ZDOCK: An Initial-stage Protein Docking Algorithm

Rong Chen[1], Li Li[1] and Zhiping Weng[1,2]

[1]Bioinformatics Program
[2]Department of Biomedical Engineering
Boston University, 44 Cummington Street, Boston, MA 02215

Corresponding Author: Zhiping Weng (zhiping@bu.edu)

Abstract

The development of scoring functions is of great importance to protein docking. Here we present a new scoring function for the initial stage of unbound docking. It combines our recently developed pairwise shape complementarity with desolvation and electrostatics. We compare this scoring function with three other functions on a large benchmark of 49 non-redundant test cases, and demonstrate its superior performance, especially for the antibody/antigen category of test cases. For 44 test cases (90% of the benchmark), we can retain at least one near-native structures within the top 2000 predictions at the 6° rotational sampling density, with an average of 52 near-native structures per test case. The remaining five difficult test cases can be explained by a combination of poor binding affinity, large backbone conformational changes and our algorithm's strong tendency for identifying large concave binding pockets. All four scoring functions have been integrated into our Fast Fourier Transform based docking algorithm ZDOCK, which is freely available to academic users at http://zlab.bu.edu/~rong/dock.

Keywords:

ZDOCK; protein docking; initial-stage docking; shape complementarity; pairwise shape complementarity; scoring function; binding free energy

Introduction

Protein docking is the prediction of the 3-dimensional (3-D) structure of a protein-protein complex from the coordinates of its component structures. It is classified as bound docking or unbound docking. For the former, a protein complex is pulled apart and re-assembled. For the latter, individually crystallized component structures are used. Unbound docking is of more interest to us and is the focus of this work. It has long been recognized that proteins undergo conformational changes upon binding, especially their surface side chains. This complicates unbound docking tremendously. With current computing power, it is infeasible to perform extensive conformational searches during docking, unless the binding site is known. Thus, a number of groups have adopted the two-stage approach[1,2]: in the *initial* stage, the receptor and ligand are treated as rigid bodies and the 6-D rotational and translational degrees of freedom are fully explored with scoring functions that are tolerant to conformational changes[3-10] ; in the *refinement* stage, a small number (tens to thousands) of structures obtained in the initial stage is refined and re-ranked using more detailed energy functions that take into account conformational

changes[11-15]. Frequently, conformational searches using side-chain rotamers and energy minimizations are performed in the refinement stage.

In this paper, we focus on the initial-stage of unbound docking. A number of algorithms have been developed for this goal, described in several reviews[1,2,16-20]. FTDock searches the grid-based shape complementarity (GSC) and electrostatics using a Fast Fourier Transform (FFT) algorithm[3]. DOT is another FFT based method that computes Poisson-Boltzmann electrostatics[4]. HEX evaluates overlapping surface skins and electrostatic complementarity with Fourier correlation[5]. GRAMM focuses on low resolution docking, evaluating GSC with FFT[6]. PPD matches critical points using geometric hashing[7]. BiGGER searches maximal surface mapping and favorable amino acid contacts using a bit-mapping method[8]. DARWIN[9] calculates molecular mechanics energies defined according to CHARMM[21] using a genetic algorithm.

We have developed an initial-stage docking algorithm called ZDOCK[10], which optimizes desolvation, GSC and electrostatics using FFT. A layer of grid points that surround the receptor is identified, and the total number of grid points in this layer that overlap any grid points corresponding to ligand atoms, minus a clash penalty, is the GSC score. We showed that the desolvation component of our scoring function was the key to ZDOCK's competitive performance, compared with several other algorithms with a similar goal[10]. Subsequently, we discovered a novel pairwise shape complementarity function (PSC), which computes the total number of receptor-ligand atom pairs within a distance cutoff, minus a clash penalty. When tested on a benchmark with 49 non-redundant test cases[22], PSC consistently identified more near-native structures and ranked them higher than GSC, and this superior performance was observed across all classes of complexes and at all prediction levels[23].

In this paper, we integrate PSC with desolvation (DE) and electrostatics (ELEC) to create a much more powerful scoring function PSC+DE+ELEC. The resulting scoring function is tested on the same benchmark[22] and proves superior to PSC alone[23] and the GSC+DE+ELEC scoring function in our previous study[10]. For 44 test cases (90% of the benchmark), ZDOCK with PSC+DE+ELEC can retain at least one near-native prediction (also called a hit) within the top 2000 predictions at a rotational sampling interval of 6°, with an average of 52 hits per test case. The improvement of PSC+DE+ELEC over GSC+DE+ELEC is most apparent in the antibody/antigen category of test cases, with the former producing more hits and better rankings for hits on practically all test cases. We also carefully examine the five test cases which ZDOCK has difficulty with, and discuss the potential applications of different scoring functions for the initial stage of unbound docking.

Scoring Functions

The basic search algorithm of ZDOCK has been described in detail[10]. In this paper, we focus on the comparison of different scoring functions. Our goal is to identify the scoring function that performs best for the initial stage of unbound docking, which entails ranking as many near-native structures as possible in the top few thousand predictions. We consider four scoring functions: combining grid-based shape complementarity GSC with desolvation and electrostatics (GSC+DE+ELEC), pairwise shape complementarity (PSC), combining PSC with desolvation (PSC+DE), combining PSC with desolvation and electrostatics (PSC+DE+ELEC). Two of these

target functions have been described previously: GSC+DE+ELEC[10] and PSC[23]. The remaining two, PSC+DE and PSC+DE+ELEC, are described as follows.

PSC+DE

We use the Atomic Contact Energy (ACE)[24] to estimate desolvation (DE). ACE is defined as the free energy change of breaking two protein-atom/water contacts and forming a protein-atom/protein-atom contact and a water-water contact. ACE scores were derived from the observed protein-atom/protein-atom contacts in 90 high-resolution crystal structures for all pairs of 18 atom types. The total desolvation score of a complex is simply the sum of the ACE scores of all receptor-ligand atom pairs within a distance cutoff of 6 Å. In order to improve the computational speed using an FFT-based search algorithm, we use 18 non-pairwise ACE scores (the e_i scores in Table 3 of [24]), representing the score between one protein atom of a specific type and another protein atom of an "averaged" type. Previously we combined this desolvation term with the grid based shape complementarity function GSC and showed a drastic improvement on docking performance[10].

PSC is composed of a favorable term and a penalty term. The favorable term calculates the total number of atom pairs between the receptor and the ligand within a distance cutoff (D plus the receptor atom radius). It is similar to the above ACE-based desolvation energy, except that ACE assigns score e_{ij} to a pair of atoms of types i and j, and PSC assigns all atom pairs the same score regardless their types[23]. The penalty term of PSC prevents clashes by assigning –81, –27 and –9 to every core-core, surface-core and surface-surface grid point overlap respectively.

The easiest way of combining PSC with ACE would be simply summing these two terms. However, positive PSC scores indicate good shape complementarity, with each atom pair receiving the score of 1, while ACE scores can be positive (unfavorable) or negative (favorable), ranging between 1.334 and -1.827. In order to make these two scores compatible, we flip the signs of the PSC scores. To keep the penalty term of PSC unaltered, we need to make the "favorable" component of the PSC+DE scoring function equal to or smaller than zero. Therefore, we decrease the PSC score for each atom pair from -1 to -1.334, to counter the most unfavorable ACE score. Then the two terms are summed. Thus, a more negative score indicates a more favorable interaction energy.

In order to compute PSC+DE efficiently using FFT, four discrete functions on an $N \times N \times N$ grid, R_{PSC}, L_{PSC}, R_{DE} and L_{DE}, are used to describe the shape and desolvation properties of the receptor and ligand, and the PSC+DE scoring function S_{PSC+DE} is expressed as correlations of these four functions:

$$R_{PSC} = L_{PSC} = \begin{cases} 3 & \text{solvent excluding surface layer of the protein} \\ 3^2 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

$$\mathrm{Re}[R_{DE}] = \mathrm{Re}[L_{DE}] = \begin{cases} \text{sum of PSC and ACE scores of all } \textit{nearby atoms} & \text{open space} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Im}[R_{DE}] = \mathrm{Im}[L_{DE}] = \begin{cases} 1 & \text{if this grid point is the nearest grid point of an atom} \\ 0 & \text{otherwise} \end{cases}$$

[1]

$$S_{PSC+DE} = \mathrm{Re}[R_{PSC} \bullet L_{PSC}] + \frac{1}{2} \times \mathrm{Im}[R_{DE} \bullet L_{DE}]$$

where R_{PSC} and L_{PSC} are real functions, and R_{DE} and L_{DE} are complex functions. Re[ ] and Im[ ] denote the *real* and *imaginary* parts of a complex function. If a protein atom has more than 1 Å^2 solvent accessible area, calculated using a water probe radius of 1.4 Å[25], it is considered a surface atom. Otherwise, it is a core atom. The "solvent excluding surface layer of the protein" is defined by the grid points corresponding to surface atoms. All other grid points corresponding to core atoms are in the "protein core". "Nearby atoms" are atoms within the distance cutoff (D plus the receptor atom radius) of a grid point. $\mathrm{Im}[R_{DE} \bullet L_{DE}]$ is divided by 2 since each atom pair has been counted twice.

