BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 5 2007, pages 555–562
doi:10.1093/bioinformatics/btl654

Structural bioinformatics

A new protein–protein docking scoring function based on interface residue properties

J. Bernauer[1], J. Azé[2], J. Janin[1] and A. Poupon[1]*

[1]Yeast Structural Genomics, IBBMC UMR CNRS 8619, Bâtiment 430 and [2]Equipe de Bioinformatique, LRI UMR CNRS 8623, Bâtiment 490, Université Paris-Sud, 91405 ORSAY, France

Received on September 15, 2006; Revised and accepted on December 21, 2006

Advance Access publication January 18, 2007

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Protein–protein complexes are known to play key roles in many cellular processes. However, they are often not accessible to experimental study because of their low stability and difficulty to produce the proteins and assemble them in native conformation. Thus, docking algorithms have been developed to provide an *in silico* approach of the problem. A protein–protein docking procedure traditionally consists of two successive tasks: a search algorithm generates a large number of candidate solutions, and then a scoring function is used to rank them.

Results: To address the second step, we developed a scoring function based on a Voronoï tessellation of the protein three-dimensional structure. We showed that the Voronoï representation may be used to describe in a simplified but useful manner, the geometric and physico-chemical complementarities of two molecular surfaces. We measured a set of parameters on native protein–protein complexes and on decoys, and used them as attributes in several statistical learning procedures: a logistic function, Support Vector Machines (SVM), and a genetic algorithm. For the later, we used ROGER, a genetic algorithm designed to optimize the area under the receiver operating characteristics curve. To further test the scores derived with ROGER, we ranked models generated by two different docking algorithms on targets of a blind prediction experiment, improving in almost all cases the rank of native-like solutions.

Availability: http://genomics.eu.org/spip/-Bioinformatics-tools-

1 INTRODUCTION

The three-dimensional structure of a protein–protein complex is a crucial information for its functional study. Whereas structural genomics projects have considerably increased our knowledge of individual protein structures, protein–protein complexes are still beyond the reach of high-throughput methods. Reliable, fast and automatic docking algorithms that can assemble individual proteins into complexes are therefore of great value.

A docking procedure comprises two tasks, generally consecutive and largely independent. The first is a rigid-body search over the rotational/translational degrees of freedom. It generates a large number of candidate solutions where the two partners contact each other in many different orientations, avoiding steric clashes. Then, the best solutions are selected by evaluating a score. Scoring functions express the geometric complementarity of the two molecular surfaces in contact, and also the strength of the interaction, based on the physico-chemical characteristics of the amino acids in contact with each other. Most procedures handle the geometric and physical-chemical criteria separately.

The formation of a complex often induces changes in the structure of both partners. These changes may concern only the side chain conformation of amino acid residues at the interface, or they may imply motions of the protein backbone, the nature and amplitude of which remains very difficult to predict. In some cases, an amino acid substitution in one of the partners can cause gross changes in the complex even though it hardly affects the structure of the protein itself (Graille *et al.*, 2005). This makes 'unbound' docking predictions much more difficult than 'bound' ones: bound docking starts from protein structures taken from the complex and ignores these motions, whereas unbound docking uses the structures of the free partners and must take the physico-chemistry into account.

One major barrier for the study of macromolecular assemblies is their sheer complexity. The problem cannot be solved without simplifications, and even with those, the best performing algorithms consume great amounts of CPU time. Biological information is often available for a given protein–protein structure, requiring 'human' post-processing (Mendez *et al.*, 2005). Yet, postgenomic functional studies need a procedure that can search for plausible protein–protein complexes in a complete genome with thousands of genes. This requires the procedure to be: (i) very reliable, which means that the score of the best solution is high only if the two proteins do form a complex, and that the best solution is close to the native complex; and (ii) very fast, since the inspection of a whole genome requires the modeling of many hundreds of thousands potential complexes.

In this work we use the Voronoï tessellation as a descriptor of the protein structure (Poupon, 2004). The tessellation, which is based on amino acid residues rather than individual atoms has been shown to yield valuable mathematical results on the molecular packing (Soyer *et al.*, 2000), molecular recognition

*To whom correspondence should be addressed.

(Bernauer *et al.*, 2005) and other structural properties of proteins. We show here that its use in the description of protein–protein interfaces leads to valuable scoring functions for docking algorithms.

To generate a scoring function, we choose a limited number of parameters that can be measured on the Voronoï tessellation of a complex. These parameters are measured in a set of native-like protein–protein complexes and in a set of decoys, and used as attributes in statistical learning methods. Three different types of learning algorithms are compared: a logistic function, SVM and a genetic algorithm called ROGER. ROGER gives the best results by far, and we use its score to re-rank models of protein–protein complexes generated by two different docking algorithms on targets of Critical Assessment of PRedicted Interactions (CAPRI), a blind prediction experiment designed to test docking procedures (Janin *et al.*, 2003). The results show that the ROGER score improves the ranking of native-like solutions and suggest that it will be of great value in early steps of a fully automated docking procedure.

