Using Kendall-τ Meta-Bagging to Improve Protein-Protein Docking Predictions

Jérôme Azé[1], Thomas Bourquard[2], Sylvie Hamel[3], Anne Poupon[4], and David W. Ritchie[2]

[1] Université de Paris-Sud, Laboratoire de Recherche en Informatique, Bâtiment 650
Équipe Bioinformatique – INRIA AMIB group
F-91405 Orsay Cedex, France
aze@lri.fr
[2] INRIA Nancy-Grand Est, LORIA
615 Rue du Jardin Botanique, 54506 Vandoeuvre-lès-Nancy, France
[3] Université de Montréal
C.P. 6128, Succ. Centre-Ville
Montréal, Québec, Canada, H3C 3J7
[4] BIOS group, INRA, UMR85
Unité Physiologie de la Reproduction et des Comportements
F-37380 Nouzilly, France
CNRS, UMR6175, F-37380 Nouzilly, France
Université François Rabelais, 37041 Tours, France

Abstract. Predicting the three-dimensional (3D) structures of macromolecular protein-protein complexes from the structures of individual partners (docking), is a major challenge for computational biology. Most docking algorithms use two largely independent stages. First, a fast sampling stage generates a large number (millions or even billions) of candidate conformations, then a scoring stage evaluates these conformations and extracts a small ensemble amongst which a good solution is assumed to exist. Several strategies have been proposed for this stage. However, correctly distinguishing and discarding false positives from the native biological interfaces remains a difficult task. Here, we introduce a new scoring algorithm based on learnt bootstrap aggregation ("bagging") models of protein shape complementarity. 3D Voronoi diagrams are used to describe and encode the surface shapes and physico-chemical properties of proteins. A bagging method based on Kendall-τ distances is then used to minimise the pairwise disagreements between the ranks of the elements obtained from several different bagging approaches. We apply this method to the protein docking problem using 51 protein complexes from the standard Protein Docking Benchmark. Overall, our approach improves in the ranks of near-native conformation and results in more biologically relevant predictions.

1 Introduction

Many biological processes and structures depend on proteins and their ability to form complexes with other proteins or macromolecules such as DNA, and

M. Loog et al. (Eds.): PRIB 2011, LNBI 7036, pp. 284–295, 2011.

RNA, for example. Solving the 3D structures of such macromolecular assemblies thus represents a key step in understanding biological machinery and how malfunctions in this machinery can lead to pathologies.

Although several structural genomic projects have led to advances in solving the structures of protein-protein complexes, this still cannot be done on a high-throughput scale. Therefore, because most proteins can often have several interaction partners [11], *in silico* techniques play a crucial role in studying molecular mechanisms. Here, we focus on the so-called protein docking problem, which aims to predict computationally how two proteins might combine to form a binary protein complex.

Several docking algorithms have been described in the literature. These use various techniques to generate and score candidate conformations [26,10,21,12,23]. Since 2001, the CAPRI experiment (Critical Assessment of PRedicted Interactions) [17] has provided an important forum in which to assess different docking algorithms in an unbiased manner, by allowing blind docking predictions to be compared with unpublished NMR and X-ray structures. Although many docking algorithms are now able to generate high quality candidate orientations, current scoring functions are still not yet sensitive enough to extract the few true positives [18] from amongst the hundreds of thousands of generated incorrect conformations.

In the present work, we introduce a novel method to score protein-protein candidate conformations based on bootstrap aggregation ("bagging") of Voronoi fingerprint representations of candidate docking orientations. We also introduce a "meta-bagging" scoring function which uses an approach based on Kendall-τ distance to minimise the pairwise disagreement between the solutions ranked by several machine learning bagging approaches. This approach is tested on the Protein Docking Benchmark set of Hwang *et al.* [16], consisting of 51 target protein complexes.

