



Predicting protein interface residues using easily accessible on-line resources

Surabhi Maheshwari and Michal Brylinski

Corresponding author: Michal Brylinski, Department of Biological Sciences, Louisiana State University, 202 Life Sciences Bldg., Baton Rouge, LA 70803, USA. Tel.: +1 225 578 2791; E-mail: michal@brylinski.org

Abstract

It has been more than a decade since the completion of the Human Genome Project that provided us with a complete list of human proteins. The next obvious task is to figure out how various parts interact with each other. On that account, we review 10 methods for protein interface prediction, which are freely available as web servers. In addition, we comparatively evaluate their performance on a common data set comprising different quality target structures. We find that using experimental structures and high-quality homology models, structure-based methods outperform those using only protein sequences, with global template-based approaches providing the best performance. For moderate-quality models, sequence-based methods often perform better than those structure-based techniques that rely on fine atomic details. We note that post-processing protocols implemented in several methods quantitatively improve the results only for experimental structures, suggesting that these procedures should be tuned up for computer-generated models. Finally, we anticipate that advanced meta-prediction protocols are likely to enhance interface residue prediction. Notwithstanding further improvements, easily accessible web servers already provide the scientific community with convenient resources for the identification of protein–protein interaction sites.

Key words: protein–protein interactions; protein interface prediction; interfacial residues; protein–protein complexes; protein models; web servers

Introduction

Proteins do not operate in isolation, rather they interact with each other either directly or indirectly to carry out their functions [1]. In fact, protein–protein interactions (PPIs) play a pivotal role in cellular functions mediating virtually all biological processes. Therefore, significant efforts have been devoted to characterize and catalog PPIs to improve our understanding of molecular recognition and reveal the mechanisms by which proteins work. Mapping these interactions facilitates the modeling of the entire functional proteome and its constituent pathways. Moreover, linking PPIs to diseased states and other phenotypes helps develop drugs that directly target protein–protein interfaces [2, 3]. Computationally inferred information about interfacial residues also aids the design of mutants for

the experimental verification of interactions [4, 5] as well as it enhances the prediction of complex structures through homology modeling and protein docking [6–8].

Given that numerous biological applications require information about surface regions involved in PPIs, a wide range of experimental techniques have been designed to identify interfacial residues, with much efforts devoted to the development of high-throughput methods [9–11]. Nonetheless, these techniques are often tedious, labor intensive and are associated with high costs of experiments. In addition, many experimental techniques have been shown to suffer from high false-positive and false-negative rates, as well as inter-study discrepancies [12–14]. On the other hand, the ongoing proteomics and structural genomics research routinely generates massive amounts

Surabhi Maheshwari is a Graduate Assistant in the Department of Biological Sciences at Louisiana State University focusing on the large-scale modeling of protein–protein interactions for applications in network and systems biology.

Michal Brylinski is an Assistant Professor in the Department of Biological Sciences and the Center for Computation & Technology at Louisiana State University interested in the modeling of biological systems for biopharmaceutical discovery.

Submitted: 1 December 2014; Received (in revised form): 13 January 2015

© The Author 2015. Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

of data, which need to be interpreted at a fast pace. Hence, there is a dire need for computational methods to effectively identify PPIs, and to assess, validate and scrutinize experimentally collected data. One of the first attempts to predict residues located at the interface was made by Jones and Thornton [15]. Since then, a number of methods for predicting protein–protein interface residues have been reported. These approaches use diverse techniques for the identification of PPI sites and may vary in terms of the attributes used to distinguish interacting sites and the implemented learning/prediction algorithms [16–19]. In general, computational methods can be broadly divided into sequence- and structure-based approaches. Sequence-based methods often use sliding window frames to calculate the specific features associated with residues based on their neighbors [20–22]. These methods employ various residue-level properties, such as the degree of evolutionary conservation, physicochemical features and energetics, to construct scoring functions. Furthermore, the availability of protein tertiary structures allows for the integration of a variety of structural information, e.g. solvent accessibility, B-factors and secondary structure, to improve the prediction accuracy [23].