2 METHODS

2.1 The Voronoï construction

Voronoï diagrams, also known as Dirichlet or Thiessen tessellations, have been used in many fields of sciences. Given a set of points in space (centroids), the Voronoï tessellation divides the space into Voronoï cells centered on each point. A Voronoï cell includes all points of space that are closer to the cell centroid than to any other centroid, and it is the smallest polyhedron defined by bisecting planes between its centroid and all others. The Delaunay tessellation is obtained by tracing the vertices joining centroids, which have a common face in their Voronoï cells. The two tessellations are dual from each other. They are uniquely related, and for efficiency, it is best to compute the Delaunay tessellation first and derive Voronoï cells from it.

We performed the computation by using as centroids the center of mass of amino acid side chains, including Cα (Fig. 1). Our procedure uses the Computational Geometric Algorithms Library (CGAL), which implements an incremental randomized algorithm (Boissonat *et al.*, 2002) of optimal $O(n^2)$ complexity, n being the number of centroids. Solvent was modeled around the protein to prevent surface residues from having unbound Voronoï cells (Fig. 1). We placed spheres of radius 6.5 Å on a water-like lattice (Soyer *et al.*, 2000) to give solvent cells a volume similar to the average residue Voronoï cell.

The tessellation of a protein–protein complex leads to the following definitions:

(i) Two residues are neighbors if their Voronoï cells share a common face.
(ii) A residue belongs to the protein interior if all its neighbors are residues of the same protein.
(iii) A residue belongs to the protein surface if one or more of its neighbors is solvent.
(iv) A residue belongs to the protein–protein interface if one or more of its neighbors belongs to the other protein.
(v) An interface residue belongs to the core of the interface if none of its neighbors is solvent.
(vi) The cell facets shared by residues of both proteins constitute the interface.

2.2 Training set

The training set consists in two subsets: positive examples, complexes of known 3D structure; and negative examples generated from the positive examples using a docking procedure.

2.2.1 Complexes of known 3D structures The set of models on which were used as positive examples in the training set for learning methods was computed from the 2004 release #1 of the Protein Data Bank (PDB) (Berman *et al.*, 2000) (Fig. 2). First, we used a BioPython module (Hamelryck and Manderick, 2003) to extract entries reporting X-ray structures with resolution higher than 3 Å and with two polypeptide chains longer than 20 residues, putative partners in a complex. Then we searched the PDB with Blastp (Altschul *et al.*, 1990) for entries containing the free partners, and retained those having more than 95% identity. Hierarchical clustering by R software (Ihaka and Gentleman, 1996) then gave a non-redundant set of 102 complexes (22 Unbound/Unbound and 80 Bound/Unbound).

2.2.2 Decoys The decoys were generated by applying a docking algorithm to each complex. The starting proteins structures were those of the free partners when available, or taken from the complex for the bound partner of unbound/bound complexes.

To generate non-native solutions (decoys), we used the program DOCK (Cherfils *et al.*, 1991; Janin and Wodak, 1985). DOCK explores the solution space using five angles and a distance as rigid-body parameters. The program was run in grid mode sampling the whole range of the five angles in 10° steps; 200 solutions were randomly chosen among the 18.10^6 thus generated. We checked that they were all non-native, that is, far from the X-ray structure.

2.3 Training attributes

We chose properties of interface residues and residue pairs as training attributes. The number of parameters that may be used in training is limited by the size of the training set. With only 102 native complexes in the set, attributes could be attached to each of the 20 amino acid residue types, but not to each pair. To define pair attributes, we grouped residue types in six categories: hydrophobic H (ILFMV), aromatic Φ (FYW), positively charged + (HKR), negatively charged − (DE), polar P (NQ) and small S (AGSTCP). The final set of attributes includes 84 parameters in five classes:

(i) P1 The Voronoï interface area, measured by summing the areas of the cell faces constituting this interface (1 parameter)
(ii) P2 the total number of core interface residues (1 parameter)
(iii) P3 the number fraction of each type of core interface residues (20 parameters)