PSC+DE+ELEC

Similar to our previous work[10], we compute the electrostatics energy using the Coulombic formula, which is expressed as a function of the electrical potential generated by the receptor and the partial charges of ligand atoms. We multiply the resulting electrostatics energy with a scaling factor β, and add it to PSC and DE scores. In practice, this sum can be directly evaluated using the FFT search algorithm. Two new discrete functions are involved: $R_{PSC+ELEC}$ and $L_{PSC+ELEC}$, in addition to R_{DE} and L_{DE} defined in Equation [1].

$$\mathrm{Re}[R_{PSC+ELEC}] = \mathrm{Re}[L_{PSC+ELEC}] = \begin{cases} 3.5 & \text{solvent excluding surface layer of the protein} \\ 3.5^2 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

$$\mathrm{Im}[R_{PSC+ELEC}] = \begin{cases} \beta \times (\text{electric potential of all receptor atoms}) & \text{open space} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Im}[L_{PSC+ELEC}] = \begin{cases} -1 \times (\text{atom charge}) & \text{if this grid point is the nearest grid point of a ligand atom} \\ 0 & \text{otherwise} \end{cases} \quad [2]$$

$$S_{PSC+DE+ELEC} = \mathrm{Re}[R_{PSC+ELEC} \bullet L_{PSC+ELEC}] + \frac{1}{2} \times \mathrm{Im}[R_{DE} \bullet L_{DE}]$$

In the above equation, the penalty component of PSC and ELEC have been assigned to the *real* and *imaginary* parts of $R_{PSC+ELEC}$, respectively. Thus, PSC+DE+ELEC has the same computational complexity as PSC+DE. The PSC penalty is increased slightly to balance the increased favorable contribution by electrostatics. β is defaulted to 3, with no major impact on the performance when varied by 50-200%. The default β value does not indicate that the electrostatics energy contributes three times as much as PSC and DE do to the final scoring function. In the original ACE publication, all ACE scores were multiplied by 1/21 to transform dimensionless contact energies into the kcal/mol unit[24]. Thus, the electrostatics energy contributes β/21=1/7 as much as PSC and DE. This is consistent with the noisy nature of the Coulombic electrostatics. In fact, FTDock could only use electrostatics as a filter and the authors indicated that it was too noisy to be a direct component of their scoring function[3].

Performance Evaluation

We used version 0.0 of a benchmark developed in our lab[22], which contained 23 enzyme/inhibitor, 16 antibody/antigen and 10 other types of test cases. For antibodies, we restricted the search to complementarity determining regions defined using only sequence information[10]. For all other proteins, we assumed no binding site information and performed a full search. The performance of different scoring functions is evaluated using success rate and hit count, as defined previously[23]. Given the number of predictions being evaluated for each test case (N_P), success rate is the percentage of test cases in the benchmark, for which at least one near-native structure (hits) has been found, and hit count is the average number of hits per test case ranked within N_P. Hits are predictions with Root Mean Square Deviation (RMSD) below 2.5 Å after superposition. Superposition and RMSD calculation only involve the C_α atoms of interface residues, which are receptor (or ligand) residues with at least one atom within 10 Å of any atoms of the ligand (or receptor).

Computational Implementation

ZDOCK is written in C, and parallelized using Message Passing Interface. We have assigned different version numbers to the various scoring functions compared in this paper: ZDOCK1.3 for GSC+DE+ELEC[10], ZDOCK2.1 for PSC[23], ZDOCK2.2 for PSC+DE (this work) and ZDOCK2.3 for PSC+DE+ELEC (this work). The average computing time for ZDOCK2.2 or ZDOCK2.3 per complex on a 16 processor IBM-SP4 is 4 minutes. The program is freely available to academic researchers at http://zlab.bu.edu/~rong/dock.

Results

Performance Averaged over the Entire Benchmark

For each test case, we obtain the number of hits ranked above some number of predictions being evaluated and the rank of the best ranked hit. Table 1 contains the results for all four scoring functions at a 6° rotational sampling interval (Δ=6°), corresponding to 54000 rotations. The GSC+DE+ELEC results are not directly comparable with those in our previous paper[10], since we have made 3 modifications: (1) Previously we rotate the ligand molecule evenly around the X-, Y-, and Z-axes; now we use a set of Euler angles corresponding to a uniformly distributed set of points on a projective sphere. (2) We used to keep the top 10 translational orientations per rotation; now we only keep one, since we have discovered that the top 10 translations are usually extremely similar and keeping only the best one helps to remove false positives without affecting the ranking of the first hit. (3) Now we randomly perturb all starting receptor and ligand orientations to avoid deliberately sampling a near-native orientation. The calculation of PSC is the same as before, except that we previously reported the results for Δ=15° [23], and now we present the results for Δ=6° in Table 1.

The average performance over the entire benchmark is best illustrated using Success Rate and Hit Count vs. Number of Predictions graphs (Figure 1). Here, the data correspond to Δ=15°. Success rate reflects the average ability of a scoring function for ranking a hit within some number of predictions being evaluated (N_P). For example, at N_P=5 the success rate is 31% (or 15 test cases) for PSC+DE+ELEC, indicating that this scoring function ranks one or more hits in the top 5 for 15 test cases. Figure 1a indicates that at most N_P values PSC+DE+ELEC performs better than PSC+DE, which outperforms PSC. Compared to the PSC family of scoring functions, GSC+DE+ELEC performs the best at N_P=1; it becomes worse than PSC+DE+ELEC for N_P>1, also worse than PSC+DE for N_P>10, and even worse than PSC for N_P>200. At a rotational sampling density of 6° (graph not shown), the above description remains largely valid except that GSC+DE+ELEC is the worst performer for N_P=1 (Table 2). Moreover, PSC+DE+ELEC is clearly the best, with a success rate higher than those of all other scoring functions by 13% (or 6 complexes) at N_P=1000.

Hit count indicates the average number of hits a target function can retain within some number of predictions being evaluated (Figure 1b). For example, at N_P=5, the hit count for PSC+DE+ELEC is 0.4, meaning that this target function retains on average 0.4 hits per test case. Figure 1b indicates that adding DE to PSC leads to more hits over all N_P values, and adding ELEC leads to even more hits. If 1000 predictions are evaluated for each test case, the hit count is 5.2, 6.7 and 7.3 for PSC, PSC+DE and PSC+DE+ELEC respectively. GSC+DE+ELEC has comparable hit count to PSC+DE+ELEC at N_P<100. For N_P>100, it has lower hit count than

PSC+DE+ELEC, but higher than or comparable to PSC+DE. At Δ=6°, the relative performance of the PSC family of scoring functions remains similar to the above description (graph not shown). GSC+DE+ELEC and PSC+DE+ELEC have comparable hit counts throughout the entire range of N_P.

Previously, we reported that on average, denser rotational sampling leads to worse success rate but much higher hit count for PSC[23]. This is also true for other scoring functions. For some test cases (such as 10 out of 49 test cases in the case of PSC+DE+ELEC), finer sampling can produce better rankings for the best ranked hits, simply because these hits were missed at coarser sampling. However, for many test cases, since the highest ranked prediction is not a hit, finer sampling tends to extend the list of false positives and thus lead to a worse rank for the best ranked hit. Since our goal here is to achieve the best performance in the initial stage of protein docking, it is important for a scoring function to retain at least one hit in a reasonably small N_P for the majority of test cases. Most post-processing methods can comfortably handle 1-2 thousand predictions. Our experience on post-processing indicates that it is best to compare scoring functions for the top 1000 predictions at Δ=15°, and for 2000 predictions at Δ=6°. Therefore, in Table 1 we have included the number of hits each scoring function can retain within the top 2000 predictions.

Similar to other docking algorithms, ZDOCK performs best on the enzyme/inhibitor category of test cases, compared to antibody/antigen and others. This applies to all four scoring functions discussed here. Nonetheless, the improvement of PSC+DE over PSC, as well as the improvement of PSC+DE+ELEC over both PSC+DE and PSC, is consistently observed across all three categories of test cases. Interestingly, the improvement of PSC+DE+ELEC over GSC+DE+ELEC differs among three categories of test cases. For the rest of the Results section, we focus on the comparison of these two scoring functions category by category. In Figure 2, we plot success rate and hit count in each category of test cases, for both 6° and 15° rotational sampling intervals.

