Protein Surface Conservation in Binding Sites

Nejc Carl, Janez Konc, and Dušanka Janežič*

National Institute of Chemistry, Hajdrihova 19, SI-1000, Ljubljana, Slovenia

Received January 23, 2008

An algorithm is described which uses the conservation of the 3D structure of protein surfaces, as opposed to their sequences, to detect protein−protein binding sites. The protein in which protein−protein binding sites are sought is compared with structures of multiple structurally related proteins and the surface that is conserved at least once is considered to be a part of the binding site. The binding site predictions obtained in this way for a set of protein−protein complexes correspond well with the actual protein−protein binding sites. A comparison of this method with an algorithm using the support vector machine approach for predicting protein−protein binding sites shows structural conservation to be an important characteristic that distinguishes binding sites from the remainder of protein surfaces.

1. INTRODUCTION

Most biological actions of proteins, including their ability to interact with one another, involve binding sites, specific regions of their three-dimensional structure. These have evolved for their ability to bind other molecules effectively and are often conserved in different proteins.[1] Identifying protein−protein binding sites in a protein that is known to interact with other proteins can provide important clues to the function of the protein and can also be used in protein−protein docking studies whose goal is to predict the structures of protein complexes. Knowledge of the location of potential binding sites can be used to reduce the search space required by docking algorithms.[2–5]

It has been shown that for some protein−protein interfaces only a small fraction of the residues contribute most of the binding energy and these so-called "hotspots" tend to have preferred residue types.[6–9] Although in some cases, protein−protein interactions have been successfully predicted based on the conservation of sequence and topology,[10] it has been argued[11] that sequence conservation is rarely sufficient for complete and accurate prediction of a protein−protein interface. Methods that search for conserved topology in proteins are able to detect functional similarity in the absence of sequence homology or fold similarity.[12,13]

We have searched in related proteins for surfaces whose topology and physicochemical properties were conserved and sought to determine whether such surfaces could be correlated with binding sites. We focused our attention on the surfaces of the proteins, employing an algorithm that efficiently identifies surface residues,[14] and in order to compare protein surfaces computationally, we chose protein graphs as a suitable representation.

Protein graphs are structures of vertices and edges, in which vertices are points in three-dimensional space, each labeled with a physicochemical property of the protein at that particular point in space.[13] For efficient comparison of protein graphs, we developed an algorithm which compares protein shapes.[15,16] In a first attempt to predict protein−protein binding sites,[17] an algorithm that detects conservation of protein surfaces by comparing surface structures in related proteins was presented. In contrast, the algorithm described in this report automatically selects multiple structural neighbors with certain characteristics and compares them to the query protein. The advantage of this algorithm, compared to methods that consider only clefts,[13] is that it can detect conserved regions anywhere on a given protein surface and thus is better able to detect protein−protein binding sites, which are often planar.[18] The residues that make up a protein−protein binding site are often adjacent structurally, but can be located on quite different parts of a protein sequence. Conservation of residues in such discontiguous sequences can only be detected with structural alignment methods that are sequence- and fold-independent. Our algorithm[17] requires that the structure of the investigated protein and at least one other, structurally similar, protein are known. These structures can share conserved surface regions that the algorithm will detect and label as possible protein−protein binding sites.

A major goal of this paper is the implementation of the idea suggested in our earlier algorithm for predicting protein−protein binding sites,[17] that multiple structurally related proteins can be aligned to a given protein and the three-dimensional structural conservation in all these proteins determines protein−protein binding sites. We apply this modified algorithm to search for protein−protein binding sites on a set of sixteen polypeptide chains involved in transient protein complexes and obtained from the Protein Data Bank.[19] Transient complexes which have been shown to be difficult to predict in other studies[20] are optimized for weak binding and the detection of protein−protein binding sites in transient protein complexes thus presents a rigorous test of the algorithm.

We also tested this modified algorithm on the set of protein structures used in our previous report[17] for comparison.

A secondary goal is to compare the results with those from an algorithm that uses a support vector machine (SVM) and

* Author to whom correspondence should be addressed. E-mail: dusa@cmm.ki.si.