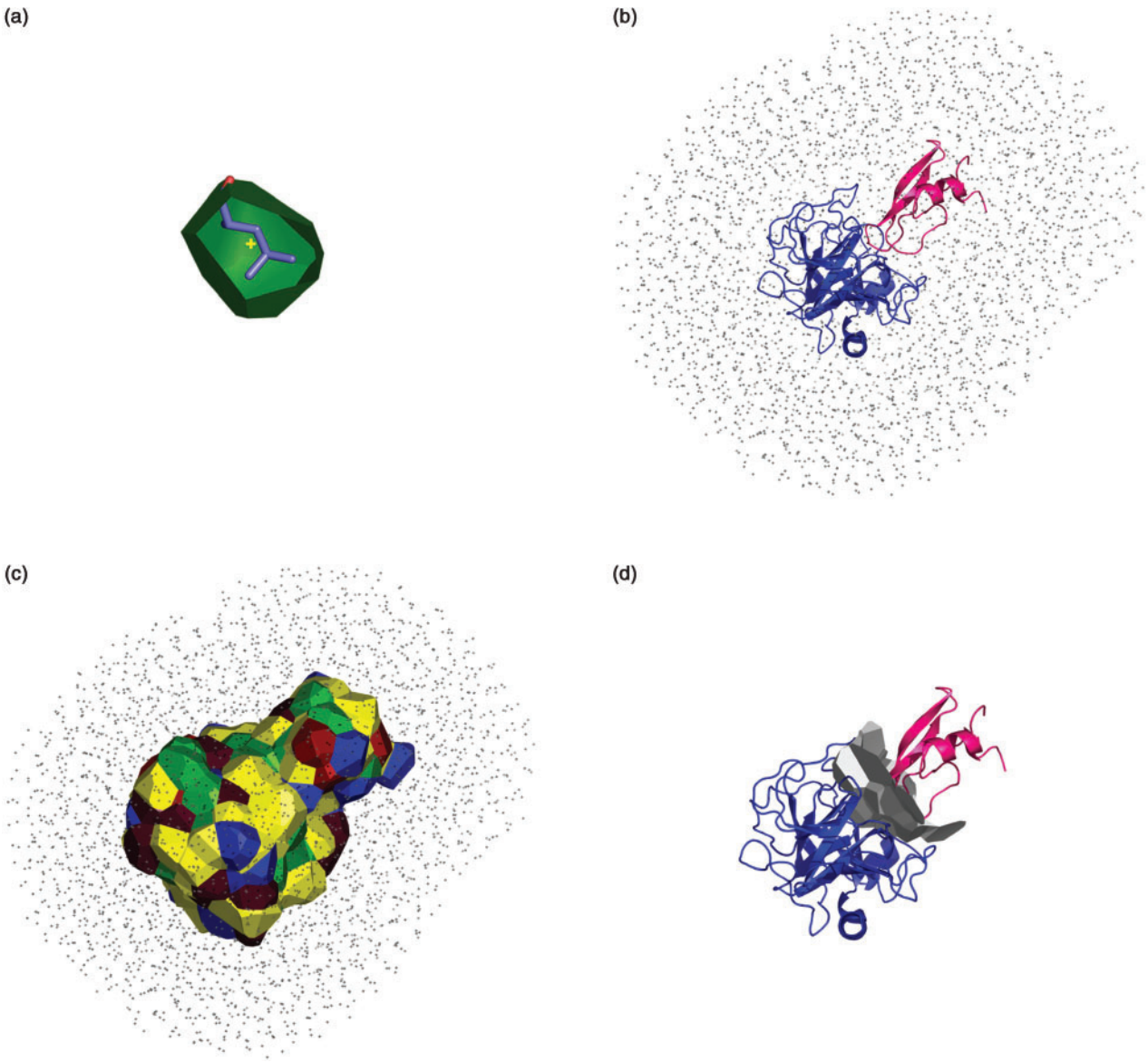

Fig. 1. Voronoi description of protein–protein interfaces. (a) The Voronoï cell of a leucine residue in complex 1p2k. The cross marks the centroid used for constructing the cell (b) Complex 1p2k; the two protein chains are in blue and pink, and solvent is drawn as dots (c) Voronoï envelope of complex 1p2k (d) Facets in gray are shared by Voronoi polyhedra of residues of the two proteins, representing the interface of the complex.

(iv) P4 the mean volume of the Voronoï cells for the core interface residues of each type (20 parameters)
(v) P5 the number fraction of pairs of each category (21 parameters)
(vi) P6 the mean centroid-to-centroid distance in pairs of each category (21 parameters)

2.4 Learning methods

The values of the 84 parameters were measured on the 102 native complexes (the positive examples) and on the decoys (the negative examples) of the training set. These values were then used to train statistical learning procedures that optimize score functions to best discriminate between the native and the decoy models.

2.4.1 Logistic function A logistic function is a linear combination of the parameters with weights optimized to yield a value close to 1 on the native models and 0 on the decoys. The vector of weights $W[w_i]$ was estimated on the learning set by the maximum likelihood method using the general linear model (GLM) of the R software (Ihaka and Gentleman, 1996).

2.4.2 SVM: Support vector machines Support vector machines (SVM) aim to divide a high-dimensional space into regions containing only positive examples or only decoys,