2 Methods and Data

In many protein docking algorithms, the initial stage of generating candidate conformations is often one of the most CPU-intensive steps. However, thanks to the tremendous computational power of modern graphics processor units (GPU), the Hex spherical polar Fourier correlation algorithm can generate and score billions of candidates conformations to produce a list of few hundred high quality candidate conformations in just a few seconds [27]. Hence, we chose to use Hex to generate the initial candidates for re-ranking. Additionally, we applied two search constraints in Hex in order to focus candidate docking orientations around the most biologically relevant positions. This ensures that at least one "acceptable" solution according to the CAPRI criteria (i.e. a solution with a root mean squared deviation from the correct X-ray structure of less than 10$\mathring{A}$) appears within the top 10 best ranked conformations for 18 out of 51 targets, and within the 50 best ranked solutions for 31 targets. We take 50 solutions to be the maximum practical number to consider for re-ranking because experimental biologists would generally not be prepared to analyse any more than this number.

For the scoring stage, it has been shown previously that by using a coarse-grained Voronoi tesselation representation and evolutionary algorithm, a scoring function may be optimized with satisfying performance [4,9]. In the present work, we use the same formalism for the representation of protein structure and parameter value computations.

2.1 Training Dataset

The Protein Data Bank (PDB) [3] currently contains some 65,000 3D protein structures[1]. However, only a relatively small fraction of these represent different structural fold families. Hence, the structures of some $1,400$ binary hetero-complexes were extracted from the 3D-Complex database [22] for training.

Following [8], we applied some additional constraints to test protein docking methods:

- The 3D structures of the individual partners have to be known in the absence of the other partner.
- Among the heterodimer structures available in PDB, only structures with a chain length greater than 30 amino acids and an X-ray resolution lower than $3.5\mathring{A}$ are retained.
- Non-biological complexes are identified and removed [24].
- The SCOP protein structure classification [25] is used to identify and remove duplicate complexes.
- The 51 complexes of the benchmark were excluded.

This gave a final set of 178 complexes which were treated as positive examples for the learning procedures. Negative examples were generated using the Hex docking algorithm. They correspond to the lowest energy incorrect conformations found by Hex, each having a root mean squared deviation (RMSD) from the native X-ray structure of at least $10\mathring{A}$, and a surface contact area between the two partners of at least $400\mathring{A}^2$. Ten decoys for each native structure were selected randomly and added to the learning set.

2.2 Voronoi Fingerprints and Learning Parameters

Voronoi fingerprint representations of both the positive and negative examples are calculated as follows. For each conformation, we first build a Delaunay triangulation (i.e. the dual of Voronoi tesselation) of the assembly using the CGAL library [7] and we assign one Voronoi node for each amino acid of the constituent proteins. Following [8], the 96 attributes used in learning procedures are: number of residues at the interface, area of the interface, frequencies of the different amino acids at the interface, average volume of the Voronoi cells of the different types of residues at the interface, frequencies of pairs of interacting residues and distances between interacting residues. For the last two parameter types, residues are binned in six categories according to their physico-chemical

[1] `http://www.rcsb.org/pdb/`

properties (large, small, hydrophobic, aromatic, polar, positively and negatively charged). All attributes are computed only for the residues belonging to the core of the interface (i.e. residues of one partner in contact with at least one residue of the other partner and not in contact with the surrounding solvent).

Since a native protein-protein interface typically contains about 30 to 35 amino acid residues with around three or four pair-wise contacts per residue, not all possible pair-wise interactions might appear in a given interface. Hence missing values can pose a problem. We have tested several different approaches to handle missing values (details not shown), and we find that the best approach is to replace any missing values with the maximal value observed in the data set. Once all of the attributes have been assigned values, they are normalised to the range $[0, 1]$ by setting $x_i = \frac{x_i - min_i}{max_i - min_i}$, where min_i and max_i represent the minimum and maximum values observed for attribute i, respectively. Using this representation of the complexes, a set of learning models was built using several state-of-the art machine learning approaches. These are briefly described in Section 2.3.