10.1021/ci8000315 CCC: $40.75
Published on Web 05/14/2008

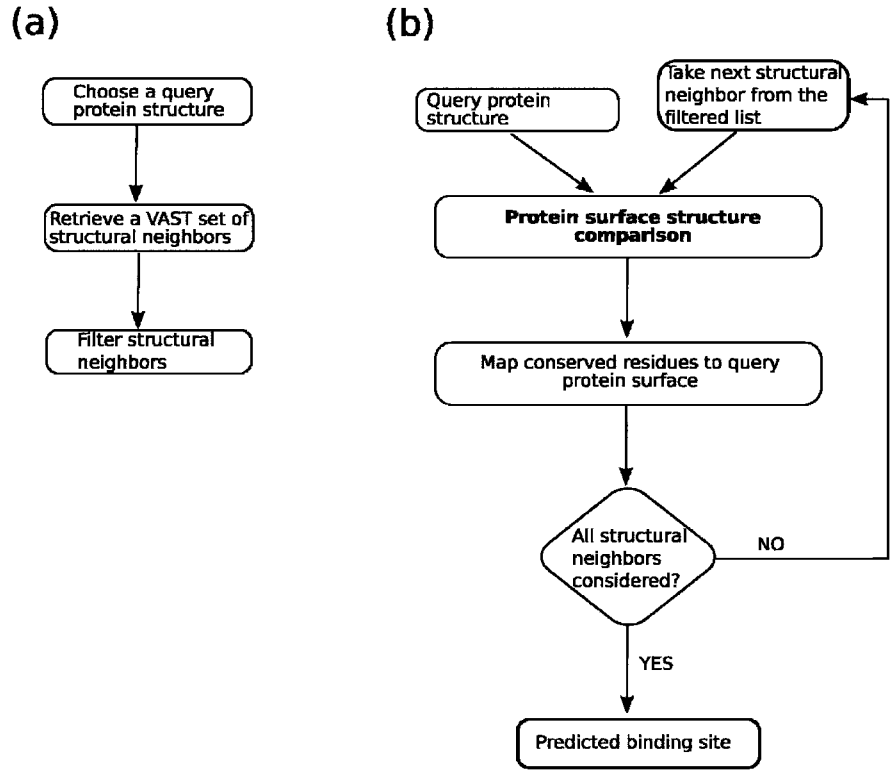

Figure 1. Algorithm for prediction of protein−protein binding sites. (a) Structural neighbors for the query protein are retrieved from the VAST server and filtered, so that only those with 15−60% sequence identity to the query protein are considered. (b) A protein is compared to its structural neighbors, and the conserved residues found in each of these neighbors are mapped to the surface of the query protein.

to provide a detailed analysis of protein surfaces to predict protein−protein binding sites on the same set of proteins.[20]

Our algorithm was found to detect both protein−protein interfaces and alternative conserved sites on protein surfaces. By comparing our algorithm with the SVM algorithm, it was shown that a significant proportion of protein−protein binding sites can be detected by structural conservation, which appears to be an important parameter separating transient interfaces from the remainder of the protein surfaces.

Table 1. Average Prediction Rate As a Function of the Distance Criteria

	distances (Å)					
averages	0.5	1	2	3	4	5
specificity (%)	13.0	22.0	25.7	31.0	37.3	43.1
sensitivity (%)	30.1	29.1	27.4	27.6	26.7	25.4
probability of random success	0.382	0.309	0.344	0.308	0.269	0.282

2. METHODS

Algorithm. In this section, we present an outline of our algorithm for the prediction of protein−protein binding sites. A detailed description of this algorithm has been published elsewhere.[17] To predict protein−protein binding sites, this earlier algorithm requires that the structures of the query protein and at least one structurally similar protein, i.e., a structural neighbor, are available. In our newer algorithm, a query protein is compared to multiple structural neighbors, selected systematically. We achieve this by filtering structural neighbors from the vector alignment search tool (VAST) structural neighbors list, which results in a more objective prediction of protein binding sites. The speed of each individual structural comparison is unchanged, but the overall computation time depends on the number of structural neighbors that are used. When comparing a protein with a structural neighbor, the algorithm first extracts the solvent accessible surface atoms from the two protein structures and replaces the functional groups of the surface residues with labeled vertices. These vertices are then connected by edges based on the distance between them to form a protein graph,[21] which is a representation of potential interactions, (aliphatic, aromatic, hydrogen bond donor, hydrogen bond acceptor, mixed acceptor/donor), that a protein can form with other molecules.[13] The algorithm then compares the query protein with each of the available structural neighbors, so that each of the labeled vertices from the investigated protein is compared with the corresponding labeled vertices from the structurally similar proteins. These comparisons are accomplished with an algorithm designed to find a maximum clique, which corresponds to the maximum similarity found between two protein surfaces.[15,16] Finally, the algorithm maps all the conserved patches that are found on the surface of the query protein. A schematic representation of the algorithm is shown in Figure 1.

Test Set. Transient protein complexes tend to be smaller than obligate complexes and have a preference for polar residues. Because proteins involved in transient interactions also exist independently, a large exposed hydrophobic patch on a protein is likely to be energetically unfavorable. Weaker interactions in these proteins allow multiply bound poses of the involved proteins and are made or broken according to the environment or external factors.[18] Since they are capable of independent existence, proteins involved in transient