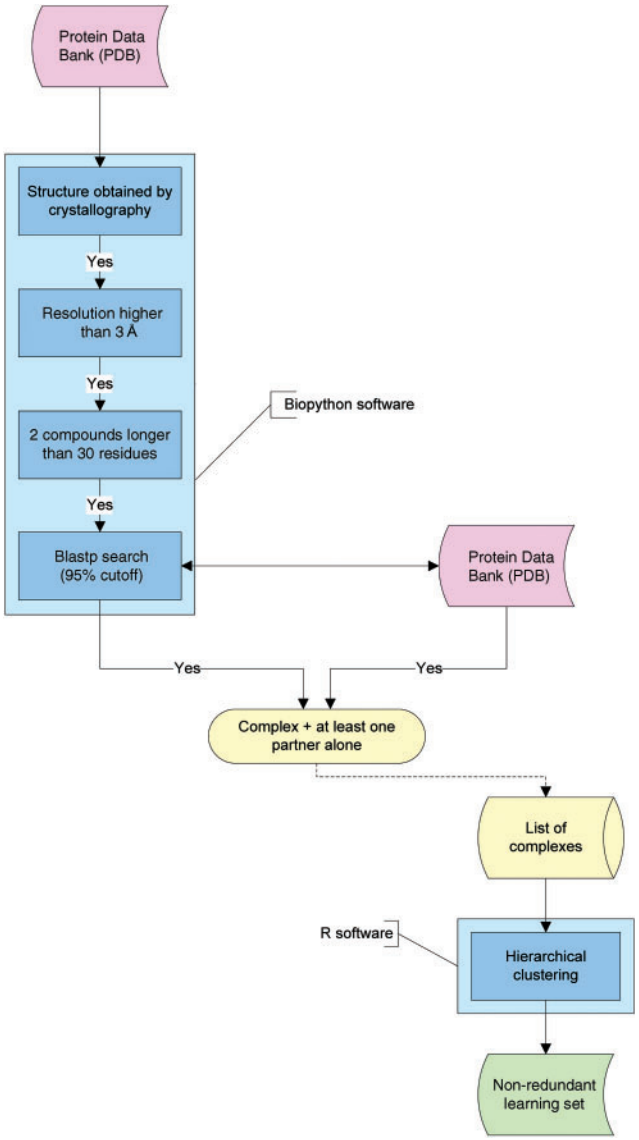

Fig. 2. Procedure for extracting the training set from the PDB.

each represented by a point with the values of the different parameters as coordinates. The functions defining these regions, which can be of different types, are called kernels. We tested SVM with linear, polynomial and radial basic function (RBF) kernels. Computations were carried out using SVMTorch (Collobert and Bengio, 2001).

2.4.3 ROGER: a ROc based GEnetic learner The Receiver Operating Characteristics (ROC) procedure is often used evaluate learning procedures by cross-validation on examples taken from the learning set. Plotting the proportion of true positives against the proportion of false positives yields a ROC curve, and the area under that curve is the ROC criterion.

ROGER (Sebag *et al.*, 2003) uses a genetic algorithm to find a family of functions that optimizes the ROC criterion. In our implementation, the function optimized was the sum of the weighted and centered parameters:

$$f(x) = \sum_i \omega_i |x_i - c_i|$$

Thus, two values are determined for each attribute: a central value c_i and a weight ω_i. We applied a ten-fold cross-validation procedure by forming 10 groups of models, each excluding 10% of the training set, and repeated the training procedure 21 times for each set. This was required by the heuristic nature of the genetic algorithm, which can end in a local minimum. Thus, the learning procedure generated 210 functions. For a given complex, the 210 functions were evaluated, and the median value was retained as score.

2.5 Missing data in learning

Interfaces in the learning set contained on average 18 core residues per partner protein. Thus, the least abundant residue types and some of the category pairs were absent from many members of the set, leading to a null value of the amino acid frequency (parameter class 3) and to missing values of the volumes (class 4) and of the pair parameters in classes 5 and 6. Because learning methods including the logistic functions and ROGER, generally cannot handle missing data, we replaced each missing value by either:

(i) The mean value on the whole set
(ii) The median value on the whole set
(iii) The mean value in the category (native or decoy) to which the example belong
(iv) The median value in the category

Missing values also had to be replaced when scoring a model generated by docking in the test phase. Here again, many combinations are possible: replacing by the mean or median values of the whole learning set, of the whole test set, by category or not.

In this study, we replaced missing values by their median on the whole learning set during both the learning procedure and the test phase. We tested other alternatives and found that they performed less well, possibly because our parameters have non-Gaussian distributions.

3 RESULTS AND DISCUSSION

3.1 Performance of the learning procedures

The ROC curve was evaluated on the training set for four different scores: the sum of the mean square deviation of the attributes from their mean values, the logistic function, the classification made with SVMs, and the scoring function obtained with ROGER. A perfect selection (100% true positives and no false positives) should make the area under the ROC curve AUC equal to 1; a random selection yielding true positives and false positives in equivalent numbers should have an AUC of 0.5.

Taking the sum of the mean square deviations as a score yielded an AUC close to 0.5, indicating that individual parameters discriminate very poorly between the native and

the decoys, and that they must be properly weighted and combined, which is what a learning procedure aims to. With the logistic function, the area under the ROC curve increased to 0.85, indicating a better discrimination.

ROGER and the SVMs did much better, achieving AUC of 0.98 and 0.99, respectively. More important perhaps, the initial slope of the ROC curve was very steep in both cases, implying that ROGER and the SVMs had very few false positives among their best scoring solutions. Thus, both learning procedures were successful, and we retained the ROGER score for further studies as the SVMs only give a binary classification, ill-suited to our problem of 'finding a needle in a hay stack'. As docking generates hundreds of thousands of non-native solutions along with only a few near-native ones, a binary classification would still yield many false positives. In contrast, the ROGER score could easily be combined with other functions if needed.

