a docking run using the geometry scoring method and the default geometry + eletrostatics scoring function, respectively, followed by 5.3 min for ZDOCK2.3.2, 10.0 min for ZDOCK3.0.2, and 14.7 min for ZDOCK2.1. The high computational efficiency of HEX and ZDOCK may be understood because HEX uses spherical polar Fourier (SPF) correlations to accelerate the calculations [38,39] and ZDOCK has recently implemented an advanced 3D convolution library to accelerate the computation in its new versions of 2.3.2 and 3.0.2 [33]. Among the evaluated docking programs, FTDock took the longest time for a docking run with 231 min for the geometry scoring only and 461 min for the default geometry + eletrostatics scoring scheme. The unusual low computational efficiency of FTDock is partly due to its finer grid with the default spacing of 0.875 Å used for docking [23]. With the development of computing hardware and software technologies, the computational efficiency of existing docking programs will become much higher and is expected to finish a rigid-docking run within seconds on a single cpu.

Concluding remarks

In summary, we have reviewed the advancement of global protein–protein docking and performed a comprehensive assessment of the 18 docking/scoring protocols of 14 global docking programs on the protein docking benchmark 4.0. It was found that the docking performance is significantly target-dependent and no program is able to perform well on all the cases of the benchmark, though some algorithms are more robust than others. The difference of the docking performances in top predictions may come from the scoring function more than from the sampling algorithm. The relative change of accessible surface areas (rΔASA) upon binding is a common factor limiting the docking performance of all docking programs and may be used to classify the difficulties of targets in terms of docking. Overall, ZDOCK3.0.2, SDOCK, and PIPER yielded the relatively higher success rates with 30.7%, 22.7%, and 21.0% for top 10 predictions, while ZDOCK3.0.2, ATTRACT, and FRODOCK are all good choices if more predictions can be evaluated which is often the case in a post-docking

approach. It was also found that the hydrophobic effect plays an important role in docking. The desolvation is an important energy term in the scoring function and knowledge-based potentials are effective for realistic unbound docking. Regarding the impact of search strategies, the Spherical Polar Fourier correlation method like HEX has a higher computational efficiency, and FFT-based algorithms like ZDOCK, SDOCK and PIPER yielded a better performance in the docking performance. The good performance of FFT-based methods like ZDOCK, PIPER, and SDOCK suggests the robustness of FFT-based docking sampling algorithms. Despite the reasonable success rates of current docking programs, there is much room to improve the docking performance by ways like optimizing and/or compounding existing sampling and scoring strategies. The current docking methods can be also computationally efficient and some programs such as HEX and ZDOCK are able to finish a rigid-docking run within minutes, which are fast enough to perform large-scale docking computations on a proteome or sub-proteome scale.

Generally speaking, the top 10 binding modes may be considered if the docking results are directly used in real applications. However, it should be noted that the global rigid-body predictions in this study may not always be used directly as the final results. They are often rescored/refined by post-processing algorithms like using a sophisticated scoring function, considering protein flexibility and/or incorporating available information about the binding interface. According to the present assessment, all the docking programs would succeed for ~50% or higher of the targets in realistic applications like CAPRI if information about the binding site or an ideal post-docking approach was available to process a few thousand predictions. Despite the advantages of using standalone programs for code customization or pipeline development, many of the programs in this study also have their web servers (e.g. GRAMM-X [85], HEX server [86], PatchDock server [87], PIPER/Cluspro [88], and ZDOCK server [35]), which provide a user-friendly web interface and are especially useful for non-expert and low-volume users.

Conflict of interest statement

None declared.

Acknowledgements

The computations for this work were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC). The support of the UMBC is greatly appreciated.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.drudis.2015.03.007.