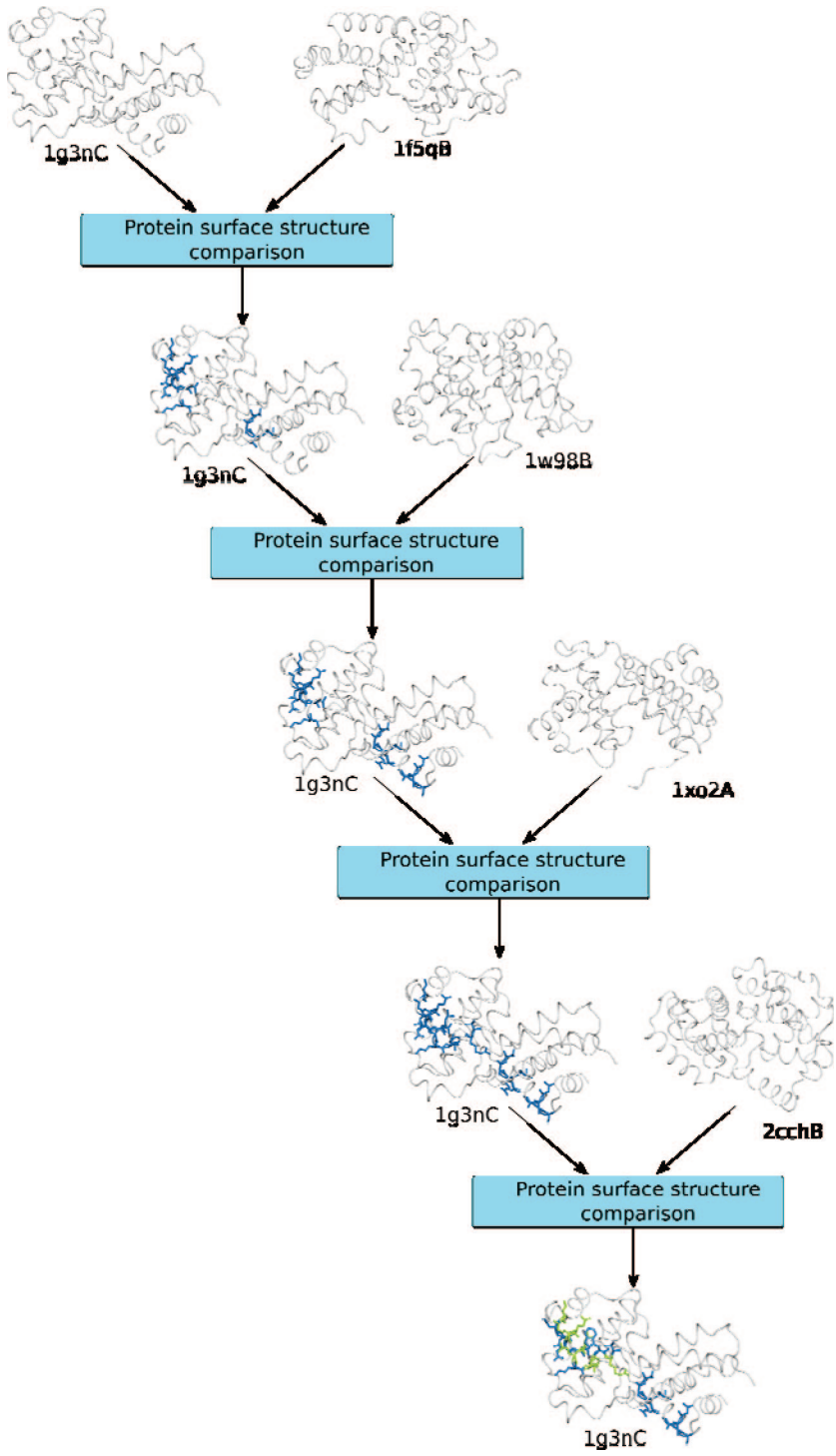

Figure 2. Protein−protein binding sites sequential prediction procedure: K-cyclin is the query protein (PDB code 1g3nC) and the structural neighbors are 1f5qB, 1w98B, 1xo2A, and 2cchB. Residues found to be conserved once are shown in blue, and those conserved twice are shown in yellow.

interactions are a compelling class of proteins with which to predict protein−protein binding sites.

A set of transient protein complexes was chosen from the Protein Data Bank,[19] and each protein from this set was split into constituent polypeptide chains. A total of 16 polypeptide chains involved in transient interactions with other chains were retained. We ensured that the interactions between the proteins in question are real and not a consequence of crystal packing, and that all of the interfaces on a polypeptide chain were considered. The unrestricted mobility of small chains makes their structural comparison with our algorithm difficult, and accordingly, we discarded the chain I of the protein with PDB code 1apm, since it is composed of only 20 residues. For the protein 1g3n, we considered only chains A, B, and C, since only these form any biologically relevant complex. The same applies for the protein 1rrp, in which we considered only chains A and B.