3.2 Weight and central value of the attributes in the ROGER scoring functions

The ROGER score of a docking model is the median value of 210 functions generated in separate trainings as described under Methods. The role of an attribute in the function depends on both the absolute value and the sign of its weight.

Because the algorithm is designed to minimize the score of native-like solutions, the most significant attributes are those having large weights and/or central values that are very different from the median value. The greater the difference between the central and median value, the more influence a parameter has.

Figure 3 and 4 show that the largest positive weights are for the interface area and number of interface residues, two attributes that measure the size of the interfaces. They are highly correlated and both have central values that are much larger than the median values observed in native complexes. Thus, large interfaces are preferred. Residue and pair frequencies also have central values that are larger than the median. The weights are small and of both signs for residue frequencies, which govern the amino acid composition of the interface. They are larger and mostly positive for pair frequencies, especially for pairs involving hydrophobic residues, implying that the score favors hydrophobic contacts.

In contrast, the central value is close to the median for most of the residue volumes and pair distances. The residue volumes and the pair distances involving hydrophobic residues have large negative weights, and therefore, they are important attributes. A tight packing of hydrophobic residues at the interface, which makes their volume and the distance to neighbors less than the average, is particularly favored by the ROGER score.

3.3 Results on the targets of CAPRI rounds 3–6

We tested the scoring functions on models of the targets of CAPRI Rounds 3–6 generated by two docking programs: DOCK (Cherfils *et al*., 1991; Janin and Wodak, 1985) run as in the learning set, and HADDOCK (Dominguez *et al*., 2003).

Models generated by the two programs were grouped into classes depending on the fraction Fnat of the residue–residue contacts in the native structure that were present in the model, and on the fraction Fint of the interface residues that were correctly predicted. As described in the legend of Table 1, classes 1 to 4 were defined by ranges of Fnat; class 5 had $Fnat = 0$ (no native contact), but $Fint > 0$ (some interface residues correctly predicted); class 6 was for incorrect models.

As all classes were not represented for each target, Table 1 cites the 'best class' that was present in a given data set. Taking target 11 as an example, the best model in the data set generated by HADDOCK had $Fnat = 0.54$. Thus, the best class in that data set was 2 ($0.5 < Fnat < 0.75$).

3.3.1 Re-ranking HADDOCK solutions Models for five CAPRI targets generated by HADDOCK (Dominguez *et al*., 2003) were kindly communicated by Dr A. Bonvin. Their ROGER score was computed and the result of their re-ranking is given in Table 1. The 'rank1' column in Table 1 indicates the rank of the first best class solution, and the 'Orig. rank' column indicates the original rank, given by HADDOCK, for the same solution. For all but one target, the rank given by our function is better than the rank given by HADDOCK.

With Target 11, a model of the best class (class 2), re-ranked 4, and one of the next best class (class 3) was 1st. The top 50 ROGER scores included 4 models of class 2, and 44 models of class 3. Thus very few if any of the top 50 were false positives. With Targets 13 and 14, the data set contained class 1 models that were ranked 1st or 2d by ROGER, although the top 50 contained more false positives than for Target 11. The results were less satisfactory on Targets 12 and 15, no model of the best class present in these data sets scoring in the top 50.

3.3.2 Re-ranking DOCK solutions We ran a complete grid search of the five angles in steps of 10 degrees on ten of the CAPRI targets of Rounds 3–6. The models generated by DOCK were clustered, the average position in each cluster was retained, and its ROGER score was evaluated.

Table 1 shows that the DOCK data sets were generally poorer than for HADDOCK. None contained models of class 1 or 2. The best models had $Fnat < 0.5$ (class 3 or 4), or even $Fnat = 0$ (class 5) in the case of Target 15. Nevertheless, the re-ranking with ROGER performed correctly. In all but one data set, the top 50 included models of the best class that was present in the set. The 1st rank was a model of the best class (class 4) in the case of Target 16, and a model of the second best class in four other cases. Moreover, for all targets but one, the rank given by our function to the first best class solution is better than the original rank given by DOCK.

Since Round 10 in CAPRI, it is possible to test scoring functions. The putative solutions generated by participants using docking algorithms are made available to other participants for re-ranking. We have tested our scoring function, and amongst the 10 solutions we have submitted, the best one had 20% of native contacts. This solution was ranked 5 with our scoring function. The fraction of interface residues correctly identified on the receptor and on the ligand were 0.43 and 0.68, respectively, and the RMSD between its interface and the correct one was 5.34 Å. There were better solutions in the original set, the best one having 54% of correct