As the overall goal is to find the best ordered list of candidate docking conformations, the performance of the different approaches were compared using the area under the curve (AUC) of receiver-operator-characteristic (ROC) plots.

2.3 Learning Algorithms

We used five state-of-the-art machine learning approaches, as implemented in Weka [15], to learn models in order to rank the candidate conformations: Naive Bayes, Decision Trees (with Weka command line options: J48, -C 0.25 -M 2), Random Forest (-I 10 -K 0 -S 1), PART Rules (-M 2 -C 0.25 -Q 1), and Ripper Rules (-F 3 -N 2.0 -O 2 -S 1 -P). Because these approaches were designed to learn models that can predict the class associated with each example but not to rank them, we used the probability distribution output by Weka to sort the examples by decreasing probability of being predicted as a positive docking conformation. In cases of ties (equal probabilities for different examples), the examples are ranked from negative to positive.

Bagged models of these five approaches were also learnt. The distribution probabilities of the bagged models, observed on our protein-protein docking dataset, is larger and the associate ranking contains less ties. We also tested a family of ROC-based evolutionary learning functions, using an in-house evolutionary algorithm optimising the area under the ROC curve. The various functions are learned by K independent runs initialized according to different random seeds. It has been shown previously that non-linear functions can out-perform linear functions in several different domains (medical data mining, text mining, etc.), as well as protein docking [4,2]. We therefore focus only on non-linear scoring functions of the form:

$$S_{w,c}(x) = \sum_{i=1}^{d} w_i |x_i - c_i|$$

where $x \in \mathbb{R}^d$ is the current example, $w \in \mathbb{R}^d$ is a weight vector, and c is a center point in $\mathbb{R}^d$. $S_{w,c}$ defines an order on the training set, and the fitness $\mathcal{F}(S_{w,c})$ is defined as the area under the ROC curve associated to the scoring function $S_{w,c}$.

2.4 Bagging

Given m scoring functions, each of which gives different rankings of the same conformations, the "consensus" score for a given example is calculated by medRank or sumRank aggregation of the m lists of scores as:

$$medRank(x) = \underset{S_{w,c} \in \mathcal{S}}{median}\, rank(S_{w,c}(x)),\ sumRank(x) = \sum_{S_{w,c} \in \mathcal{S}} rank(S_{w,c}(x))$$

2.5 Finding a Median under the Kendall-τ Distance

Using the above ranking functions, m different ranks can be associated with each example. A good distance measure for combining or comparing the ranked lists from two bagging approaches, which are considered in the following section as permutations, is the Kendall-τ distance [19]. This measure counts the number of pairwise disagreements between the positions of elements in permutations.

More formally, we define the Kendall-τ distance, denoted d_{KT}, as

$$d_{KT}(\pi, \sigma) = \#\{(i,j) : i < j \text{ and } ((\pi[i] < \pi[j] \text{ and } \sigma[i] > \sigma[j]) \text{ or} (\pi[i] > \pi[j] \text{ and } \sigma[i] < \sigma[j]))\} \quad (1)$$

where $\pi[i]$ is the position of integer i in permutation π and $\#\mathcal{A}$ the cardinality of set $\mathcal{A}$.

One way to generate a consensus permutation for a given set of permutations is to find a *median permutation* for the set – i.e. to find a permutation that minimizes the sum of Kendall-τ distances between it and all other permutations in the given set. More formally, let $\mathcal{S}_n$ denoted, as usual, the set of all permutation of $\{1, 2, 3, \ldots, n\}$. Given any set of permutations $A \subseteq \mathcal{S}_n$ and a permutation π, we have $d_{KT}(\pi, A) = \sum_{\sigma \in A} d_{KT}(\pi, \sigma)$. Then, finding a consensus permutation for a given set under the Kendall-τ distance can be stated formally as follow: Given $A \subseteq \mathcal{S}_n$, we want to find a permutation π^* of $\{1, 2, 3, \ldots, n\}$ such that $d_{KT}(\pi^*, A) \leq d_{KT}(\pi, A)$, for all $\pi \in \mathcal{S}_n$

The problem of finding the median of a set of m permutations of $\{1, 2, 3, \ldots, n\}$ under the Kendall-τ distance is a NP-hard problem when $m \geq 4$. However, this problem has been well-studied in recent years, and several good heuristics are known to exist [1,5,6,13,20,28].