We defined two residues as being a part of a protein−protein interface if the distance between any two atoms from different chains was less than the sum of their van der Waals radii plus 3.0 Å. A value of 0.5 Å has been used by other authors,[22] but we found this to be too restrictive. Conserva-

Table 2. Results from Predictions for 16 Proteins Involved in Transient Interactions and 8 Proteins Used in ref 17

PDB code and chain ID	surface size (residues)	interface size (residues)	probability of random success	details of predicted patch		
				size (residues)	specificity (%)	sensitivity (%)
transient						
1apmE	290	41	0.529	76	14.5	26.8
1efuA	339	83	0.211	77	28.6	26.5
1efuB	266	89	<0.001	40	75.0	33.7
1g3nA	275	65	0.389	54	25.9	21.5
1g3nB	141	28	0.194	53	24.5	46.4
1g3nC	214	38	0.001	24	45.8	28.9
1gotA	322	36	0.012	57	21.1	33.3
1gotB	309	123	0.005	38	60.5	18.7
1k9oE	198	38	0.002	87	28.7	65.8
1k9oI	354	21	0.362	36	8.3	14.3
1rrpA	184	72	0.179	36	47.2	23.6
1rrpB	129	68	1.000	13	7.7	1.5
1ughE	211	35	1.000	6	0.0	0.0
1ughI	82	34	1.000	6	0.0	0.0
1ytfA	169	18	0.010	33	24.2	44.4
1ytfD	100	74	0.026	49	83.7	55.4
mixed types[a]						
1allA	153	43	0.005	80	37.5	69.8
1hcgA	216	33	0.824	111	13.5	45.4
1lucA	309	65	<0.001	73	52.1	58.5
1tcoB	169	66	0.350	75	41.3	47.0
1bncA	387	42	0.014	113	16.8	45.2
1daaA	260	63	0.085	75	30.7	36.5
1azeA	54	27	0.986	27	37.0	37.0
1lw6I	62	18	0.336	37	32.4	66.7

[a] Protein structures from ref 17.

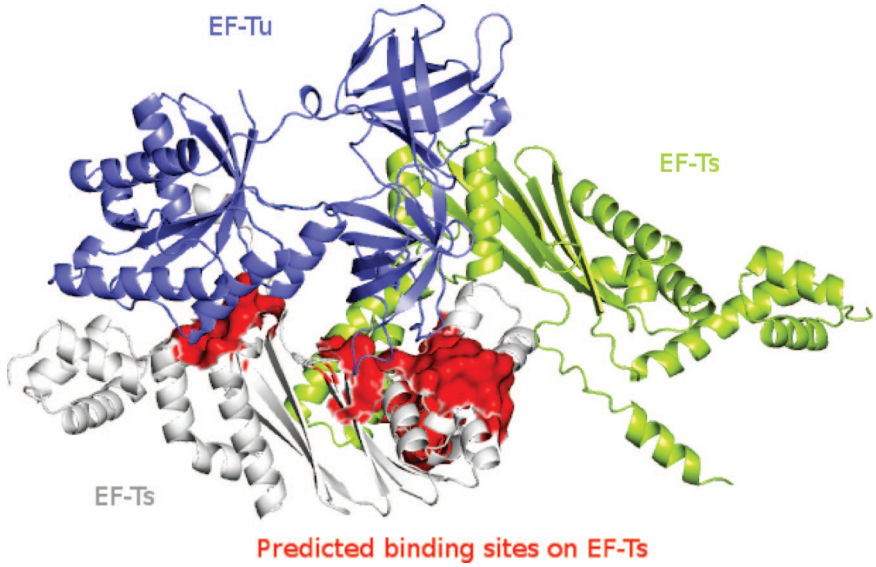

Figure 3. Predicted binding site (red) between elongation factor EF-Ts (white) and elongation factor EF-Tu (blue). The homodimeric complex between the two EF-Ts (white and green) is also shown.

tion is not limited to only those residues in direct contact with one another, and we also wanted to sample nearby residues, which provide a structural scaffold to the interface and which may also be considered by the algorithm. We tested this by calculating the parameters that describe predicted binding sites, e.g., specificity, sensititvity, and probability of random success (see below) for different distance parameters: 0.5, 1.0, 2.0, 3.0, 4.0, and 5.0 Å. The average values of these parameters calculated using the test set proteins are presented in Table 1. The prediction rate is significantly affected by varying the distance criteria: the specificity increases and the sensitivity and the probability of random success decreases with increasing distance. This indicates that at low values of the distance parameter (0.5 Å) the prediction rate decreases while at high values of this parameter (5.0 Å) it is generally increased, while the sensitivity is reduced. A distance of 3.0 Å accounts well for residues surrounding the binding site, which can be structurally conserved, while not compromising the sensitivity of prediction.

Prediction of Binding Sites. Each of the 16 polypeptide chains was compared with several structural neighbors, and the surface that was found to be conserved between the chain and its neighbor structures was predicted to be the binding