Many recently published reviews provide insights into the fundamentals of protein binding and docking and discuss the mechanics of PPI prediction. Zhou and Qin give a comprehensive overview of the underlying principles used by different methods and discuss the challenges faced by the community [24]. Vries and colleagues provide a critical assessment of the state-of-the-art in PPI prediction, compare different approaches and explain difficulties in assessing the absolute and relative performance of various predictors owing to differences in the choice of data and evaluation criteria [25]. A review by Ezkurdia *et al.* examines the weak points of current PPI prediction methods arising from the incomplete structural information on transient complexes, which remain largely under-represented in the Protein Data Bank (PDB) [26]. Finally, Wang and colleagues focus on machine learning-based techniques and outline the key components of an effective prediction pipeline to infer protein interaction sites [19]. Because the majority of research studies concentrate on the experimental structures of target proteins in their bound and unbound conformations, significantly fewer reviews touch on issues related to using protein models in the structure-based prediction of PPI sites. Certainly, the unavailability of structural data may impose constraints on research projects involving PPIs. Using protein models mitigates this issue, however, assuming that PPI prediction methods tolerate imperfections in the target structures. Therefore, in this communication, we describe 10 freely accessible web servers for PPI site prediction and comparatively evaluate their performance on a common data set assessing the effect of using computer-generated models on the prediction accuracy.

Types of protein complexes

Protein–protein complexes can be divided into obligatory and transient assemblies based on their overall interaction strength and stability. Obligatory complexes are functional only in their coupled state, and the monomers do not exist as stable structures *in vivo*. The interaction partners also have a high shape complementarity, and their interface residues resemble the hydrophobic core of globular proteins. In contrast, transient complexes are formed by proteins that may be functional even in their unbound monomeric state. The interface of such complexes is stabilized by weak interactions, the partners have a lower geometrical complementarity and the interface area

between them is relatively small compared with obligatory complexes. Also, the hydrophobicity of residues that make up the interface of transient associations is indistinguishable from the remaining protein surface. With respect to the sequence identity between monomers, protein assemblies can be divided into homo- and hetero-complexes. The former consist of two or more identical chains, while the latter are composed of protein chains with different amino acid sequences. Obligatory associations can be homo- and hetero-complexes, whereas the majority of transient assemblies are hetero-complexes that comprise different chains. In general, interfacial sites in obligatory complexes are easier to detect, as they are generally larger, flatter, more hydrophobic and more conserved than transient interfaces [27–29].

Interfacial regions of protein surface

Proteins interact with one another via interfacial sites predominantly composed of surface residues. Interface residues tend to be more conserved than other positions; however, this signal is weakened for residues below a certain solvent accessibility. Therefore, the definition of surface residues plays a pivotal role in the creation of databases for methods exploiting evolutionary conservation. The prediction accuracy also strongly depends on how surface residues are defined; as a common practice, residues are classified as surface residues if their relative solvent accessibility (RSA) is above some threshold. Different studies use different cutoffs, which typically range from 5 to 16% [26, 28], with higher thresholds leading to a lower number of solvent-exposed residues. Based on the three-dimensional structure of a protein complex, PPI sites are identified from the subset of surface residues either using interatomic distances between non-hydrogen atoms in different protein chains or by calculating the change in the solvent accessible area upon complex formation. In both cases, empirically optimized thresholds are often used; for instance, distance-based methods typically use cutoff values of 4 Å, 4.5 Å or 5 Å [30–32], whereas surface-based approaches define interfacial residues as those whose accessible surface area changes by $>20\text{Å}^2$ [28].

Characteristic features of interface residues

Comparison of interfacial and non-interfacial regions on protein surfaces reveals a number of intrinsic characteristics of residues involved in the formation of quaternary structures. These features are commonly used by PPI prediction algorithms, and can be broadly classified into the following three categories:

- **Sequence-based features** are derived from the amino acid sequence alone and use various physicochemical properties of residues to identify the interface regions. Examples of these features are interface propensity [33, 34], hydrophobicity and electrostatic desolvation [35], as well as structural attributes predicted from sequence, such as secondary structure and solvent accessibility [23, 36].
- **Structure-based features** are derived from the tertiary structures of target proteins. These attributes include, but are not limited to, solvent accessible surface area [37, 38], secondary structure [39], crystallographic B-factors [40], local geometries [41], as well as the spatial distribution of hydrophobic and polar surface patches [42].
- **Evolutionary features** are calculated by comparing the sequence of a query protein to the sequences of its homologs. Interface residues tend to be highly conserved, in contrast to non-

interfacial surface residues that are subjected to a notably lower selection pressure [43, 44]. Thus, the sequence conservation reflects the evolutionary selection at interfacial sites to maintain protein function. These attributes have a high discriminatory power toward interfacial residues; for example, protein sequence entropy is a conservation score that estimates sequence variability, thus it is often used in PPI site prediction [45].