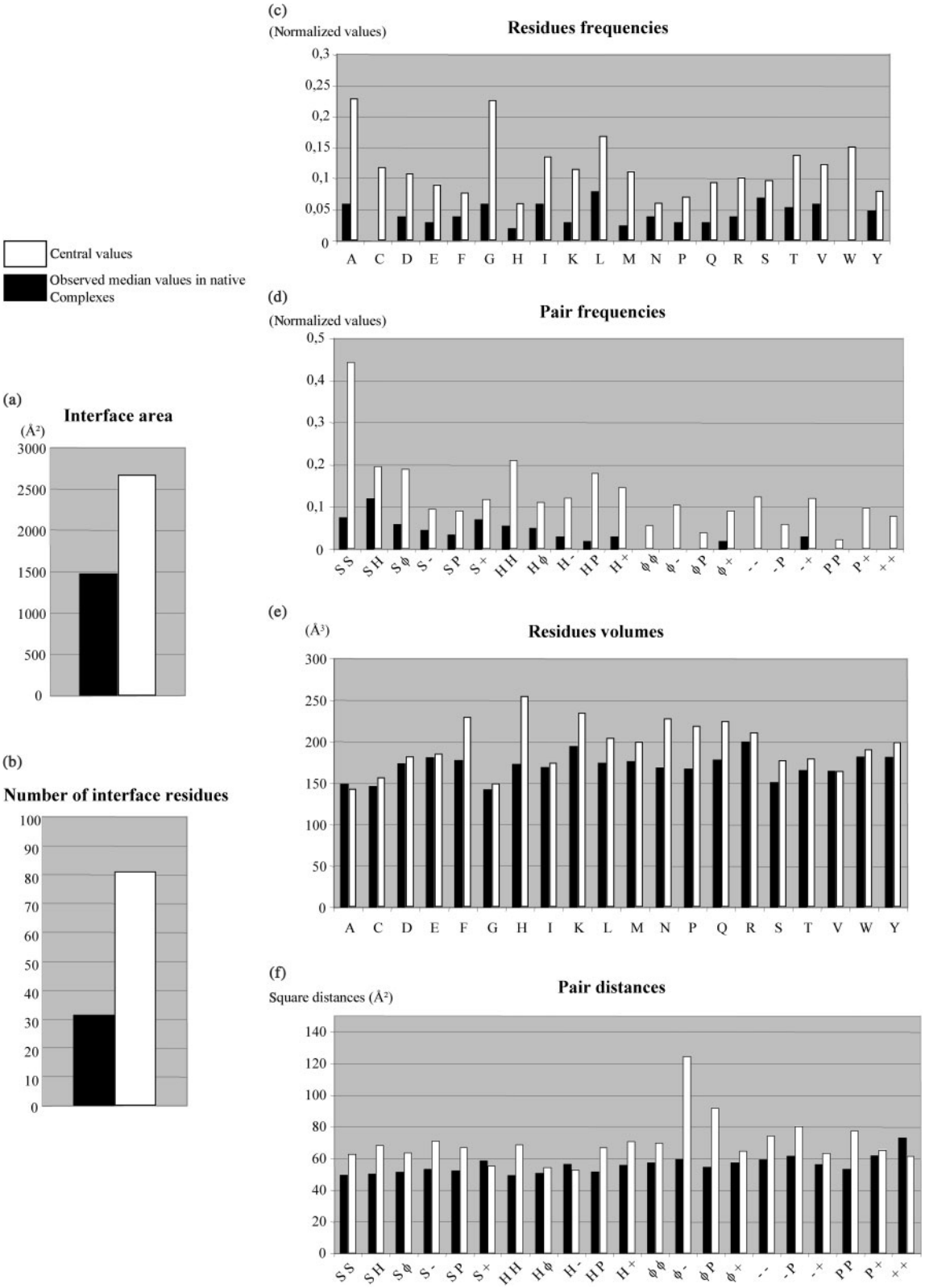

Fig. 3. Comparision of parameter values observed on the learning set and central value obtained after the learning procedure. The root square deviation on central values was 0.17 on average over 10 runs in 5-fold cross validation. (a) P1 Voronoï interface area (Å^2) (b) P2 number of interface residues (c) P3 Residues frequencies (normalized) (d) P5 Pair frequencies (normalized) (e) P4 Residues Voronoï volumes (Å3) (f) P6 Pair square distances.