Antibody/Antigen

The superior performance of PSC+DE+ELEC over GSC+DE+ELEC in Figure 1 can be largely attributed to the antibody/antigen category of test cases. Figure 2a indicates that at Δ=15°, PSC+DE+ELEC has drastically higher success rates than GSC+DE+ELEC for all N_P values except N_P=1. The exception is due to 1AHW, for which GSC+DE+ELEC ranks a hit as the number 1 prediction, while PSC+DE+ELEC only ranks a hit at 10 for this test case. Interestingly, the relative performance for these two scoring functions on 1AHW is quite different at Δ=6° (Table 1): the highest rank for a hit is only 11 for GSC+DE+ELEC, and PSC+DE+ELEC ranks a hit at 7, indicating that its poorer performance at Δ=15° was due to under-sampling. At Δ=6°, PSC+DE+ELEC has higher success rates for all N_P values (Figure 2a). Figure 2b indicates that at either sampling density, PSC+DE+ELEC produces approximately twice as many hits as GSC+DE+ELEC, across the entire N_P range.

A case-by-case comparison indicates that PSC+DE+ELEC beats GSC+DE+ELEC on almost all test cases, both in terms of the ranking of the first hit and in terms of the number of hits retained (Table 1). Impressively, except for one test case (1DQJ, with 1415 and 9249 being the best rank

for a hit at Δ=15° and 6° respectively), PSC+DE+ELEC is able to produce at least one hit within some reasonable number of predictions (we use N_P=1000 at Δ=15° and N_P=2000 at Δ=6°) for all other 19 antibody/antigen test cases. This represents a significant improvement over previous first-stage unbound docking algorithms, which either examined very few antibody/antigen, or reported much worse performance on this class of test cases.

Enzyme/Inhibitor

In Figure 2c, we show that PSC+DE+ELEC consistently achieves higher success rates than GSC+DE+ELEC at both rotational sampling densities. The only exception is that at Δ=15°, GSC+DE+ELEC ranks a hit as the number 1 prediction for 9 test cases, while PSC+DE+ELEC only succeeds in doing so for 7 test cases. Figure 2d indicates GSC+DE+ELEC on average produces more hits than PSC+DE+ELEC, especially at Δ=6°. Upon close examination of individual test cases in Table 1, we discover that the elevated hit count for GSC+DE+ELEC is due to its ability to retain many hits for six test cases (2SNI, 2SIC, 1ACB, 1MAH, 1UGH and 1STF). PSC+DE+ELEC also performs very well for the last five test cases, with over a dozen hits within the top 1000 predictions for each test case. 2SNI seems to be the only enzyme/inhibitor test case, with which the PSC family of scoring functions has some difficulty. In contrast, at Δ=6°, GSC+DE+ELEC struggles with four test cases (2KAI, 1BRS, 1FSS and 1TAB). Especially for 1TAB, GSC+DE+ELEC is not able to retain a hit within the top 1000 predictions at Δ=15°, nor within the top 2000 predictions at Δ=6°.

Others

Figure 2e indicates that PSC+DE+ELEC prevails in the low N_P range, while GSC+DE+ELEC takes over at larger N_P. Figure 2f indicates PSC+DE+ELEC produces many more hits than GSC+DE+ELEC at Δ=6°, while they perform comparably at Δ=15°. Since test cases in this category are diverse, close examination of individual ones is important. All scoring functions have failed on 1AVZ and 1MDA. In addition, GSC+DE+ELEC fails on 1IGC. PSC+DE+ELEC cannot find any hits for this test case at Δ=15° due to under-sampling. The particular random starting orientation happens to produce poor results. We have rerun the program at Δ=15° with 10 random starting orientations, and obtained at least one hit in the top 1000 predictions for 8 runs. At Δ=6°, PSC+DE+ELEC successfully retains 3 hits in the top 2000 for 1IGC, with the best rank being 153. However, PSC+DE+ELEC fails on two other test cases (2PCC and 1GLA), for which GSC+DE+ELEC performs well for both Δ values. For the remaining five test cases (1WQ1, 1ATN, 1SPB, 2BTF and 1A0O), PSC+DE+ELEC consistently performs better than GSC+DE+ELEC, indicated by a better rank for the first hit and/or more hits. PSC+DE+ELEC generates a large number of hits for 1WQ1 at Δ=6°, which accounts for its high hit count in Figure 2f.

Discussion

We have developed a new scoring function PSC+DE+ELEC for the initial stage of unbound docking. It combines our recently developed shape complementarity scoring function PSC[23] with desolvation and electrostatics. We compared PSC+DE+ELEC with three other scoring functions – PSC, PSC+DE and GSC+DE+ELEC – on a large benchmark of test cases. We have

implemented all of these scoring functions in our FFT-based docking algorithm ZDOCK. Our results demonstrate that with PSC+DE+ELEC we are getting close to solving the initial stage of the unbound docking problem. Out of 49 test cases, only three proved difficult for all scoring functions (1DQJ, 1AVZ and 1MDA). The best scoring function PSC+DE+ELEC failed on two more test cases (2PCC and 1GLA). Therefore, for 90% test cases, ZDOCK with PSC+DE+ELEC can retain at least one hit within the top 2000 predictions at Δ=6°, with an average of 52 hits per test case.

The three most difficult test cases are: Hyhel-63 Fab / Lysozyme (1DQJ), HIV-1 NEF / FYN tyrosine kinase SH3 domain (1AVZ) and Methylamine dehydrogenase / Amicyanin (1MDA). The other two test cases that PSC+DE+ELEC had difficulty with were Cytochrome C Peroxidase / Iso-1-Cytochrome C (2PCC) and Glycerol kinase / GSF III (1GLA). The first possible explanation that comes to mind is that these are low affinity complexes. Indeed, the binding free energy is -11.5 kcal/mol for 1DQJ[26], -10.4 kcal/mol for 1AVZ[27], -7.2 kcal/mol for 1MDA[28], -10.0 kcal/mol for 2PCC[29] and -7.1 kcal/mol for 1GLA[30], respectively. These are all within the weaker half of the affinity range in the benchmark.

Poor bound docking results on a test case can suggest explanations for the poor performance of unbound docking on the same test case. Applying PSC+DE+ELEC with default parameters for unbound docking to the bound components of the crystal complexes in all 49 test cases, we were able to rank a hit as the number 1 prediction for 29 test cases, and at least one hit in the top 10 for 11 additional test cases. However, the best rank of a hit for 1MDA was 1377, much worse than all other test cases. 2PCC had the second worst rank at 658. Both 1MDA and 2PCC are electron transfer complexes. Close inspection of the crystal complexes reveals many cavities at the interface, perhaps important for the electron transfer function. Therefore, we conclude that the poor performance for 1MDA and 2PCC is due to their weak binding affinities.

The second possible explanation is conformational flexibility, especially for the test cases that bound docking performs well on. Bound docking on 1DQJ and 1AVZ produces the best ranked hits as the number 1 and 6 predictions, respectively. Therefore, these two test cases are not inherently difficult. Close inspection reveals significant backbone conformational changes for both of these two test cases: the RMSD between the bound and unbound conformations of residues 99-103 of lysozyme in 1DQJ is 3.35 Å. After replacing these residues in the unbound structure with their bound conformations, PSC+DE+ELEC is able to identify one hit ranked at 742. The N-terminal tail of Nef in 1AVZ, which forms part of the binding site for the Fyn SH3 domain, is highly flexible in the unbound state (residues 71-73 disordered and residues 74-78 with RMSD of 1.78 Å compared to the bound structure). After replacing these residues with their bound conformations, PSC+DE+ELEC was able to identify a hit ranked at 1667. Both of the above two calculations were performed at Δ=6°.

The poor performance on 1GLA is a bit puzzling. Unlike the four test cases described above, 1GLA is an unbound/bound test case. Unbound docking had a rank for the first hit at 181, somewhat poor. Its weak binding free energy (-7.1 kcal/mol) could take the blame. We have noticed that PSC is particularly capable of identifying large concave binding pockets[23]. Visual inspection indicates that numerous top ranked false positives form clusters at three large concave pockets of the glycerol kinase, with one of them being its deep funnel-like active site;

unfortunately none of these is the binding site for GSF-III. Therefore, two reasons could account for 1GLA: poor binding affinity and the high tendency of PSC+DE+ELEC in docking molecules into large pockets.