Feature integration and the prediction of PPI sites

While a number of discriminatory features have been explored, individual attributes provide only a weak signal, thus no single feature can be used to unambiguously identify the interaction regions in proteins [24]. Because these attributes may provide a complementary discriminatory power with respect to each other, many PPI residue predictors combine different features to more effectively identify interfacial regions. Individual features are often integrated using scoring functions and machine learning techniques. The optimization of a relatively small number of attributes can be done by constructing a discriminant function that either linearly or nonlinearly combines individual features [15, 28, 38, 46]. More recently, machine learning strategies have become popular, especially for the optimal combination of a large number of attributes. The most commonly used machine learning algorithms include Neural Networks (NNs) [17, 20, 32, 47], Support Vector Machines (SVMs) [30, 31, 48], Random Forests (RFs) [22] and Naïve Bayesian Classifiers (NBCs) [39].

Most PPI site predictors fall into two major categories, residue- and patch-based methods. Residue-based techniques assign each residue in the target protein with a score corresponding to the probability to be a part of the interface [39, 31, 49, 50]. These residues need not necessarily be adjacent on the protein surface; however, clustering algorithms are sometimes used to impose a spatial proximity. The output from such methods often contains raw interface/non-interface scores calculated for all residues in the target protein as well as a separate list of predicted interface residues that have their score above some predefined threshold. Methods that use machine learning usually adopt the residue-based approach, as the input data can be conveniently mapped to the feature space. On the other hand, patch-based methods partition a target protein surface into a set of discrete patches/clusters [15]. These surface patches are then analyzed and ranked based on a combined score calculated using individual features with the top-ranked group taken as the predicted interface. In addition to interface/non-interface scores assigned to individual residues, the output from patch-based methods often contains a confidence score derived for the entire cluster of residues. A weak point of many patch-based strategies is that the predicted patches are generally circular, whereas biological interfaces tend to be rather irregular in shape. Furthermore, these methods also require estimating the size of a putative interfacial site; nevertheless, this information can be reliably obtained from a correlation between the number of interfacial residues and the target protein length [15, 51].

Intrinsic disorder in protein interactions

While the main focus of this review is on the structure-based prediction of interface residues, other methods for the identification of PPIs involving intrinsically disordered proteins attract significant attention owing to the fact that the flexibility and disorder play an important role in molecular recognition.

Briefly, the term ‘intrinsic disorder’ refers to those proteins and protein segments that fail to self-fold into fixed tertiary structures [52]. Attributed to unique characteristics of interactions mediated by intrinsically disordered proteins, the involvement of disordered regions in complex PPI networks has become increasingly apparent in recent years. For instance, these molecules can recognize multiple partners on the adoption of different conformations contributing to binding diversity [53]. Moreover, owing to a relatively lower binding affinity compared with classical binding, interactions involving intrinsically disordered segments are fully reversible while maintaining the high specificity [54]. Interestingly, binding motifs located in longer intrinsically disordered protein regions, called Molecular Recognition Features (MoRFs), undergo disorder-to-order transitions on binding [55]. Several prediction methods have been developed to identify MoRFs from protein primary sequence, e.g. SLiMPred [56], MoRFPred [57] and ANCHOR [58]. The implications of the protein intrinsic disorder in molecular recognition and binding functions are comprehensively discussed in a recent review [59].

Web servers for PPI site prediction

A number of algorithms for PPI site prediction are freely available to the scientific community as user-friendly web servers. Here, we selected 10 resources (listed in Table 1) that represent a variety of methods and were up and running at the time of this study. Moreover, these web servers offer a possibility to process data sets of moderate sizes in the order of a couple of hundreds of proteins using either web-based interfaces or command line tools that can query remote services. The selected web servers are arranged in four groups: (1) primarily sequence profile-based techniques that additionally use the accessible solvent area (ASA), (2) those approaches using residue-level characteristics, (3) algorithms using sub-residue physicochemical and structural features and (4) template-based methods that incorporate global structure alignments. Below, we review the design of individual web servers according to this classification.