In this work, we choose to apply the heuristic presented in [6] to find our medians: Let $A = \{\pi^1, \ldots \pi^m\}$ be a set of m permutations for which we want to find a median. We apply the heuristic of [6] taking either the identity permutation or each $\pi^\ell \in A$, $1 \leq \ell \leq m$ as a starting point, and we take the "median" to be the best result obtained from these $m + 1$ runs.

The idea of this heuristic is to apply a series of "*good*" movements on some starting permutations in order to bring them closer to a median. A movement is considered good if when applying it to a permutation, the Kendall-τ distance of the obtained permutation to the considered set is less than the previous distance. See [6] for more details.

3 Results

Our approach has been evaluated using three-fold cross-validation (and each experiment is repeated 10 times) and has been validated on the learning set. We first compared the performance of the different approaches using the ROC AUC criterion.

Table 1. Total number of complexes with at least one near-native solution in the top N solutions, and number of near-native solutions in the top N solutions (between brackets, total number of near-native solutions is 109) for different values of N. For example, when N=15, for 23 out of 31 complexes, Kendall-τ ranks an acceptable solution 15 or better, and over all the targets, 48 near-native solutions are ranked 15 or better.

top N	Hex	Kendall-τ	sumRank	medRank	b-RF	b-PART	b-JRip	b-J48	b-NaiveBayes
5	13(15)	13(20)	13(18)	9(16)	11(20)	12(19)	11(14)	10(12)	12(14)
10	18(30)	19(37)	20(35)	18(36)	19(36)	17(31)	20(29)	18(27)	13(20)
15	22(39)	23(48)	21(46)	23(51)	23(50)	20(39)	22(41)	23(40)	20(37)
25	26(58)	27(71)	25(67)	26(72)	26(69)	27(60)	27(62)	25(61)	25(64)
35	28(77)	29(90)	28(86)	28(85)	28(86)	29(85)	30(82)	28(78)	28(87)
45	30(96)	31(105)	30(105)	30(105)	31(106)	30(102)	30(102)	29(99)	30(106)

As expected, the bagging versions of the machine learning approaches are always better than the non-bagging versions. Consequently, in the subsequent experiments, only bagging versions will be used. We also found that sumRank (AUC 0.79) and medRank (AUC 0.782) give slightly better results than the Weka bagging approaches.

3.1 Ranking Hex Docking Solutions

For each of the 31 complexes of the benchmark (those complexes for which Hex doesn't place a good solution in the 50 top-ranking solutions were excluded), the 50 top-ranking conformations given by Hex were re-ranked using the Weka algorithms, *sumRank*, *medRank*. We also used Kendall-τ to bag all the ranks proposed by the seven bagging approaches. When two conformations have the same rank for a given method, these two conformations are ordered by their Hex scores. For each method, we counted the total number of solutions within $10\mathring{A}$ of the native structure in the top N solutions, and the number of complexes for

which at least one such solution is found in the top N, with N varying from 1 to 50 (see Table 1).

Evolutionary-based scoring functions increase the number of complexes for which a near-native solution is found in the top N solutions, together with the overall number of near-native solutions ranked N or better. Results show that Kendall-τ meta-bagging method performs as well or better than any other method for every value of N, which makes it the best performing method on the overall test.