Ultimately, ZDOCK must be combined with a refinement method. If the refinement method can handle 2000 predictions per test case, we recommend using the top 2000 predictions generated at Δ=6°, since we observe under-sample at Δ=15° for some test cases. The number of hits that each scoring function can retain within the top 2000 predictions at Δ=6° is 44 for PSC+DE+ELEC, 42 for PSC, 40 for PSC+DE and 39 for GSC+DE+ELEC respectively (Table 1). The difference can be almost completely explained by the performance on the antibody/antigen category of test cases. Even though PSC+DE+ELEC is the best, the other three scoring functions are not too far behind at N_P=2000. Since the performance of a refinement method can be heavily influenced by the type of false positives produced by ZDOCK, and the scoring functions discussed here generate different types of false positives, it is likely that a refinement method works best with PSC, and not with PSC+DE+ELEC. Therefore, we have made all four scoring functions available as different versions of ZDOCK. The top predictions for all test cases in the benchmark are also available at http://zlab.bu.edu/~rong/dock, and they should be helpful for the development of refinement methods.

Acknowledgements

We thank Julian Mintseris for critically reading the manuscript, Dr. Julie C. Mitchell for providing uniform Euler angle sets, Dr. Natalie Strynadka for providing the coordinates of BLIP crystal structures, Prof. Kadin Tseng and Bob Walkup for the help in MPI implementation. We are grateful to the Scientific Computing Facilities at Boston University and the Advanced Biomedical Computing Center at NCI, NIH for computing supports. This work was funded by NSF grants DBI-0078194, DBI-0133834 and DBI-0116574.

References:

1. Sternberg MJ, Gabb HA, Jackson RM, Moont G. Protein-protein docking. Generation and filtering of complexes. Methods Mol Biol 2000;143:399-415.
2. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol 2002;12(1):28-35.
3. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272(1):106-20.
4. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF. Protein docking using continuum electrostatics and geometric fit. Protein Eng 2001;14(2):105-113.
5. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39(2):178-94.
6. Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. Proteins 1997;Suppl 1:226-30.
7. Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. Protein Sci 2001;10(11):2147-61.
8. Palma PN, Krippahl L, Wampler JE, Moura JJ. BiGGER: A new (soft) docking algorithm for predicting protein interactions. Proteins 2000;39(4):372-84.
9. Taylor JS, Burnett RM. DARWIN: a program for docking flexible molecules. Proteins 2000;41(2):173-91.
10. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins: Structure, Function and Genetics 2002;47:281-294.
11. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci U S A 2001;98(19):10636-41.
12. Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein-protein docking. Proteins 2000;40(3):525-37.
13. Weng Z, Vajda S, Delisi C. Prediction of protein complexes using empirical free energy functions. Protein Science 1996;5:614-26.
14. Jackson RM, Sternberg MJE. A continuum model for protein-protein interactions: Application to the docking problem. Journal of Molecular Biology 1995;250:258-275.
15. Jackson RM, Gabb HA, Sternberg MJ. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276(1):265-85.
16. Cherfils J, Janin J. Protein docking algorithms: simulating molecular recognition. Current Opinion in Structural Biology 1993;3:265-269.
17. Janin J. Protein-protein recognition. Progress in Biophysics and Molecular Biology 1995;64(2-3):145-66.
18. Shoichet BK, Kuntz ID. Predicting the structure of protein complexes: a step in the right direction. Chem Biol 1996;3(3):151-6.
19. Sternberg MJ, Gabb HA, Jackson RM. Predictive docking of protein-protein and protein-DNA complexes. Curr Opin Struct Biol 1998;8(2):250-6.
20. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guid to scoring functions. Proteins: Structure, Function and Genetics 2002;47:409-443.

21. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry 1983;4:187-217.
22. Chen R, Mintseris J, Janin J, Weng Z. A protein-protein docking benchmark. Proteins: Structure, Function and Genetics In Press.
23. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. Proteins: Structure, Function and Genetics In Press.
24. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 1997;267:707-26.
25. Lee B, Richards FM. The interpretation of protein structures: Estimation of static accessibility. Journal of Molecular Biology 1971;55:379-400.
26. Li Y, Li H, Smith-Gill SJ, Mariuzza RA. Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63(,). Biochemistry 2000;39(21):6296-309.
27. Arold S, Franken P, Strub MP, Hoh F, Benichou S, Benarous R, Dumas C. The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. Structure 1997;5(10):1361-72.
28. Davidson VL, Graichen ME, Jones LH. Binding constants for a physiologic electron-transfer protein complex between methylamine dehydrogenase and amicyanin. Effects of ionic strength and bound copper on binding. Biochim Biophys Acta 1993;1144(1):39-45.
29. Corin AF, McLendon G, Zhang Q, Hake RA, Falvo J, Lu KS, Ciccarelli RB, Holzschu D. Effects of surface amino acid replacements in cytochrome c peroxidase on complex formation with cytochrome c. Biochemistry 1991;30(49):11585-95.
30. Novotny MJ, Frederickson WL, Waygood EB, Saier MH, Jr. Allosteric regulation of glycerol kinase by enzyme IIIglc of the phosphotransferase system in Escherichia coli and Salmonella typhimurium. J Bacteriol 1985;162(2):810-6.

Table 1: Docking performance with 6° rotational sampling interval [a]

Test Case[b]	GSC+DE+ELEC		PSC		PSC+DE		PSC+DE+ELEC		
	Hits[c]	Rank[d]	Hits[c]	Rank[d]	Hits[c]	Rank[d]	Hits[c]	Rank[d]	RMSD[e]
1CGI	77	3	54	4	77	7	**77**	**4**	2.41
1CHO	93	22	66	1	82	1	**99**	**3**	1.57
2PTC	62	65	2	1655	20	434	48	193	1.83
1TGS	86	5	107	3	145	1	**109**	**3**	2.22
2SNI	60	169	0	7434	1	1544	1	1262	2.22
2SIC	115	2	24	241	53	46	**52**	**11**	2.37
1CSE	87	3	3	1537	14	429	29	198	2.20
2KAI	1	1772	3	1399	19	339	16	388	1.61
1BRC	24	52	16	173	42	109	54	24	2.32
1ACB	199	3	38	25	79	12	**93**	**18**	1.33
1BRS	3	1019	34	61	28	67	21	65	2.13
1JTG	72	1	69	3	76	3	**82**	**1**	1.52
1MAH	58	9	6	849	23	97	28	24	1.29
1UGH	58	14	4	305	28	6	**20**	**8**	2.25
1DFJ	43	2	15	37	11	6	**51**	**1**	2.48
1FSS	2	1066	5	731	11	204	15	50	1.52
1AVW	2	704	28	45	29	12	**52**	**3**	2.07
1PPE*	318	1	272	1	364	1	**393**	**1**	0.90
1TAB*	0	10783	47	65	8	565	50	79	1.21
1UDI*	41	198	16	31	34	2	**35**	**5**	1.19
1STF*	152	1	42	1	87	1	**83**	**1**	0.88
2TEC*	226	1	77	1	180	1	**185**	**1**	0.76
4HTC*	73	2	54	1	62	3	**57**	**3**	2.46
1MLC	16	134	3	1106	7	433	17	128	1.65
1WEJ	1	1940	4	1396	0	2597	22	183	1.04
1AHW	64	11	28	26	27	76	**67**	**7**	1.82
1DQJ	0	46002	1	1341	0	6055	0	9249	2.37
1BVK	0	40864	2	974	4	496	2	821	2.34
1FBI*	3	561	2	1786	1	1827	5	642	2.03
2JEL*	0	4296	62	112	42	91	35	233	1.46
1BQL*	114	4	16	172	33	127	**70**	**13**	1.07
1JHL*	0	4259	15	404	0	2275	12	333	1.37
1NCA*	9	211	55	2	56	1	**67**	**1**	1.06
1NMB*	3	1108	6	693	3	1473	9	135	0.98
1MEL*	32	9	52	12	108	4	**71**	**3**	1.19
2VIR*	0	3003	3	476	1	1896	3	1101	1.03
1EO8*	0	8420	0	4366	0	5801	2	1497	0.96
1QFU*	4	606	10	407	12	307	18	388	1.14
1IAI*	3	905	0	2525	2	1151	3	997	1.70
2PCC	6	702	0	-	0	-	0	22338	2.49
1WQ1	10	131	26	5	24	28	**54**	**15**	1.31
1AVZ	0	39047	0	-	0	-	0	53466	1.61
1MDA	0	16183	0	33988	0	32051	0	18034	2.29
1IGC*	0	5088	6	22	15	38	3	153	1.20
1ATN*	47	13	1	360	9	118	**24**	**7**	0.80
1GLA*	19	214	0	-	0	28601	0	9794	1.55
1SPB*	106	1	75	1	103	1	**112**	**1**	0.61
2BTF*	17	27	13	32	6	166	**35**	**2**	0.95
1A0O*	9	619	2	833	0	7889	4	284	2.45

[a] For bolded test cases, PSC+DE+ELEC ranked at least one hit within the top 20 predictions.
[b] 4-letter Protein Data Bank (PDB) code for the crystal complex of a test case.
[c] Number of hits in the top 2000 predictions. Hits are defined as docked structures with interface RMSD ≤ 2.5 Å from the crystal complex; see Methods for more details.
[d] Rank of the best ranked hit. "–" indicates that no hit was found in the first 54000 predictions.
[e] RMSD for the best ranked hit. "–" indicates that no hit was found in the first 54000 predictions.
* Unbound/bound complexes.