Group I

We assigned two algorithms to this group, cons-PPISP and PSIVER. The original PPISP (Protein-Protein Interaction Site Predictor) algorithm [47] was developed to effectively exploit evolutionary information from sequence profiles constructed by PSI-BLAST [61] and the residue solvent exposure calculated by DSSP [62]. It uses an NN classifier, in which the nodes are fed with a series of scores including those calculated for spatial neighbors on the protein surface. It is noteworthy that PPISP was demonstrated to maintain its accuracy when unbound structures are used as the targets for interfacial residue prediction. The problem of over- and underpredictions was subsequently addressed by using a consensus classification by multiple NN models. This improved method, called cons-PPISP, uses a series of models ranging from a high accuracy with low coverage to a low accuracy with high coverage, and a new procedure for the spatial clustering of predicted interface residues [32]. Cons-PPISP not only offers a higher accuracy at an increased coverage compared with PPISP, but also shows a good agreement with experimental data as demonstrated for several proteins whose protein-protein complexes were characterized by NMR chemical shift perturbation.

The second method in this group is a sequence-based approach, PSIVER (Protein-protein interaction Sites prediction

Table 1. Summary of the design and implementation of 10 web servers for the prediction of protein interface residues

Group	Web server	Local features ^a		Global features ^b		Classifier	Clustering	Reference
		Propensity level	ASA ^c	Sequence profiles	Structure alignments			
I	Cons-PPISP	–	DSSP	PSI-BLAST	–	NN	+	[32]
	PSIVER	–	SABLE ^d	PSI-BLAST	–	NBC	–	[36]
II	InterProSurf	Residue	GetArea	–	–	Product	+	[37]
	SPPIDER	Residue	DSSP	PSI-BLAST	–	NN	–	[17]
	VORFFIP	Residue	DSSP	AL2CO	–	RF	–	[40]
	WHISCY	Residue	NACCESS	HSSP	–	LR	–	[46]
III	PIER	Sub-residue	ICM	BLAST, ZEGA	–	PLS-R	+	[28]
	ProMate	Atom	Connolly's MS	PSI-BLAST	–	NBC	+	[39]
IV	eFindSite ^{PPI}	Residue	NACCESS	PSI-BLAST	Fr-TM-align	SVM, NBC	–	[31]
	PredUs	–	SURFace	–	Ska	SVM	–	[60]

^aDerived for amino acids, groups of atoms or individual atoms. ^bDerived from sequence or structure alignments of the target protein and its homologs or structural neighbors. ^cAccessible Solvent Area. ^dPredicted from sequence.

NN = neural network; NBC = Naive Bayesian classifier; Product = average interface propensity weighted by ASA; RF = Random Forest; LR = Linear Regression; PLS-R = Partial Least Square Regression; SVM = support vector machines.

seVER) [36]. It uses an NBC and a set of sequence features to predict protein interaction sites, focusing on transient and heterodimer complexes. Two separate classification models are implemented in PSIVER for sequence profiles obtained from PSI-BLAST [61] and ASA. Because PSIVER is a sequence-based method, rather than calculating ASA directly from structure, these values are predicted for target sequences using SABLE [63]. Both NBCs calculate conditional probabilities using the kernel density estimation method. Leave-one-out cross-validation demonstrated that combining individual sequence profile- and ASA-based classifiers significantly improves the overall performance of PSIVER. Evaluated on an independent data set of proteins selected from the Protein Docking Benchmark Set 3.0 [64], PSIVER outperformed the ISIS server [29] and the sequence-based version of SPPIDER [17].

Group II

Among many residue-level attributes, interface propensities derived for individual amino acids are frequently used in interfacial residue prediction, as exemplified by several methods in this group. For instance, InterProSurf [37] uses interfacial propensities for amino acids calculated from a data set of 72 dimer structures [65]. Different from other approaches, InterProSurf first partitions the target protein surface defined by the GetArea program [66] using either a cluster or a patch analysis, and then applies a scoring function to find surface regions with high interface propensities. The number of high-ranking clusters in the clustering method and a radius in the patch analysis were optimized empirically to balance the sensitivity and precision of interface residue prediction. In addition to benchmarking simulations, InterProSurf successfully predicted interaction sites for the Anthrax toxin and measles virus hemagglutinin protein as validated by sequence analysis and mutagenesis experiments [37].