The targets of the benchmark have been classified in *Difficult*, *Medium* and *Easy* by its authors depending mostly on the RMSD between the partners composing the complex in their free and bound forms. We have slightly revised this classification by considering as *Difficult* the complexes having more than 2 partners (since our method was trained strictly on heterodimers), and complexes at which interfaces are found large ligands such as FAD or NAD (since such molecules are ignored by our method). Table 2 lists the rank of the first acceptable solution for each method. Results show that on average this rank is better for Kendall-τ than for any other method. Looking only at results obtained for *easy* targets with our method, for all targets an acceptable solution in found in the top 10, and the average rank is 3.7 (against 5.9 for Hex). Finally, for 4 out of these 14 targets, an acceptable solution is found at rank 1, and for 11 out these 14 targets, an acceptable solution is found in the top 5.

3.2 Further Insight on Difficult Targets

We then gave further attention to the 12 targets for which we were unable to rank an acceptable solution in the top 10. Removing the structural redundancy (conformations closer than 5 $\mathring{A}$ RMSD) allows to place a near-native solution in the top 10 for 3 of these targets (1RV6, 2A9K, 1CLV).

Amongst the remaining 9 complexes, 3 include large non-peptidic molecules at the interface (1LFD, 1R6Q, 2O3B), 1 has more than two subunits (3CPH), and 5 exhibit important fold modifications upon complexation (1JK9, 2A5T, 2I9B, 2IDO, 2Z0E). Although some complexes presenting more than two subunits (1RV6, 2ABZ), or large molecules at the interface (1F6M, 1ZHH, 2A9K, 2G77, 2J7P) can be successfully predicted, our method cannot be reliably used in these cases. Fold modifications are more problematic since they cannot be easily predicted from the structures of the partners alone.

From a biologist's point of view, a docking prediction is useful if it allows the residues lying at the interface and the interactions between those residues to be identified. Inspection of the top 10 conformations of the 9 unsuccessful targets show that for 8 of them, a solution with the two correct epitopes is present in the top 10. Figure 1 shows some biologically interesting conformations of three different complexes which are ranked within the first 10 conformations by the Kendall-τ method. For these complexes, none of the top 10 conformations is near-native. However, inspection of Figure 1 shows that the interacting regions

Table 2. Rank of the first acceptable solution found for each docking target according to each approach

Target	Hex	Kendall-τ sumRank	medRank	b-RF	b-PART	b-JRip	b-NaiveBayes	b-J48
Difficult and medium targets								
1CLV	2	12	17	12	14	6	12	6	13
1F6M	31	5	8	9	7	2	10	4	20
1JK9	17	17	9	17	7	20	1	36	10
1LFD	33	29	33	30	21	20	20	26	31
1R6Q	1	16	27	23	14	15	17	3	7
1RV6	24	12	4	7	8	21	48	26	50
1ZHH	6	10	3	3	5	31	21	1	29
2A5T	17	45	47	46	43	24	22	46	12
2A9K	1	11	16	12	34	44	10	11	1
2ABZ	8	8	11	13	5	14	4	13	9
2G77	4	2	3	7	1	1	1	3	1
2I9B	11	13	22	12	13	19	8	11	47
2IDO	49	29	38	37	17	16	35	28	12
2J7P	16	7	3	5	8	9	10	13	10
2O3B	12	39	36	37	36	12	7	37	32
2Z0E	37	22	20	14	26	16	27	15	14
3CPH	38	24	28	28	39	26	7	25	3
Easy targets								
1FFW	8	1	1	1	4	2	4	2	9
1FLE	4	1	1	2	10	1	5	5	5
1GL1	3	5	6	8	2	3	4	1	3
1H9D	14	1	1	1	2	1	8	22	3
1JTG	2	4	9	6	1	10	25	3	3
1OC0	9	1	1	2	1	10	1	4	37
1OYV	2	4	3	7	1	1	10	3	10
1SYX	3	4	7	3	17	5	8	20	4
2AYO	11	2	4	6	9	1	11	11	4
2J0T	7	10	7	18	13	46	34	17	7
2OUL	1	2	2	3	7	5	2	25	11
3D5S	2	7	5	6	5	7	1	12	22
3SGQ	3	2	6	6	1	3	2	2	1
4CPA	5	8	4	4	6	1	1	2	8
av.	12,29	11,39	12,32	12,42	12,16	12,65	12,13	13,97	13,81