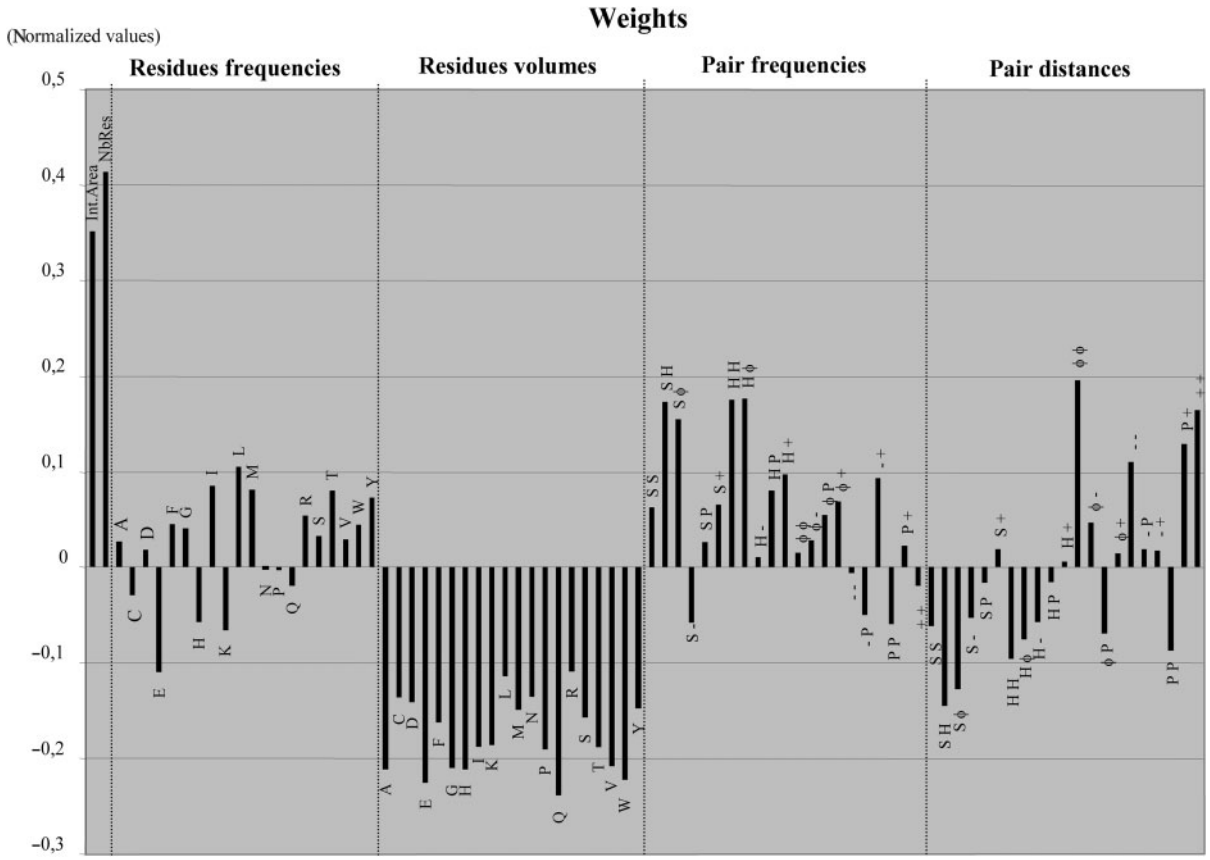

Fig. 4. Weights of the different parameters obtained after the learning procedure. The root square deviation on weights was 0.21 on average over 10 runs in 5-fold cross validation.

native interactions. However, none of these were ranked amongst the 10 first with our scoring function. Moreover, as the coordinates of the real structure are not yet available, we could not know the ranks of these better solutions.

3.3.3 Comparison with other docking programs Recent experiment in protein–protein docking such as CAPRI has shown improvements in docking procedures now able to take certain flexibility into account. Lately scoring functions appeared to play a greater role in rescoring the solutions (Mendez *et al*., 2005), especially during a refinement stage.

They usually rely on shape complementarity supplemented by additional energy terms such as van der Waals Coulomb or desolvation in different and ingenious ways (Carter *et al*., 2005; Daily *et al*., 2005; Fernandez-Recio *et al*., 2005; Law *et al*., 2005; Lee *et al*., 2005; Mustard and Ritchie, 2005; Schneidman-Duhovny *et al*., 2005; Terashi *et al*., 2005; van Dijk *et al*., 2005; Wiehe *et al*., 2005). Taking into account residue conservation and biological information from literature has also been added to the process with varying results (Ma *et al*., 2005; Tress *et al*., 2005).

Lately novel energy scoring functions involving statistical methods have made a breakthrough, first in predicting protein structure in the CASP (Rohl *et al*., 2004) but now also for the protein–protein docking problem with very promising first attempts (Daily *et al*., 2005; Zhang *et al*., 2005).

Our work falls within this scope of statistical method but affords both the measure of geometric criteria and protein structure knowledge described by previous energy-based methods.

4 CONCLUSION

The residue-based Voronoi tessellation provides a convenient low-resolution description of protein structure and protein–protein interfaces. We built a set of parameters derived from that description and a data set of native or decoys models obtained by docking, and used these sets to train the ROGER statistical learning procedure. It returned a score that we tested on docking models of CAPRI targets generated by two different docking programs.

For most targets, a best or second best class solution was found in the top 10 ranking solutions, and in more than half of the cases the top ranking solution belonged to the best or second best class. Moreover, for all targets but one, the rank given to the first best class solution by our scoring function is better than the rank of the same solution given by the original method (DOCK or HADDOCK). We may thus hope that further refinement of the parameters and of the scoring function will have a class 1 or 2 solution as top ranking in all cases. The quick and efficient selection of docking models

Table 1. Summary of rescoring results on DOCK and HADDOCK data sets. Solutions: number of solutions scored, Best class: best class present in set

Target	Nb. Sol.	Best class	rank1	Orig. rank	rank2	rank3	Number of hits
HADDOCK							
11	200	2	4	50	1	9	4
12	200	1	78	119	78	49	0
13	200	1	1	8	7	5	1
14	200	1	2	3	1	3	10
15	200	3	57	29	1	20	0
DOCK							
9	3446	4	4	49	1	3	10
11	32	4	11	28	3	2	3
12	139	3	27	64	14	20	4
13	2263	4	84	1740	7	1	0
14	2460	4	5	244	352	1	2
15	9	5	9	8	1	8	1
16	165	4	1	65	2	5	2
17	72	4	16	39	1	2	1
18	972	3	24	226	1	4	1
19	1979	4	22	94	25	1	2