SPPIDER (Solvent accessibility based Protein-Protein Interface iDentification and Recognition) [17] is an NN method that uses a set of 19 attributes derived from the sequence and structure of a query protein, and its evolutionary profiles. Predicted solvent accessibility fingerprints are a novel feature implemented in SPPIDER. Interestingly, the difference between the observed and predicted ASA is highly informative and can be used to increase the predictive power of solvent accessibility-based features. The integration of the enhanced RSA

predictions by SABLE [63] with high-resolution structural data led to the development of RSA-based fingerprints of protein interactions, which were found to significantly improve the discrimination between interacting and noninteracting sites. Similar to cons-PPISP, SPPIDER is a consensus-based classifier that combines 10 cross-validated NN models with a *k*-nearest neighbor selection procedure to filter out misclassified residues.

A recent study indicated that Voronoi diagrams provide more accurate descriptions of the exposed residue environment than techniques based on Euclidian distances and sequence sliding windows [40]. This observation led to the development of VORFFIP (Voronoi Random Forest Feedback Interface Predictor), a novel method for protein binding site prediction. It integrates heterogeneous data including various residue-level structural and energetic characteristics, the evolutionary sequence conservation calculated by AL2CO [67] and crystallographic B-factors. VORFFIP uses a two-step RF classifier and a set of residue- and environment-based features to assign surface residues with interfacial scores. Cross-validation benchmarks performed on a data set derived from the Protein Docking Benchmark Set 3.0 [64] demonstrated that combining different features with Voronoi diagrams used as the environment descriptor yields the best performance. VORFFIP was also found to outperform other methods for binding interface prediction, SPPIDER [17] and WHISCY [46].

The last method in this group, WHISCY (What Information does Sequence Conservation Yield?) [46], uses a linear regression (LR) method to combine residue conservation and structural information to effectively discriminate between interfacial and non-interfacial residues. The conservation is computed from multiple sequence alignments obtained from the HSSP database [68]. WHISCY takes into account structural information such as interface propensities and considers the properties of surface neighbors to remove isolated high-scored residues. The implemented simple LR model offers a high flexibility by allowing users to choose which characteristic should be included in the prediction procedure. In a validation study, WHISCY and ProMate [39] were used to generate input for a data-driven protein docking program, HADDOCK [69]. Near-native structures constructed by docking simulations using unbound receptor conformations from the Protein Docking Benchmark Sets 1.0 and 2.0 [64, 70] demonstrate that incorporating the predicted PPI sites in data-driven docking yields an improved accuracy of the protein quaternary structure modeling.

Group III

Statistical properties are usually derived for individual amino acids; however, these can be also calculated at the sub-residue level of atomic groups. For example, PIER (Protein IntERface Recognition) [28] applies a partial least square regression (PLS-R) algorithm to optimize desolvation parameters [71] for 12 significant atomic groups whose ASA is calculated by ICM [72]. PIER initially divides the surface of a target protein into a set of individual patches. In the alignment-independent mode, a decision score indicating the likelihood of being at the protein interaction site is computed as a linear combination of the physical descriptors. Furthermore, sequence alignment information was incorporated to evaluate the strength of evolutionary signal. Specifically, in the alignment-dependent mode, surface patches are additionally assigned several features calculated from sequence alignments constructed by the Zero End-gap Global Alignment (ZEGA) method [73]. Interestingly, adding evolutionary information only marginally influenced the prediction performance of PIER and for certain classes of proteins, the evolutionary signal even deteriorated the prediction accuracy [28].

Atomic level descriptors are implemented in ProMate [39], an NBC method that identifies interface regions using composite probabilities derived from protein sequences and structures. ProMate uses Connolly's MS program [74] to identify surface atoms, which are subsequently extended to so-called circles. To classify these regions as interfacial, non-interfacial or boundary, an optimal combination of scoring terms was identified from a set 13 different properties comprising the chemical composition of binding interfaces, geometric properties, and specific information obtained from crystallographic data. Based on this classification, the neighboring circles are merged and clustered to predict interface patches. The algorithm was demonstrated to successfully predict the interface location for the majority of benchmarking transient hetero-complexes. Importantly, the identified biophysical properties were found to be largely independent of a particular receptor conformation; therefore, the success rate of ProMate was almost equal for target proteins experimentally solved in their bound and unbound states.