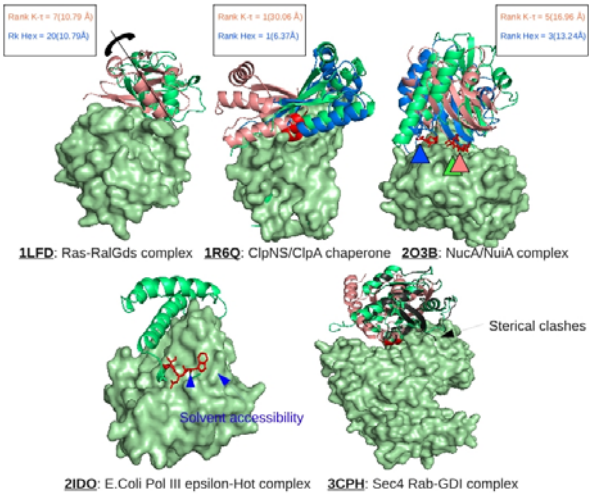

Fig. 1. Three examples for which meta-bagging of Voronoi fingerprints correctly identifies the location of the native interface, although with a relatively high RMSD (wheat) of solutions obtained by the meta-bagging scoring algorithm (wheat) and Hex (blue), compared to the correct native structures (green). The catalytic site residues are indicated in red and are located by colored arrows in order to highlight the position of the biological interface. The ranks of the represented conformations using Kendall-τ and Hex are given, and the corresponding RMSD is shown between brackets. Highest ranking near-native solution found by Hex are shown in blue, and native solution in green.

on both monomers have been correctly identified, and even some of the catalytic residues involved in the interaction have been correctly positioned. The complex between Polymerase III epsilon and the Hot protein (PDB code 2IDO) is also a special case in which we were not able to find a near-native solution. In this complex, important residues of the native interface remain accessible to the solvent instead of being occluded by the docking partner.

As shown in Figure 2, Hex is able to rank in the top 20 a near-native conformation of the NR1/NR2A complex, which plays an crucial role into mammalian central nervous system [14], whereas this conformation is ranked 45 by the Kendall-τ meta-bagging method. While the superposition of this solution with the native structure is generally satisfying, closer inspection of the interface moderates this conclusion. More specifically, the ion channel activity of this complex is regulated by Van der Waals contacts between residues Y535 and P527. In the conformation proposed by Hex, this interaction is lost. However, this crucial interaction is correctly predicted in another conformation, ranked 9 by the Kendall-τ method, even though the global RMSD is rather high.

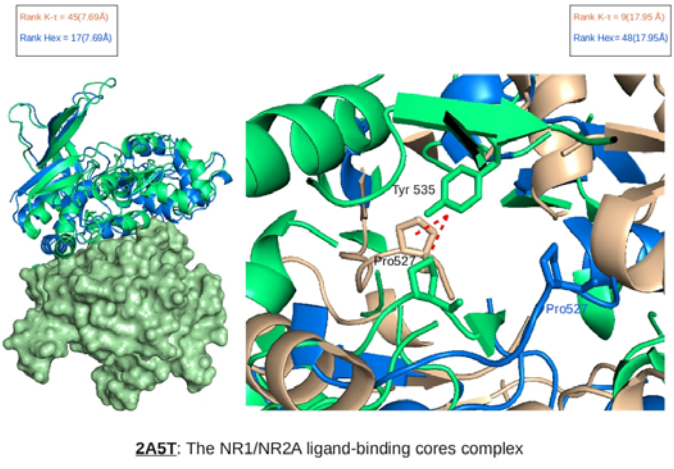

Fig. 2. NR1/NR2A complex (PDB code 2A5T). Left: superposition of native structure (green) and near-native structure given by Hex (ranked 17 by Hex and 45 by Kendall-τ meta-bagging method). Right: zoom on the core of the interface, including a conformation ranked 9 by the Kendall-τ meta-bagging algorithm (wheat) and 48 by Hex. The key interaction between tyrosine (TYR535) and Proline (PRO527) is represented. TYR 535 belongs to the A chain of the complex/conformations and have thus been represented only once, since these chains are identical and superposed for all conformations.