References

1 Arkin, M.R. and Wells, J.A. (2004) Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* 3, 301–317

2 Kann, M.G. (2007) Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief. Bioinform.* 8, 333–346

3 Schreiber, G. and Fleishman, S.J. (2013) Computational design of protein–protein interactions. *Curr. Opin. Struct. Biol.* 23, 903–910

4 Naveed, H. and Liang, J. (2014) Weakly stable regions and protein–protein interactions in beta-barrel membrane proteins. *Curr. Pharm. Des.* 20, 1268–1273

5 Li, H. *et al.* (2014) Drug design targeting protein–protein interactions (PPIs) using multiple ligand simultaneous docking (MLSD) and drug repositioning: discovery of raloxifene and bazedoxifene as novel inhibitors of IL-6/GP130 interface. *J. Med. Chem.* 57, 632–641

6 Zhou, H.X. (2014) Theoretical frameworks for multiscale modeling and simulation. *Curr. Opin. Struct. Biol.* 25, 67–76

7 Fuller, J.C. *et al.* (2009) Predicting druggable binding sites at the protein–protein interface. *Drug Discov. Today* 14, 155–161

8 Huang, S.Y. *et al.* (2009) Molecular modeling of the heterodimer of human CFTR's nucleotide-binding domains using a protein–protein docking approach. *J. Mol. Graph. Model.* 27, 822–828

9 Zhang, G. *et al.* (2010) Ion sensing in the RCK1 domain of BK channels. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18700–18705

10 Zhang, L.Q. *et al.* (2008) Interactions between PBEF and oxidative stress proteins – a potential new mechanism underlying PBEF in the pathogenesis of acute lung injury. *FEBS Lett.* 582, 1802–1808

11 Grosdidier, S. and Fernandez-Recio, J. (2012) Protein–protein docking and hot-spot prediction for drug discovery. *Curr. Pharm. Des.* 18, 4607–4618

12 Huang, S.Y. (2014) Search strategies and evaluation in protein–protein docking: principles, advances and challenges. *Drug Discov. Today* 19, 1081–1096

13 Wodak, S.J. and Janin, J. (1978) Computer analysis of protein–protein interaction. *J. Mol. Biol.* 124, 323–342

14 Smith, G.R. and Sternberg, M.J. (2002) Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.* 12, 28–35

15 Halperin, I. *et al.* (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409–443

16 Ritchie, D.W. (2008) Recent progress and future directions in protein–protein docking. *Curr. Protein Pept. Sci.* 9, 1–15

17 Moreira, I.S. *et al.* (2010) Protein–protein docking dealing with the unknown. *J. Comput. Chem.* 31, 317–342

18 Janin, J. (2010) Protein–protein docking tested in blind predictions: the CAPRI experiment. *Mol. Biosyst.* 6, 2351–2362

19 Vajda, S. *et al.* (2013) Sampling and scoring: a marriage made in heaven. *Proteins* 81, 1874–1884

20 Zacharias, M. (2003) Protein–protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* 12, 1271–1282

21 Zacharias, M. (2005) ATTRACT: protein–protein docking in CAPRI using a reduced protein model. *Proteins* 60, 252–256

22 Duhovny, D. *et al.* (2002) Efficient unbound docking of rigid molecules. In *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI)* (Guido, R. and Gusfield, D., eds), In Rome, Italy. *Lect. Notes Comput. Sci.* 2452, 185–200

23 Gabb, H.A. *et al.* (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 272, 106–120

24 Vakser, I.A. (1997) Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins Suppl.* 1, 226–230

25 Katchalski-Katzir, E. *et al.* (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U. S. A.* 89, 2195–2199

26 Heifetz, A. *et al.* (2002) Electrostatics in protein–protein docking. *Protein Sci.* 11, 571–587

27 Berchanski, A. *et al.* (2004) Hydrophobic complementarity in protein–protein docking. *Proteins* 56, 130–142

28 Mandell, J.G. *et al.* (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* 14, 105–113

29 Roberts, V.A. *et al.* (2013) DOT2: macromolecular docking with improved biophysical models. *J. Comput. Chem.* 34, 1743–1758