Group IV

The last group of methods for protein interface residue prediction comprises template-based predictors, eFindSite^{PPI} and PredUs. eFindSite^{PPI} capitalizes on the tendency of the location of binding sites to be highly conserved across evolutionarily related protein dimers [31]. It uses a collection of effective algorithms, including meta-threading by eThread [75], structural alignments by Fr-TM-align [76] and machine learning using SVMs and NBCs [77]. Each residue in a query protein is assigned a probability to be at the interface using residue-level attributes as well as structure and sequence conservation scores derived from evolutionarily related complexes. In addition, eFindSite^{PPI} effectively detects specific molecular interactions at the interface, such as hydrogen bonds, aromatic interactions, salt bridges and hydrophobic contacts. Previous comparative benchmarks demonstrated that it outperforms Protein IntERface residUe Prediction (PINUP) [38] using experimental protein structures as well as computer-generated models. The performance of eFindSite^{PPI} was also better than several other PPI site prediction programs, including PrISE (Prediction of protein-protein Interface residues using Structural Elements) [41], ET (Evolutionary Trace) [21] and JET (Joint Evolutionary Trees) [78].

Interface conservation is most significant among proteins that have a clear evolutionary relationship; however, it has been shown that a notable level of conservation exists among remote structural neighbors as well [49]. These structural insights are exploited by PredUs, a structure-based method that predicts surface residues likely to participate in the binding of other proteins [60]. For a given protein of interest, PredUs uses a structure alignment program Ska [79] to identify those structural neighbors forming complexes according to the Protein Quaternary Structures database [80] and the PDB [81]. Interfaces from neighbors are used to calculate contact frequencies, which along with ASAs computed by SURface [82] make a feature vector for SVMs [83]. PredUs offers several unique interactive features so that a prediction can be tailored to a particular hypothesis. For example, users can upload the structure of a binding partner to include structural neighbors of the partner in PPI residue prediction. Moreover, as proteins may interact with different partners at distinct regions to perform various molecular functions [84], the list of structural neighbors can be filtered based on functional information according to Gene Ontology [85], Structural Classification of Proteins [86], Pfam [87] and InterPro [88]. Comparative benchmarks demonstrated that PredUs outperforms several other algorithms, including PINUP [38], cons-PPIISP [32] and ProMate [39].

Head-to-head comparison of web servers

Comparing the performance of various algorithms for PPI site prediction reported in literature may not be straightforward as their accuracy was often assessed using different data sets and evaluation metrics. Moreover, most benchmarking studies focus on experimental structures in their bound and/or unbound conformations with significantly fewer assessments carried out for close and remote homology models. Yet, using protein models as the targets in PPI interface prediction is particularly relevant for across-proteome studies, where only sequence information is available for the vast majority of proteins. Therefore, in this review, we include a direct comparison of 10 web servers using a common testing data set composed of experimental and computer-generated structures.

Target proteins were selected from the Protein Docking Benchmark Set 4.0 [89]. We followed similar criteria to those used in our previous study [31], i.e. we excluded multimeric complexes, in which the receptor is either <50 or >600 amino acids, the interface is made up of <20 residues or multiple interfaces are present. This procedure resulted in a set of 90 target proteins forming heterodimers (42 enzyme/inhibitor or enzyme/substrate, one antibody/antigen and 47 other complexes). In addition to the experimental structures, we constructed high- and moderate-quality protein models for each target. Specifically, weakly homologous models were generated by eThread [75, 90] excluding closely related templates whose sequence similarity to the target is >40%. High-quality models have a TM-score [91] to native of >0.7, whereas the TM-score of moderate-quality models is within a range of 0.4–0.7. These sets of crystal structures and high- and moderate-quality models are referred to as BM90C, BM90H and BM90M, respectively. We queried the web servers with all BM90 structures using either web interfaces that allow for multiple target submissions or command-line tools and scripts. Because PSIVER is a sequence-based method, we queried it using BM90 sequences. The predictions were collected and assessed using several commonly accepted evaluation metrics that are derived from a confusion matrix as described below.