4 Conclusion

Despite the fact that the Weka machine learning approaches used in this work were designed to separate examples into positive and negative instances rather than to rank them, for the current learning set, we find that using probabilities for ranking gives as good results as those achieved by the evolutionary ROC-based learning approaches. The different machine learning approaches let us find at least one acceptable solution within the top 10 for all easy targets.

The improvement of the rank of the first quasi-native conformation can seem modest. However, it should be borne in mind that this method is intended to be used by experimental biologists as a guide for experimentation. Laboratory-based experiments can be costly and time-consuming. Therefore, reducing the number of experiments is a crucial objective, and will give the method an effective practical interest. Our method not only guarantees that a near-native conformation is present in the top 10 solutions, it even places a near native solution in the top 5 for 11 out of 14 easy targets. For the experimenter, this means that the number of experiments can be cut by half with a risk of "losing" predictions for only 3 targets. Moreover, we show that the conformations selected by our method, even when their structural similarity with the native conformation is low, have a real biological relevance. Our method also ranks a near-native conformation in the top 10 for 8 of the 17 difficult targets, and a

conformation with the correct epitopes for 8 of the remaining 9 targets. For the experimenter, this means that considering the top 10 conformations guarantees having the correct epitopes (which is sufficient for directed mutagenesis for example) for 30 of the 31 targets. However, the heterogeneity observed in some of the individual predictions means it is not obvious how to reach the best overall ranking from the different approaches.

The problem we aimed to address in this paper is how best to aggregate such lists of ranks from an ensemble of ranking functions. We propose that the notion of Kendall-τ distance provides one way to achieve this goal. The driving heuristic in the Kendall-τ approach is to find good cyclic movements of ranks towards the median permutation in order to obtain a consensus ranking which minimises the Kendall-τ distance between the individual rankings. Thus, the Kendall-τ approach quickly dominates all of the individual approaches tested here by aggregating the best predictions from each of them. The present study focused on the 50 best conformations found by Hex because the Kendall-τ approach is currently limited to work with permutations containing the same elements. Lifting this limitation would allow the Kendall-τ approach to focus on the best predictions from each learning algorithm, either directly by comparing Kendall-τ distances or, as was done here, by finding the optimal combination of predictions from the independent bagging algorithms.

References

1. Ailon, N., Charikar, M., Newman, A.: Aggregating inconsistent information: Ranking and clustering. J. ACM 55(5) (2008)
2. Azé, J., Roche, M., Sebag, M.: Bagging evolutionary roc-based hypotheses application to terminology extraction. In: Proceedings of ROCML (ROC Analysis in Machine Learning), Bonn, Germany (2005)
3. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissing, H., Shindyalov, I., Bourne, P.: The protein data bank. Nucleic Acids Research 28, 235–242 (2000)
4. Bernauer, J., Azé, J., Janin, J., Poupon, A.: A new protein-protein docking scoring function based on interface residue properties. Bioinformatics 5(23), 555–562 (2007), http://bioinformatics.oxfordjournals.org/cgi/content/full/23/5/555
5. Betzler, N., Fellows, M.R., Guo, J., Niedermeier, R., Rosamond, F.A.: Fixed-parameter algorithms for kemeny scores. In: Fleischer, R., Xu, J. (eds.) AAIM 2008. LNCS, vol. 5034, pp. 60–71. Springer, Heidelberg (2008)
6. Blin, G., Crochemore, M., Hamel, S., Vialette, S.: Median of an odd number of permutations. Pure Mathematics and Applications 21(2), 161–175 (2011)
7. Boissonnat, J.D., Devillers, O., Pion, S., Teillaud, M., Yvinec, M.: Triangulations in CGAL. Comput. Geom. Theory Appl. 22, 5–19 (2002)
8. Bourquard, T., Bernauer, J., Azé, J., Poupon, A.: Comparing Voronoi and Laguerre tessellations in the protein-protein docking context. In: Sixth International Symposium on Voronoi Diagrams (ISVD), pp. 225–232 (2009)
9. Bourquard, T., Bernauer, J., Azé, J., Poupon, A.: A Collaborative Filtering Approach for Protein-Protein Docking Scoring Functions. PLoS ONE 6(4), e18541 (2011), doi:10.1371/journal.pone.0018541