Figure Captions

Figure 1. The performance of PSC+DE+ELEC (◆ and solid orange line), PSC+DE (■ and solid cyan line), PSC (● and dash black line) and GSC+D+E (▲ and solid red line) are compared according to success rate (a) and hit count (b). The rotational sampling interval used here is 15°.

Figure 2. The performance of PSC+DE+ELEC (◆ and solid orange line for 15°, ■ and dash blue line for 6°), and GSC+DE+ELEC (▲ and solid red line for 15°, ● and dash black line for 6°) are compared within 16 antibody/antigen test cases (a and b), 23 enzyme/inhibitor test cases (c and d) and 10 test cases in the others category (e and f). The comparison is based on success rate for (a), (c) and (e), and based on hit count for (b), (d) and (f).

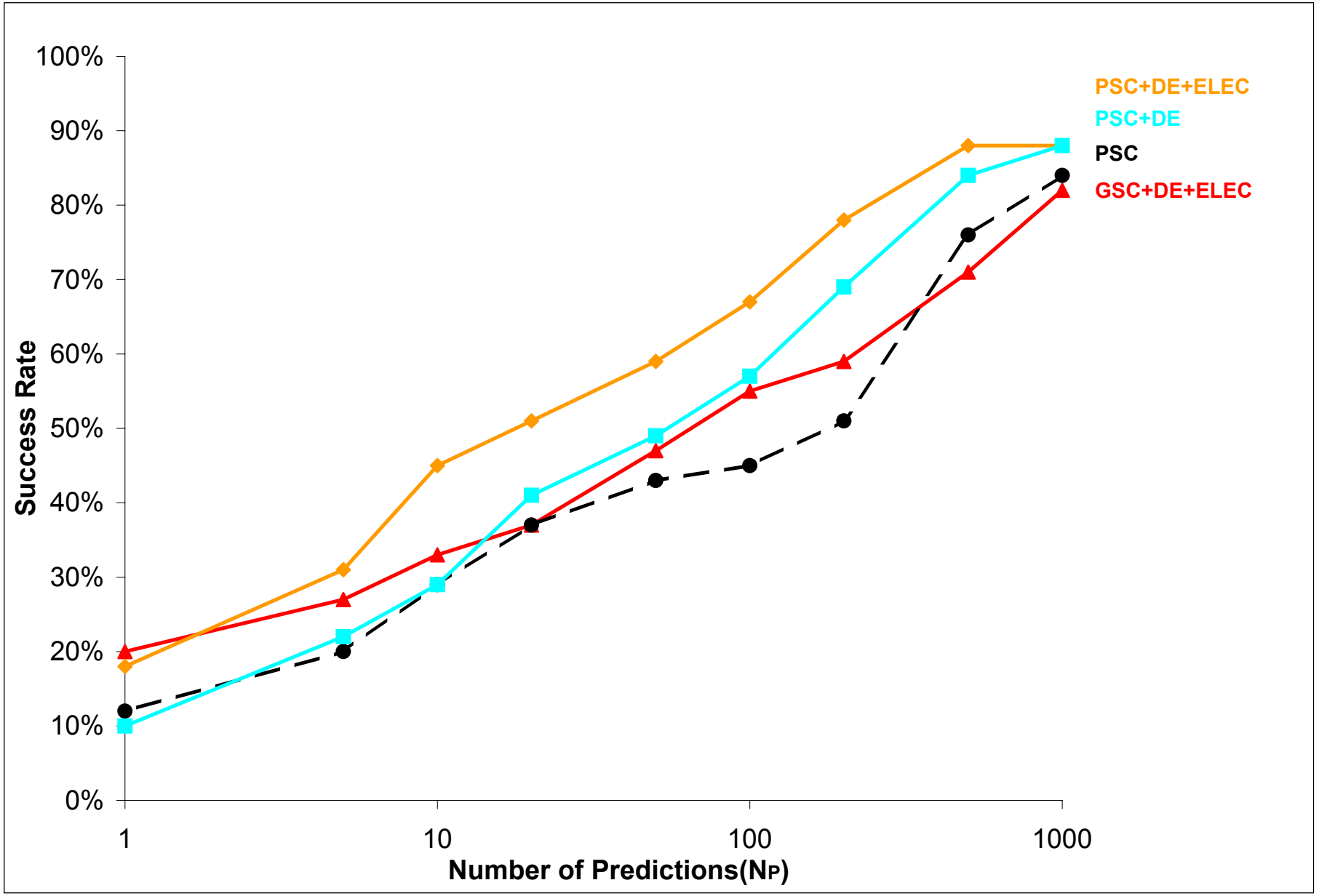
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Success Rate
1
10
100
1000
Number of Predictions(NP)
PSC+DE+ELEC
PSC+DE
PSC
GSC+DE+ELEC

Figure 1a

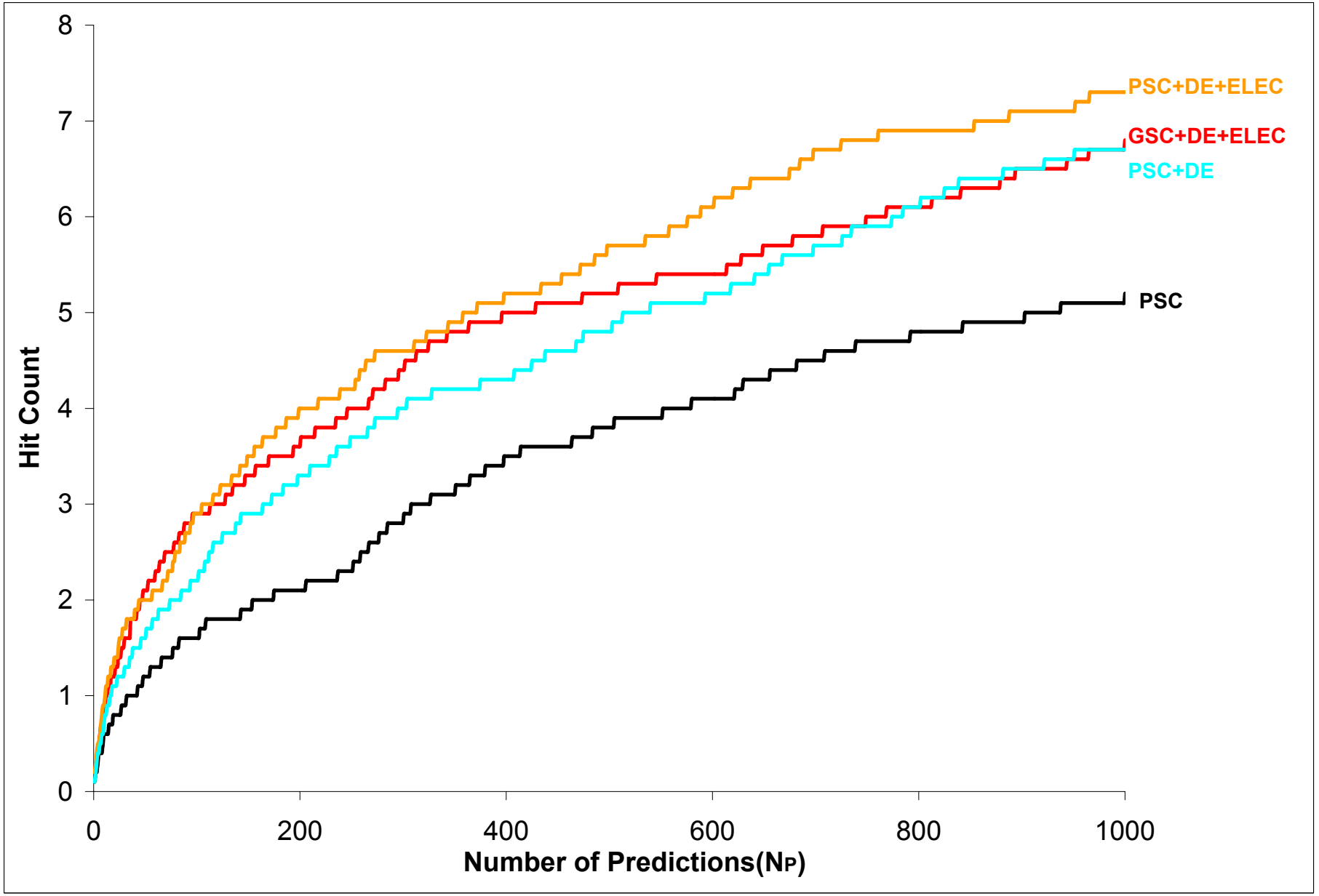
PSC+DE+ELEC
GSC+DE+ELEC
PSC+DE
PSC
Hit Count
Number of Predictions(NP)
0
1
2
3
4
5
6
7
8
0
200
400
600
800
1000