Accuracy measures for PPI residue prediction

Predicting interfacial residues can be formulated as a binary classification problem, where each protein residue can be either interfacial (positive, P) or non-interfacial (negative, N). Evaluation of the classification performance generally considers those cases that are correctly and incorrectly predicted for each class, which is quantified by the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Several metrics are commonly used to represent these four figures as a single measure of the binary classification performance:

- **Accuracy (ACC)** evaluates the effectiveness of a predictor by the fraction of correct predictions:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

- **Precision** (also **Positive Predictive Value, PPV**) evaluates the fraction of predicted interface residues forming an interface in the experimental complex structure:

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

- **Sensitivity** (also **True Positive Rate, TPR**) and **Specificity (SPC)** evaluate the effectiveness of the predictor for each class. TPR measures the fraction of correctly predicted interface residues, while SPC evaluates the fraction of correctly predicted non-interface residues:

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

$$SPC = \frac{TN}{FP + TN} \quad (4)$$

- **Fall-out** (also **False Positive Rate, FPR**) evaluates the fraction of predicted interface residues, which are not at the interface:

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

- **Matthew's Correlation Coefficient (MCC)** is a measure that balances the sensitivity and specificity, evaluating the strength of the correlation between predicted and the actual classes. Its values range from -1 to 1 , where 1 corresponds to a perfect prediction, 0 to a random prediction and -1 to a perfectly inverse prediction:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + TN)(FP + FN)(TN + FN)}} \quad (6)$$

- **Receiver Operating Characteristic (ROC)** plot, representing the relation between FPR and TPR on a single graph, is another widely used performance assessment method for binary classification problems.

Performance of web servers using experimental structures

We carried out a comparative assessment of the performance of 10 freely available PPI prediction servers using experimental target structures (BM90C) as well as their high- and moderate-quality models (BM90H and BM90M, respectively). Full ROC plots were constructed for those servers that provide continuous

residue scores; here, we also found the optimal threshold values that maximize MCC. Additionally, some servers use post-processing procedures, e.g. clustering and re-ranking, to compile a list of predicted residues; therefore, the performance was also assessed using the default list of predicted interfacial residues when this information was available. For these servers, the better performance (either optimized or default) was used in the comparative analysis.

Table 2 shows that using BM90C, the ranking of web servers based on MCC is PredUs, eFindSite^{PPI}, cons-PPISP, SPPIDER, ProMate, WHISCY, PIER, VORFFIP, PSIVER and InterProSurf. PredUs with MCC of 0.384 is the best performing server on this data set, eFindSite^{PPI} is second with MCC of 0.376 and cons-PPISP is third with MCC of 0.247. While MCC for PredUs is slightly better than that for eFindSite^{PPI}, SPC, PPV and ACC for eFindSite^{PPI} are higher than those for PredUs by 0.111, 0.156 and 0.075, respectively. Moreover, we point out that post-processing procedures implemented in several web servers often considerably improve their performance for crystal structures; note that diamonds representing the default predictions in Figure 1A are

Table 2. Comparison of the performance of 10 web servers for the prediction of protein interface residues using different quality target structures

Data set	Web server	MCC	TPR	FPR	SPC	PPV	ACC
BM90C	Pseudo-meta	0.481	0.692	0.094	0.905	0.417	0.887
	PredUs	0.383	0.701	0.156	0.843	0.302	0.831
	eFindSite ^{PPI}	0.375	0.396	0.045	0.954	0.459	0.905
	cons-PPISP	0.247	0.279	0.052	0.947	0.338	0.888
	SPPIDER	0.173	0.340	0.125	0.875	0.208	0.827
	ProMate	0.165	0.526	0.295	0.704	0.210	0.684
	WHISCY	0.164	0.130	0.025	0.975	0.334	0.900
	PIER	0.118	0.066	0.012	0.987	0.342	0.906
	VORFFIP	0.117	0.531	0.401	0.598	0.337	0.579
	PSIVER	0.103	0.645	0.463	0.536	0.118	0.546
BM90H	InterProSurf	0.100	0.435	0.291	0.709	0.163	0.677
	Pseudo-meta	0.443	0.680	0.108	0.891	0.380	0.872
	eFindSite ^{PPI}	0.340	0.377	0.051	0.948	0.414	0.898
	PredUs	0.309	0.571	0.147	0.852	0.272	0.827
	cons-PPISP	0.207	0.251	0.058	0.941	0.294	0.881
	SPPIDER	0.164	0.464	0.216	0.783	0.171	0.755
	PIER	0.137	0.234	0.088	0.911	0.204	0.852
	ProMate	0.132	0.463	0.278	0.721	0.189	0.689
	WHISCY	0.127	0.101	0.023	0.976	0.291	0.899
	PSIVER	0.103	0.645	0.463	0.536	0.118	0.546
BM90M	VORFFIP	0.092	0.681	0.576	0.423	0.284	0.488
	InterProSurf	0.075	0.405	0.293	0.706	0.145	0.673
	Pseudo-meta	0.290	0.563	0.158	0.841	0.225	0.816
	eFindSite ^{PPI}	0.242	0.303	0.064	0.935	0.312	0.880
	PredUs	0.135	0.366	0.177	0.822	0.165	0.782
	cons-PPISP	0.077	0.152	0.076	0.923	0.160	0.855
	PSIVER	0.103	0.645	0.463	0.536	0.118	0.546
	ProMate	0.101	0.571	0.417	0.582	0.162	0.580
	SPPIDER	0.096	0.537	0.371	0.628	0.122	0.620
	WHISCY	0.078	0.072	0.025	0.974	0.215	0.895
PIER	0.070	0.362	0.251	0.749	0.121	0.714	
VORFFIP	0.058	0.625	0.555	0.445	0.245	0.485	
InterProSurf	0.034	0.354	0.302	0.697	0.115	0.663	