Definitions of the classes:
Class 1: Fnat > 0.75
Class 2: $0.5 <$ Fnat ≤ 0.75
Class 3: $0.25 <$ Fnat ≤ 0.5
Class 4: $0 <$ Fnat ≤ 0.25
Class 5: Fint > 0 and Fnat $= 0$
Class 6: Fint $= 0$ and Fnat $= 0$
Fnat: fraction of contacts in the native structure that are present at the docking model
Fint: fraction of correctly predicted interface residues

rank1, rank2, rank3: rank1 is the best ROGER rank achieved by a model that belong to the best class present in the set; rank2 is the best rank achieved by of model of the next best class, and so on for rank3.
Orig. rank: original rank of the rank1 solution
Number of hits: number of models with a ROGER rank <50 that belong to the best class present in the set.

based on whole residues will be an asset in whole-genome studies. On the other hand, these models remain crude and we cannot expect to have very accurate solutions unless go back to an atomic model and an appropriate scoring function.

Last, a study that deals only with the scoring function is obviously dependent on the quality of the original data set. Further refinement of the method will require the implementation of a new exploration method that could also be based on Voronoï tessellation.

ACKNOWLEDGEMENTS

We thank Dr A. Bonvin for kindly communicating data sets of CAPRI models, and the EIDIPP program of Action Concertée Incitative IMPBio for financial support.

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
Bernauer,J. *et al.* (2005) A docking analysis of the statistical physics of protein-protein recognition. *Phys. Biol.*, **2**, S17–S23.
Boissonat,J. *et al.* (2002) Triangulations in CGAL. *Comput. Geom. Theory Appl.*, 225–319.
Carter,P. *et al.* (2005) Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins*, **60**, 281–288.
Cherfils,J. *et al.* (1991) Protein-protein recognition analyzed by docking simulation. *Proteins*, **11**, 271–280.
Collobert,R. and Bengio,S. (2001) SVMTorch: Support Vector Machines for large-scale regression problems. *J. Mach. Learn. Res.*, **1**, 143–160.
Daily,M.D. *et al.* (2005) CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock. *Proteins*, **60**, 181–186.
Dominguez,C. *et al.* (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
Fernandez-Recio,J. *et al.* (2005) Improving CAPRI predictions: optimized desolvation for rigid-body docking. *Proteins*, **60**, 308–313.
Graille,M. *et al.* (2005) Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. *J. Biol. Chem.*, **280**, 14780–14789.
Hamelryck,T. and Manderick,B. (2003) PDB file parser and structure class implemented in Python. *Bioinformatics*, **19**, 2308–2310.
Ihaka,R. and Gentleman,R. (1996) R: A language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
Janin,J. *et al.* (2003) CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins*, **52**, 2–9.
Janin,J. and Wodak,S.J. (1985) Reaction pathway for the quaternary structure change in hemoglobin. *Biopolymers*, **24**, 509–526.
Law,D. *et al.* (2005) Progress in computation and amide hydrogen exchange for prediction of protein-protein complexes. *Proteins*, **60**, 302–307.
Lee,K. *et al.* (2005) Study of protein-protein interaction using conformational space annealing. *Proteins*, **60**, 257–262.
Ma,X.H. *et al.* (2005) Biologically enhanced sampling geometric docking and backbone flexibility treatment with multiconformational superposition. *Proteins*, **60**, 319–323.
Mendez,R. *et al.* (2005) Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, **60**, 150–169.
Mustard,D. and Ritchie,D.W. (2005) Docking essential dynamics eigenstructures. *Proteins*, **60**, 269–274.
Poupon,A. (2004) Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr. Opin. Struct. Biol.*, **14**, 233–241.
Rohl,C.A. *et al.* (2004) Modeling structurally variable regions in homologous proteins with rosetta. *Proteins*, **55**, 656–677.
Schneidman-Duhovny,D. *et al.* (2005) Geometry-based flexible and symmetric protein docking. *Proteins*, **60**, 224–231.
Soyer,A. *et al.* (2000) Voronoi tessellation reveals the condensed matter character of folded proteins. *Phys. Rev. Lett.*, **85**, 3532–3535.
Terashi,G. *et al.* (2005) Searching for protein-protein interaction sites and docking by the methods of molecular dynamics, grid scoring, and the pairwise interaction potential of amino acid residues. *Proteins*, **60**, 289–295.
Tress,M. *et al.* (2005) Scoring docking models with evolutionary information. *Proteins*, **60**, 275–280.
van Dijk,A.D. *et al.* (2005) Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins*, **60**, 232–238.
Wiehe,K. *et al.* (2005) ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. *Proteins*, **60**, 207–213.
Zhang,C. *et al.* (2005) Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. *Proteins*, **60**, 314–318.