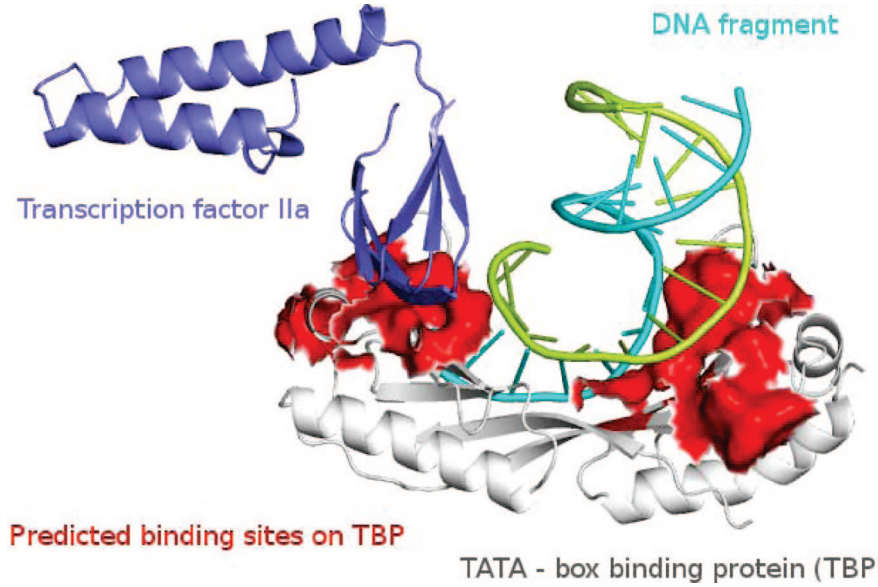

Figure 4. Predicted binding sites (red) on the TATA-box-binding protein (white) in complex with transcription factor IIA, chain D (blue), and a TATA-element DNA fragment (cyan and green).

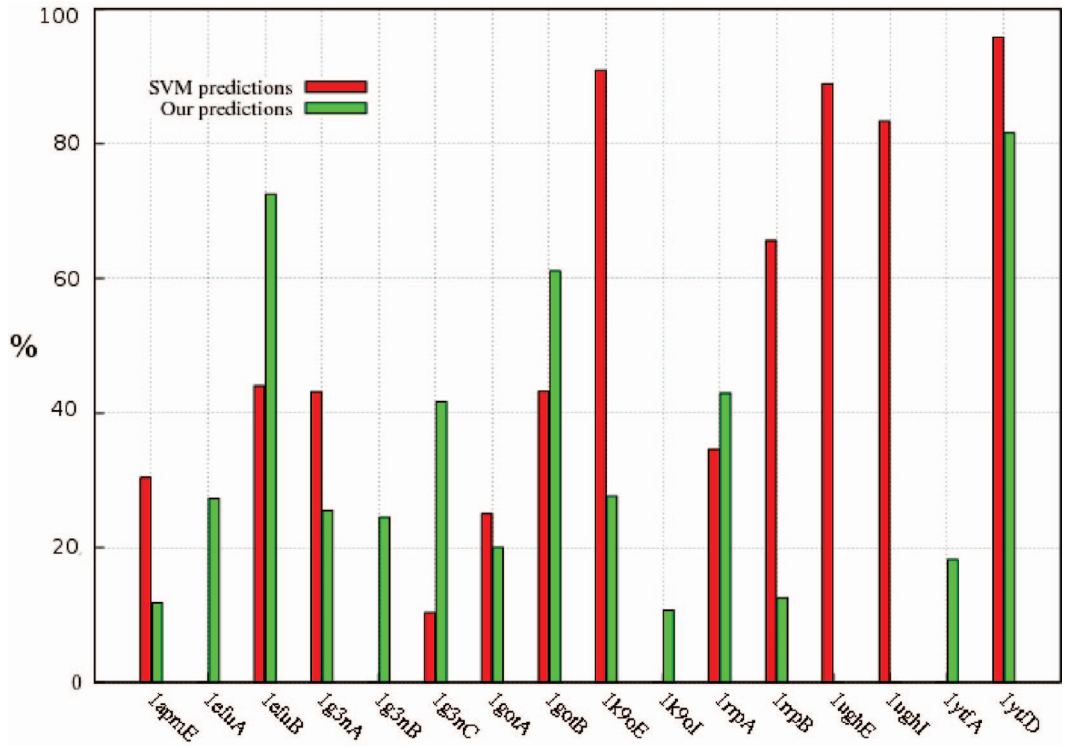

Figure 5. Specificities of predicted protein−protein binding sites on each of the polypeptide chains: our algorithm (green) and the SVM algorithm (red).

site for the partner chains. A list of structural neighbors was provided by the VAST (vector alignment search tool, http://www.ncbi.nlm.nih.gov/Structure/VAST),[23] which offers structure−structure alignments of publicly available protein structures. We used a medium redundancy list to avoid excessively similar structures, the use of which could lead to biased predictions. All of the protein structures in the list share some sequence identity with the polypeptide chain for which the prediction is performed. The difference from our earlier approach[17] is that we compared each polypeptide chain sequentially with multiple structural neighbors, while previously, we considered only one or a few structural neighbors, selected nonsystematically. Each of the automatically selected structural neighbors had a 15−60% identical sequence with the query polypeptide chain in the structurally aligned region. In each step of the algorithm, the detected conserved surface regions were mapped on those identified previously. Each residue on the surface of a query protein was assigned a conservation score, which counts the number of times this residue is conserved in selected structural neighbors. A surface residue is considered to be a part of a protein−protein binding site if it is conserved at least once.