30 Chen, R. and Weng, Z.P. (2002) Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* 47, 281–294

31 Chen, R. and Weng, Z. (2003) A novel shape complementarity scoring function for protein–protein docking. *Proteins* 51, 397–408

32 Chen, R. *et al.* (2003) ZDOCK: an initial-stage protein docking algorithm. *Proteins* 52, 80–87

33 Pierce, B.G. *et al.* (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* 6, e24657

34 Mintseris, J. *et al.* (2007) Integrating statistical pair potentials into protein complex prediction. *Proteins* 69, 511–520

35 Pierce, B.G. *et al.* (2014) ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* 30, 1771–1773

36 Kozakov, D. *et al.* (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65, 392–406
37 Zhang, C. and Lai, L. (2011) SDOCK: a global protein–protein docking program using stepwise force-field potentials. *J. Comput. Chem.* 32, 2598–2612
38 Ritchie, D.W. and Kemp, G.J. (2000) Protein docking using spherical polar Fourier correlations. *Proteins* 39, 178–194
39 Ritchie, D.W. *et al.* (2008) Accelerating and focusing protein–protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics* 24, 1865–1873
40 Garzon, J.I. *et al.* (2009) FRODOCK: a new approach for fast rotational protein–protein docking. *Bioinformatics* 25, 2544–2551
41 Venkatraman, V. *et al.* (2009) Protein–protein docking using region-based 3D Zernike descriptors. *BMC Bioinform.* 10, 407
42 Li, L. *et al.* (2011) ASPDock: protein–protein docking algorithm using atomic solvation parameters model. *BMC Bioinform.* 12, 36
43 Moal, I.H. and Bates, P.A. (2010) SwarmDock and the use of normal modes in protein–protein docking. *Int. J. Mol. Sci.* 11, 3623–3648
44 Bajaj, C. *et al.* (2011) F2Dock: fast Fourier protein–protein docking. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8, 45–58
45 Janin, J. *et al.* (2003) CAPRI: a critical assessment of predicted interactions. *Proteins* 52, 2–9
46 Mendez, R. *et al.* (2005) Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. *Proteins* 60, 150–169
47 Lensink, M.F. *et al.* (2007) Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins* 69, 704–718
48 Lensink, M.F. and Wodak, S.J. (2010) Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins* 78, 3085–3095
49 Lensink, M.F. and Wodak, S.J. (2013) Docking, scoring and affinity prediction in CAPRI. *Proteins* 81, 2082–2095
50 Chen, R. *et al.* (2003) A protein–protein docking benchmark. *Proteins* 52, 88–91
51 Mintseris, J. *et al.* (2005) Protein–protein docking benchmark 2.0: an update. *Proteins* 60, 214–216
52 Hwang, H. *et al.* (2008) Protein–protein docking benchmark version 3.0. *Proteins* 73, 705–709
53 Hwang, H. *et al.* (2010) Protein–protein docking benchmark version 4.0. *Proteins* 78, 3111–3114
54 Douguet, D. *et al.* (2006) DOCKGROUND resource for studying protein–protein interfaces. *Bioinformatics* 22, 2612–2618
55 Gao, Y. *et al.* (2007) DOCKGROUND system of databases for protein recognition studies: unbound structures for docking. *Proteins* 69, 845–851
56 Huang, S-Y. and Zou, X. (2010) MDockPP: a hierarchical approach for protein–protein docking and its application to CAPRI rounds 15–19. *Proteins* 78, 3096–3103
57 Huang, S-.Y. *et al.* (2013) Inclusion of the orientational entropic effect and low-resolution experimental information for protein–protein docking in CAPRI. *Proteins* 81, 2183–2191
58 Vreven, T. *et al.* (2011) Integrating atom-based and residue-based scoring functions for protein–protein docking. *Protein Sci.* 20, 1576–1586
59 Pierce, B. and Weng, Z. (2007) ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* 67, 1078–1086
60 Pierce, B. and Weng, Z. (2008) A combination of rescoring and refinement significantly improves protein docking performance. *Proteins* 72, 270–279