Figure 1b

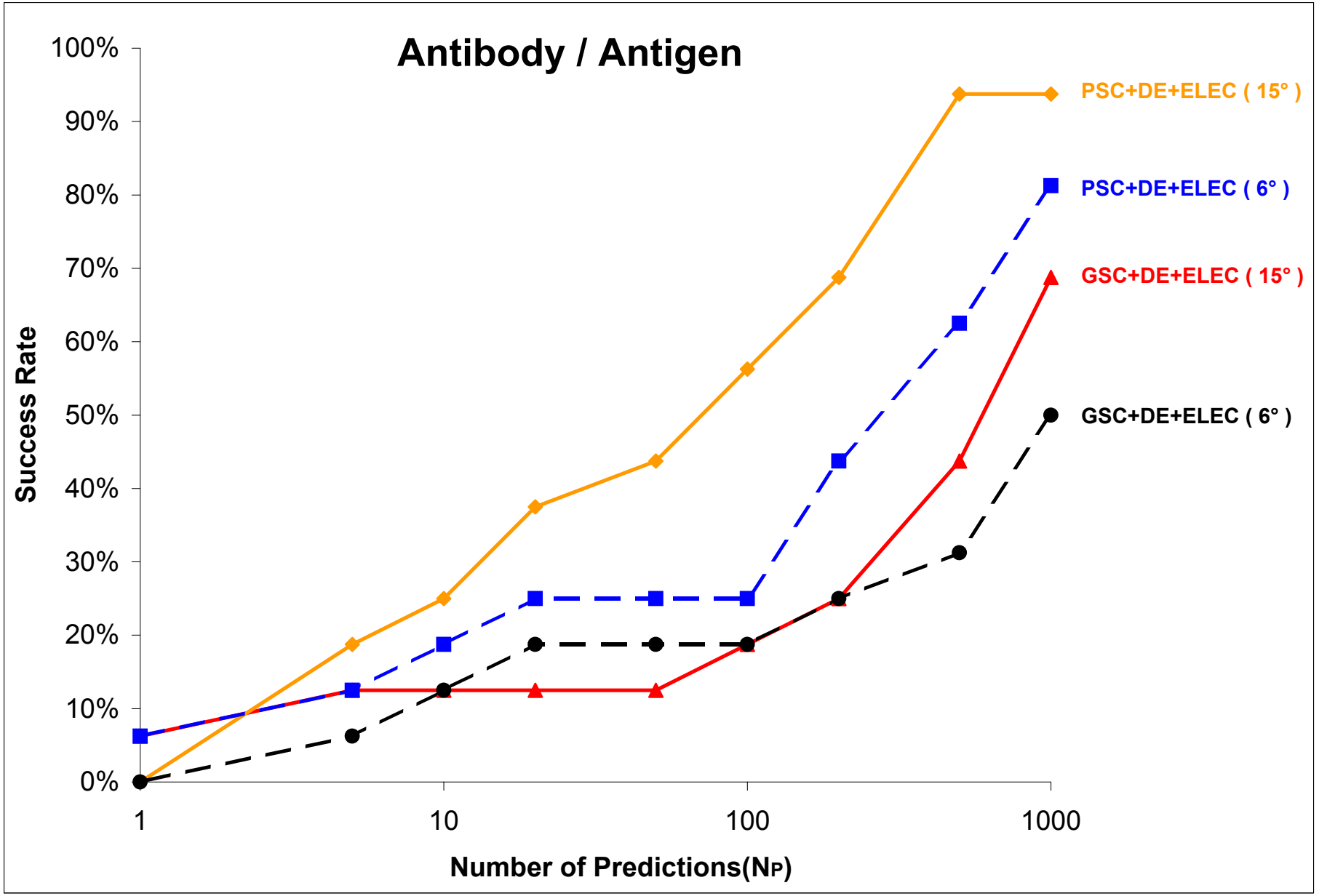
Antibody / Antigen
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Success Rate
1
10
100
1000
Number of Predictions(NP)
PSC+DE+ELEC (15°)
PSC+DE+ELEC (6°)
GSC+DE+ELEC (15°)
GSC+DE+ELEC (6°)

Figure 2a

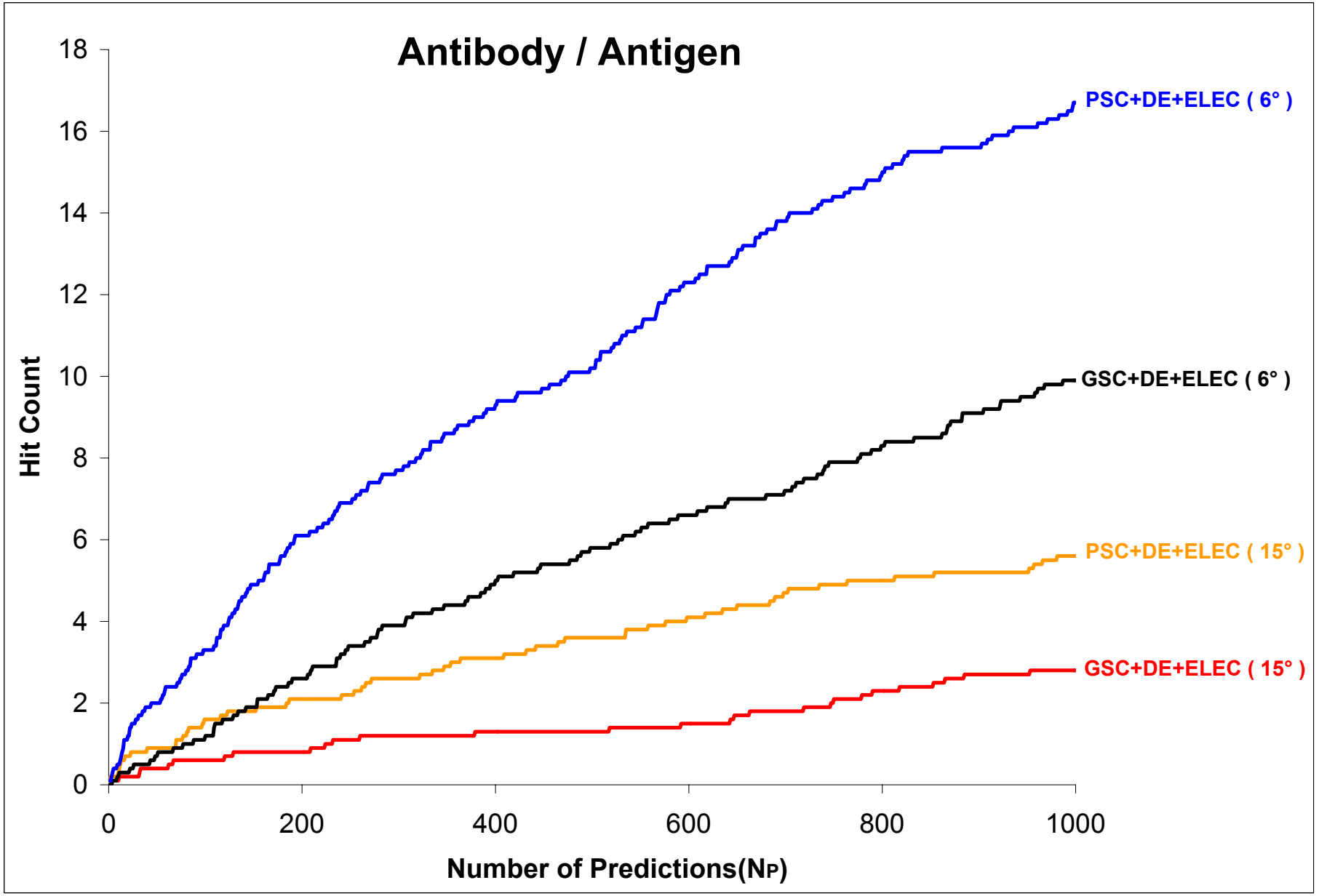
Antibody / Antigen
PSC+DE+ELEC (6°)
GSC+DE+ELEC (6°)
PSC+DE+ELEC (15°)
GSC+DE+ELEC (15°)
Hit Count
Number of Predictions(NP)
0
2
4
6
8
10
12
14
16
18
0
200
400
600
800
1000