The agreement of predicted binding sites with the actual binding sites was measured in terms of

1. Specificity. This indicates the proportion of the predicted residues that are also interface residues and is defined as

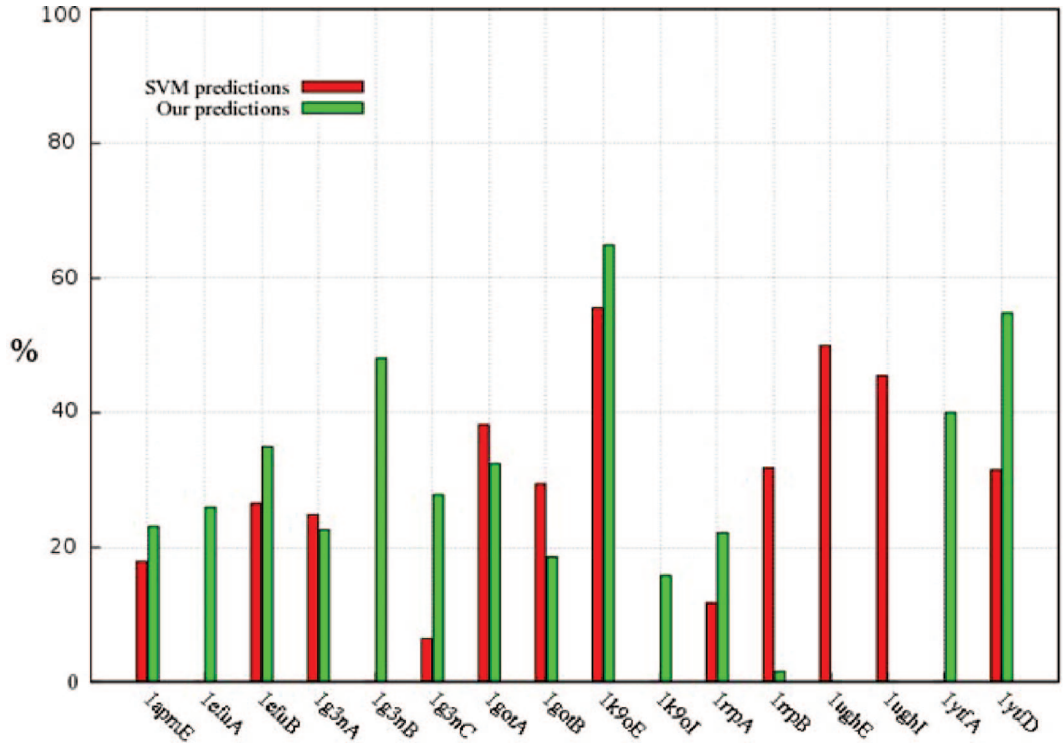

Figure 6. Sensitivities of predicted protein−protein binding sites on each of the polypeptide chains: our algorithm (green) and the SVM algorithm (red).

specificity = number of predicted residues in the actual interface/number of predicted residues.

2. Sensitivity. This is the proportion of the interface that was predicted and is defined as sensitivity = number of predicted residues in the actual interface/number of interface residues. The protein−protein binding sites prediction procedure is demonstrated with the K-cyclin (1g3nC) as the query protein and is shown in Figure 2. This protein is compared with the structural neighbors and the residues that are found conserved in each step are mapped to the surface of the query protein.

Significance of Predictions. To measure the significance of predictions, we introduce the probability P of randomly guessing a patch of surface residues of equal size as the predicted patch, but with equal or better overlap with the actual binding site than the predicted patch, which is described by eq 1,

$$P = 1/\binom{S_s}{P_s}\sum_{i=O_s}^{\min(P_s,I_s)}\binom{I_s}{i}\binom{S_s - I_s}{P_s - i} \quad (1)$$

where S_s (surface size) is the total number of surface residues, I_s (interface size) is the number of residues in the actual protein−protein binding site, P_s (predicted size) is the number of residues in the predicted protein−protein binding site, and O_s (overlap size) is the number of predicted residues that overlap with the actual protein−protein binding site. For example, the P-value of 0.5 for a patch of predicted residues indicates that in 50 out of 100 attempts a patch of randomly chosen residues with the same size will lead to a better prediction of the actual binding site.

Use of the Support Vector Machine Algorithm. We compared our algorithm for predicting protein−protein binding sites with an algorithm that uses a support vector machine (SVM) approach with surface patch analysis.[20] The two algorithms were tested on the same set of 16 polypeptide chains involved in transient complexes and the specificities and sensitivities of the predicted protein−protein binding sites were determined. In the SVM approach, several properties describing the surface of a protein, e.g., surface shape, sequence conservation, electrostatic potential, hydrophobicity, and residue interface propensity, are calculated and used to distinguish protein−protein binding sites from the rest of the protein surface. This approach gives three patches ranked by the probability of each being a part of a protein−protein binding site. Because our algorithm predicts only one patch, we discarded the two alternatives and considered only the most probable predicted patch.

3. RESULTS

Our algorithm finds conserved regions on protein surfaces and was tested for its ability to predict protein−protein binding sites on a set of transient protein complexes. We considered only polypeptide chains that naturally form a transient complex; polypeptide chains of less than 20 residues were not considered and interfaces present in the crystal structures only as a result of crystal packing were discarded. The 16 polypeptide chains that form transient interfaces and the associated predicted binding sites are presented in Table 2.