61 Cheng, T.M. *et al.* (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins* 68, 503–515
62 Liang, S. *et al.* (2007) A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. *Proteins* 69, 244–253
63 Huang, S-.Y. and Zou, X. (2008) An iterative knowledge-based scoring function for protein–protein recognition. *Proteins* 72, 557–579
64 Huang, S.Y. and Springer, G.K. (2013) How the folding funnel depends on size and structure of proteins? A view from the scoring function perspective. *Tsinghua Sci. Technol.* 18, 462–468
65 Li, L. *et al.* (2003) RDOCK: refinement of rigid-body protein docking predictions. *Proteins* 53, 693–707
66 Andrusier, N. *et al.* (2007) FireDock: fast interaction refinement in molecular docking. *Proteins* 69, 139–159
67 Mashiach, E. *et al.* (2010) FiberDock: flexible induced-fit backbone refinement in molecular docking. *Proteins* 78, 1503–1519
68 Venkatraman, V. and Ritchie, D.W. (2012) Flexible protein docking refinement using pose-dependent normal mode analysis. *Proteins* 80, 2262–2274
69 Krol, M. *et al.* (2007) Implicit flexibility in protein docking: cross-docking and local refinement. *Proteins* 69, 750–757
70 Liang, S. *et al.* (2009) Refining near-native protein–protein docking decoys by local re-sampling and energy minimization. *Proteins* 76, 309–316
71 Viswanath, S. *et al.* (2013) Improving ranking of models for protein complexes with side chain modeling and atomic potentials. *Proteins* 81, 592–606
72 Shen, Y. *et al.* (2008) Protein docking by the underestimation of free energy funnels in the space of encounter complexes. *PLoS Comput. Biol.* 4, e1000191
73 Shen, Y. (2013) Improved flexible refinement of protein docking in CAPRI rounds 22–27. *Proteins* 81, 2129–2136
74 Omori, S. and Kitao, A. (2013) CyClus: a fast, comprehensive cylindrical interface approximation clustering/reranking method for rigid-body protein–protein docking decoys. *Proteins* 81, 1005–1016
75 Oliva, R. *et al.* (2013) Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins* 81, 1571–1584
76 Qin, S. and Zhou, H.X. (2010) Selection of near-native poses in CAPRI rounds 13–19. *Proteins* 78, 3166–3173
77 Qin, S. and Zhou, H.X. (2013) Using the concept of transient complex for affinity predictions in CAPRI rounds 20–27 and beyond. *Proteins* 81, 2229–2236
78 Lorenzen, S. and Zhang, Y. (2007) Monte Carlo refinement of rigid-body protein docking structures with backbone displacement and side-chain optimization. *Protein Sci.* 16, 2716–2725
79 Kozakov, D. *et al.* (2008) Discrimination of near-native structures in protein–protein docking by testing the stability of local minima. *Proteins* 72, 993–1004
80 Kowalsman, N. and Eisenstein, M. (2009) Combining interface core and whole interface descriptors in postscan processing of protein–protein docking models. *Proteins* 77, 297–318
81 Xue, L.C. *et al.* (2014) DockRank: ranking docked conformations using partner-specific sequence homology-based protein interface prediction. *Proteins* 82, 250–267
82 Dominguez, C. *et al.* (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737
83 Gray, J.J. *et al.* (2003) Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331, 281–299
84 Ritchie, D.W. and Venkatraman, V. (2010) Ultra-fast FFT protein docking on graphics processors. *Bioinformatics* 26, 2398–2405
85 Tovchigrechko, A. and Vakser, I.A. (2006) GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res.* 34, W310–W314
86 Macindoe, G. *et al.* (2010) HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* 38, W445–W449
87 Schneidman-Duhovny, D. *et al.* (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367
88 Comeau, S.R. *et al.* (2004) ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res.* 32, W96–W99