Figure 2b

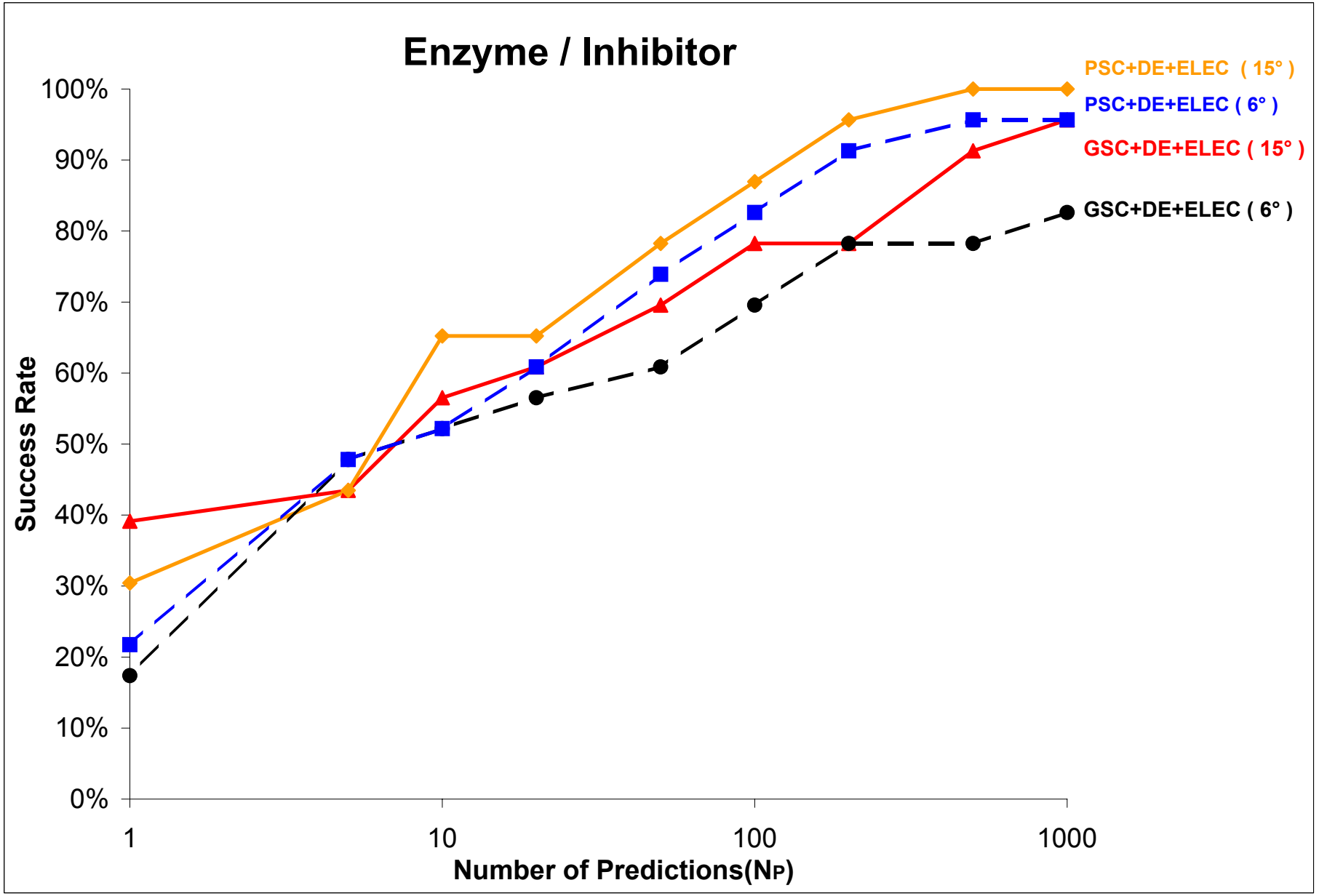
Enzyme / Inhibitor
PSC+DE+ELEC (15°)
PSC+DE+ELEC (6°)
GSC+DE+ELEC (15°)
GSC+DE+ELEC (6°)
Success Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
10
100
1000
Number of Predictions(NP)

Figure 2c

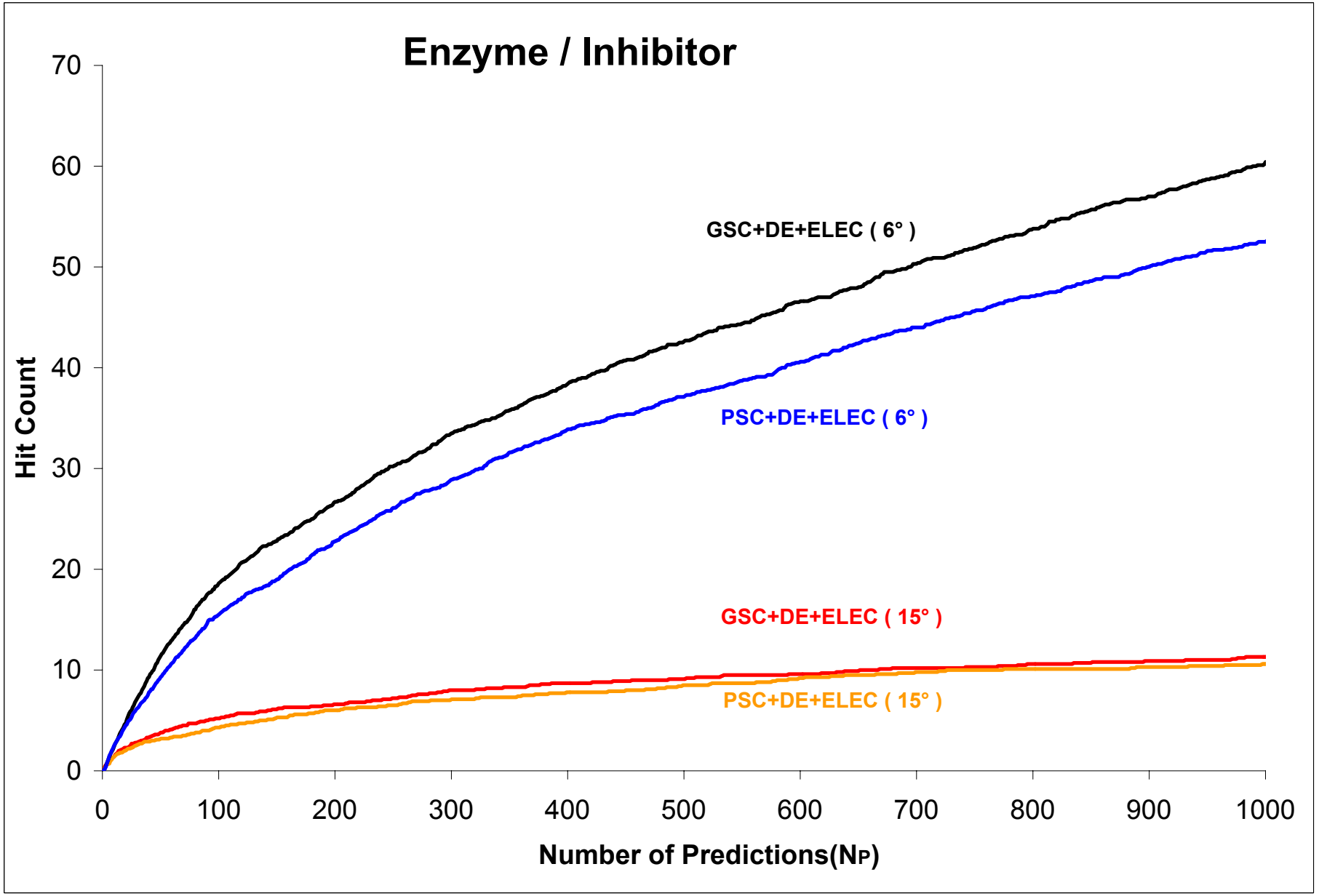
Enzyme / Inhibitor
Hit Count
Number of Predictions(NP)
GSC+DE+ELEC (6°)
PSC+DE+ELEC (6°)
GSC+DE+ELEC (15°)
PSC+DE+ELEC (15°)
0
10
20
30
40
50
60
70
0
100
200
300
400
500
600
700
800
900
1000