We considered a prediction to be significant when the P-value (the probability for random success) was < 0.2. On this basis, a significant prediction of a protein−protein binding site was achieved in 9 of the 16 cases. The best prediction, with a P-value < 0.001, is of the transient interface between polpeptide chains EF-Ts (1efuB) and EF-Tu. This interface is predicted with 75.0% specificity and 33.7% sensitivity and is shown in Figure 3. The tetrameric complex contains two subunits of each of the elongation factors, i.e., two EF-Ts molecules and two EF-Tu molecules.

Besides the transient interface between EF-Ts and EF-Tu, the two EF-Ts molecules also form a tight dimer, but there is little contact between the two EF-Tu molecules.[24] The agreement between the predicted and real binding site at rigid parts of the EF-Ts protein is excellent, and the homodimeric interface between the two EF-Ts molecules is also predicted; but, the flexible C-terminal tail, which is also a part of this binding site, was not identified in the prediction.

In the case of TATA-box-binding protein (1ytfA) complexed with transcription factor IIA, the binding site is predicted with 24% specificity and 44% sensitivity.With such low specificity, this prediction would be classified as poor, if it were not for the low *P*-value of 0.010. Examination of the results revealed two conserved but disconnected patches, one of which overlapped with the protein−protein interface with 85% specificity and 46% sensitivity. Surprisingly, the larger of the two conserved patches almost completely covers the binding site for TATA-element DNA fragment.[25] Both predicted binding sites are shown in Figure 4.

We further examined uracil−DNA glycosylase (1ughE) and uracil−DNA glycosylase inhibitor (1ughI), two predictions which failed to reveal the interface seen in the crystal structure. For each of these two proteins, the algorithm found only one structural neighbor with low sequence identity with the considered proteins, insufficient for successful structural comparison. Very small predicted binding sites, e.g. six residues, for these two proteins could indicate that what was found is not a binding site but a randomly similar surface patch. For example, a structural similarity of residues belonging to the alpha helices can be expected, since these are common motifs in the secondary structures of proteins. This highlights the problem of choosing structural neighbors on the predictions of our method, which might be overcome by a more thorough inspection of the structural alignments obtained with our algorithm. A query protein structure could be compared to some unrelated proteins and the similarites found could be used to distinguish meaningful predicted binding sites from random similarities, when comparing the query protein to the structural neighbors.

Our present algorithm, which takes into account multiple structural neighbors, was also tested on the set of protein structures used in our previous report,[17] and the results of these predictions are presented in Table 2. The quality of predicted binding sites is comparable to that obtained in our previous work, but the results can not be easily compared, since previously we did not have a criterion with which to select structural neighbors and the predictions could vary greatly with the choice of structural neighbor. The main advantage of the present method is that the conserved surface is found by structural comparisons of several structural neighbors, and this increases the reliability of the predicted binding sites.

Comparison with Other Algorithms. We compared our algorithm with one that combines a support vector machine (SVM) approach with a surface patch analysis for the prediction of protein−protein binding sites. We compared our algorithm and the SVM algorithm for their ability to predict protein−protein binding sites with the same set of 16 polypeptide chains involved in transient interactions. Our algorithm's predicted binding sites achieved, on average, a specificity of 31% and a sensitivity of 28%. The first patch predicted by the SVM algorithm had, on average, a specificity of 41% and sensitivity of 23%.

In view of the relative simplicity of the idea underlying our method, which relies on structural conservation as the only parameter to predict protein−protein binding sites, this is an encouraging result which suggests that the conservation of protein structures is an important factor in the prediction of protein−protein binding sites. Furthermore, because our algorithm predicts alternative conserved sites, e.g., protein−DNA binding sites, in a nondiscriminatory fashion, it is at a disadvantage when compared with algorithms that predict only protein−protein binding sites. The results of the comparison between our method and the SVM-based procedure are presented in Figures 5 and 6.

4. CONCLUSIONS

An algorithm combining bioinformatics and graph theory to detect protein−protein binding sites using multiple structural alignments has been developed. The algorithm was tested on a set of transient protein complexes and compared to a standard algorithm for its ability to predict protein−protein binding sites. The new algorithm also detects other types of binding sites and may be used for purposes other than detecting protein−protein binding sites. The new algorithm represents a considerable advance with its recognition that structural conservation is an important parameter distinguishing a binding site from the rest of the protein surface and that it can be used to detect protein−protein binding sites in proteins involved in transient interactions.

The important advantage of the new algorithm is that the conserved parts of two proteins are found by many local alignments of relatively small parts of surface structures and not by only one global alignment of protein structures. This may reduce the problem of structural flexibility which presents a problem for current techniques that search for conservation by globals structural alignment.[26]

ACKNOWLEDGMENT

The financial support through grants P1-0002 of the Ministry of Higher Education, Science, and Technology of Slovenia and the Slovenian Research Agency is acknowledged.