10. Camacho, C.: Modeling side-chains using molecular dynamics improve recognition of binding region in capri targets. Proteins 60(2), 245–251 (2005)
11. Devos, D., Russell, R.B.: A more complete, complexed and structured interactome. Current Opinion in Structural Biology 17, 370–377 (2007)
12. Dominguez, C., Boelens, R., Bonvin, A.: HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125(7), 1731–1737 (2003)
13. Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the Web. In: World Wide Web, pp. 613–622 (2001), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.8702
14. Furukawa, H., Singh, S., Mancusso, R., Gouaux, E.: Subunit arrangement and function in nmda receptors. Nature 438(7065), 185–192 (2005)
15. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explorations 11(1), 10–18 (2009)
16. Hwang, H., Vreven, T., Janin, J., Weng, Z.: Protein-protein docking benchmark version 4. 0. Proteins 78(15), 3111–3114 (2010)
17. Janin, J., Henrick, K., Moult, J., Eyck, L., Sternberg, M., Vajda, S., Vakser, I., Wodak, S.: CAPRI: a Critical Assessment of PRedicted Interactions. Proteins 52(1), 2–9 (2003)
18. Janin, J., Wodak, S.: The Third CAPRI Assessment meeting. Structure 15, 755–759 (2007)
19. Kendall, M.: A New Measure of Rank Correlation. Biometrika 30(1/2), 81–93 (1938), http://www.jstor.org/stable/2332226
20. Kenyon-Mathieu, C., Schudy, W.: How to rank with few errors. In: Johnson, D., Feige, U. (eds.) STOC, pp. 95–103. ACM (2007)
21. Komatsu, K., Kurihara, Y., Iwadate, M., Takeda-Shikata, M.: Evaluation of the third solvent clusters fitting procedure for the prediction of protein-protein interactions based on the results at the capri blind docking study. Proteins 52(1), 15–18 (2003)
22. Levy, E., Pereira-Leal, J., Chothia, C., Teichmann, S.: 3d complex: a structural classification of protein complexes. PLoS Comput. Biol. 2(11) (2006)
23. Mihalek, J., Res, I., Lichtarge, O.: A structure and evolution-guided monte carlo sequence selection strategy for multiple alignement-based analysis of proteins. Bioinformatics 22(2), 149–156 (2006)
24. Mintseris, J., Weng, Z.: Atomic contact vectors in protein-protein recognition. Proteins 53(3), 214–216 (2003)
25. Murzin, A., Brenner, S., Hubbard, T., Chothia, C.: Scop: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540 (1995)
26. Ritchie, D.: Evaluation of protein docking predictions using hex 3.1 in capri rounds 1 and 2. Proteins 52(1), 98–106 (2003)
27. Ritchie, D., Venkatraman, V.: Ultra-fast FFT protein docking on graphics processors. Bioinformatics 26(19), 2398–2405 (2010)
28. van Zuylen, A., Williamson, D.P.: Deterministic algorithms for rank aggregation and other ranking and clustering problems. In: Kaklamanis, C., Skutella, M. (eds.) WAOA 2007. LNCS, vol. 4927, pp. 260–273. Springer, Heidelberg (2008)