Figure 2d

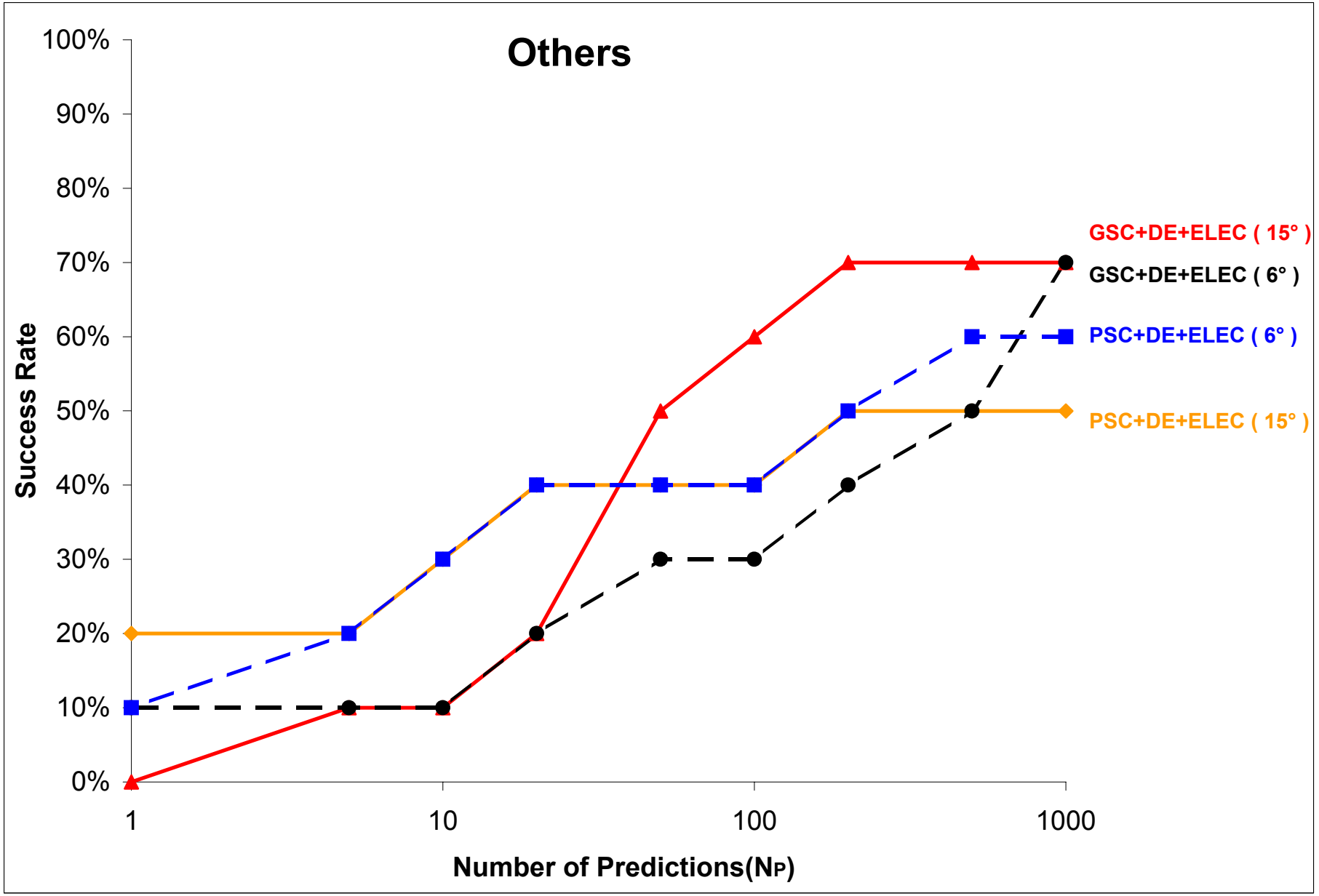
Others
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Success Rate
1
10
100
1000
Number of Predictions(NP)
GSC+DE+ELEC (15°)
GSC+DE+ELEC (6°)
PSC+DE+ELEC (6°)
PSC+DE+ELEC (15°)

Figure 2e

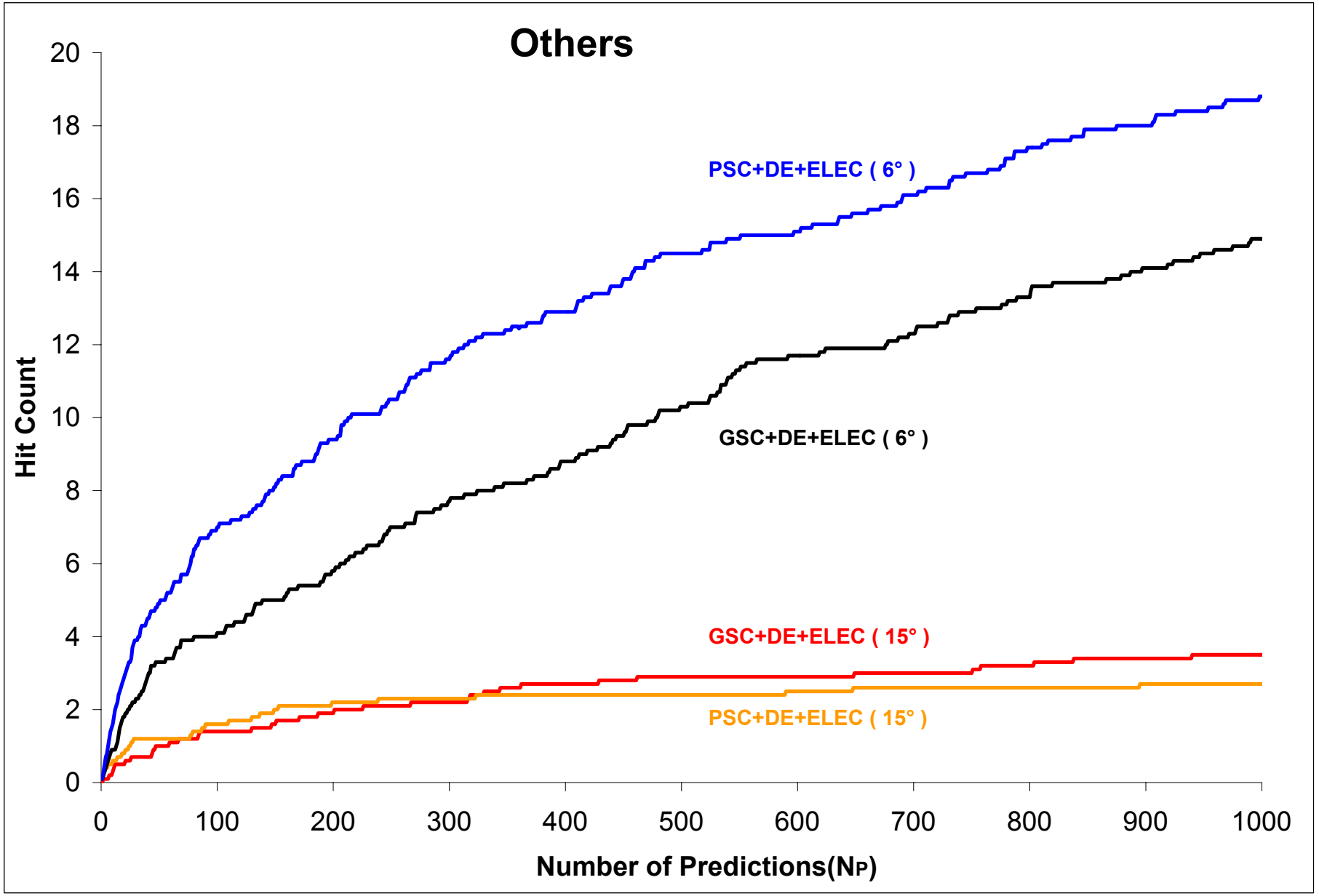
Others
Hit Count
Number of Predictions(NP)
PSC+DE+ELEC (6°)
GSC+DE+ELEC (6°)
GSC+DE+ELEC (15°)
PSC+DE+ELEC (15°)
0
2
4
6
8
10
12
14
16
18
20
0
100
200
300
400
500
600
700
800
900
1000

Figure 2f