REFERENCES AND NOTES

(1) Lichtarge, O.; Sowa, M. E. Evolutionary Predictions of Binding Surfaces and Interactions. *Curr. Opin. Struct. Biol.* **2002**, *12*, 21–27.
(2) Fahmy, A.; Wagner, G. TreeDock: A Tool for Protein Docking Based on Minimizing van der Waals Energies. *J. Am. Chem. Soc.* **2002**, *124*, 1241–1250.
(3) Mendez, R.; Leplae, R.; De Maria, L.; Wodak, S. J. Assessment of Blind Predictions of Protein-Protein Interactions: Current Status of Docking Methods. *Proteins* **2003**, *52*, 51–67.
(4) Smith, G. R.; Sternberg, M. J. Prediction of Protein-Protein Interactions by Docking Methods. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28–35.
(5) Zhong, S.; MacKerell, A. D. Binding Response: A Descriptor for Selecting Ligand Binding Site on Protein Surfaces. *J. Chem. Inf. Model.* **2007**, *47*, 2303–2315.
(6) Clackson, T.; Wells, J. A. A Hot Spot of Binding Energy in a Hormone Receptor Interface. *Science* **1995**, *267*, 383–386.
(7) Keskin, O.; Ma, B.; Nussinov, R. Hot Regions in Protein-Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* **2005**, *345*, 1281–1294.
(8) Li, X.; Keskin, O.; Ma, B.; Nussinov, R.; Liang, J. Protein-Protein Interactions: Hot Spots and Structurally Conserved Residues Often Locate in Complemented Pockets that Pre-organized in the Unbound States: Implications for Docking. *J. Mol. Biol.* **2004**, *344*, 781–795.

(9) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. Protein-Protein Interactions: Structurally Conserved Residues Distinguish Between Binding Sites and Exposed Protein Surfaces. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5772–5777.
(10) Aytuna, A. S.; Gursoy, A.; Keskin, O. Prediction of Protein-Protein Interactions by Combining Structure and Sequence Conservation in Protein Interfaces. *Bioinformatics* **2005**, *21*, 2850–2855.
(11) Caffrey, D. R.; Somaroo, S.; Hughes, J. D. Are Protein-Protein Interfaces More Conserved in Sequence than the Rest of the Protein Surface. *Protein Sci.* **2004**, *13*, 190–202.
(12) Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westhead, D. R. Ligand Binding: Functional Site Location, Similarity and Docking. *Curr. Opin. Struct. Biol.* **2003**, *13*, 389–395.
(13) Schmitt, S.; Kuhn, D.; Klebe, G. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *J. Mol. Biol.* **2002**, *323*, 387–406.
(14) Konc, J.; Hodošček, M.; Janežič, D. Molecular Surface Walk. *Croat. Chem. Acta* **2006**, *79*, 237–241.
(15) Konc, J.; Janežič, D. An Improved Branch and Bound Algorithm for the Maximum Clique Problem. *MATCH Commun. Math. Comput. Chem.* **2007**, *58*, 569–590.
(16) Konc, J.; Janežič, D. A Branch and Bound Algorithm for Matching Protein Structures. *Lect. Notes. Comp. Sci.* **2007**, *4432*, 399–406.
(17) Konc, J.; Janežič, D. Protein-Protein Binding-Sites Prediction by Protein Surface Structure Conservation. *J. Chem. Inf. Model.* **2007**, *47*, 940–944.
(18) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13–20.
(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(20) Bradford, J. R.; Westhead, D. R. Improved Prediction of Protein-Protein Binding Sites Using a Support Vector Machines Approach. *Bioinformatics* **2005**, *21*, 1487–1494.
(21) Janežič, D.; Miličević, A.; Nikolić, S.; Trinajstić, N. *Graph Theoretical Matrices in Chemistry*; Mathematical Chemistry Monographs; University of Kragujevac and Faculty of Science: Kragujevac, 2007; No. 3.
(22) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. A Dataset of Protein-Protein Interfaces Generated With a Sequence-Order-Independent Comparison Technique. *J. Mol. Biol.* **1996**, *260*, 604–620.
(23) Gibrat, J. F.; Madej, T.; Bryant, S. H. Surprising Similarities in Structure Comparison. *Curr. Opin. Struct. Biol.* **1996**, *6*, 377–385.
(24) Kawashima, T.; Berthet-Colominas, C.; Wulff, M.; Cusack, S.; Leberman, R. The Structure of the Escherichia Coli EF-Tu.EF-Ts Complex at 2.5 Å Resolution. *Nature* **1996**, *379*, 511–518.
(25) Tan, S.; Hunziker, Y.; Sargent, D. F.; Richmond, T. J. Crystal Structure of a Yeast TFIIA/TBP/DNA Complex. *Nature* **1996**, *381*, 127–134.
(26) Chung, J. L.; Wang, W.; Bourne, P. E. Exploiting Sequence and Structure Homologs to Identify Protein-Protein Binding Sites. *Proteins* **2006**, *62*, 630–640.

CI8000315