Note: For each data set, web servers are sorted by MCC values. A pseudo-meta approach combines the best predictions produced by individual methods. BM90C = crystal structures; BM90H = high-quality models; BM90M = moderate-quality models; FPR = false positive rate; TPR = sensitivity; ACC = accuracy; SPC = specificity; PPV = precision; MCC = Matthew's correlation coefficient.

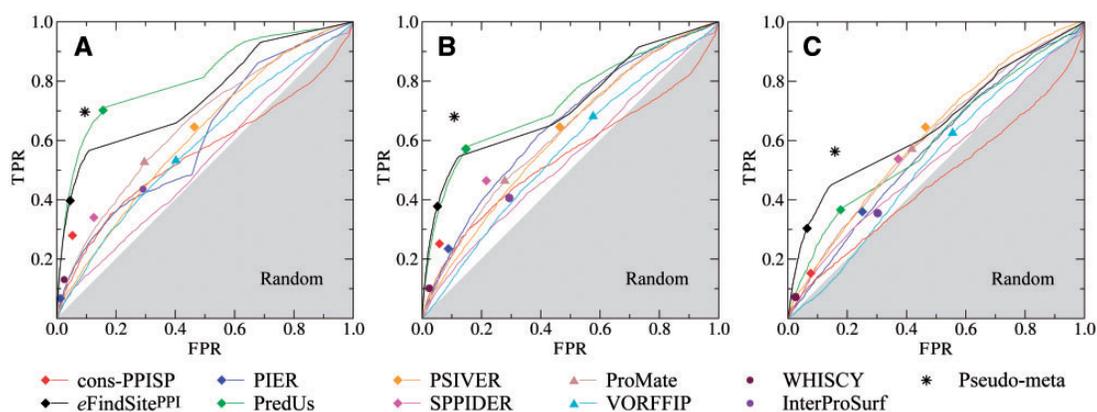


Figure 1. ROC plots assessing the accuracy of interface residue prediction by 10 web servers across three BM90 data sets. (A) Crystal structures, BM90C; (B) high-quality models, BM90H; and (C) moderate-quality models, BM90M. Continuous ROC lines are calculated using raw residue scores with triangles corresponding to the best performance of raw scores. Default predictions by web servers, including post-processing, are shown as diamonds and circles; circles are used for those web servers that do not provide continuous residue scores. Asterisks mark the accuracy of a pseudo-meta approach that combines the best predictions produced by individual algorithms.

above the corresponding continuous lines calculated from raw residue scores. For example, the improvement in MCC for SPIDDER (cons-PPISP) on the BM90C data set is 0.093 (0.078).

Performance of web servers using computer-generated models

Nine of 10 web servers described in this review are structure-based methods, i.e. they require the structure of a target protein. The performance of these predictors certainly depends on the quality of input structures. Despite a continuous growth of protein structure databases, there is still a huge gap between the number of known sequences and the number of solved structures. When the experimental structures of query proteins are unavailable, computer-generated models can be used in structure-based PPI residue prediction, however, assuming that the predictor tolerates distortions in modeled structures. To assess the impact of the quality of input structures on the prediction accuracy, we submitted high- (BM90H) and moderate-quality (BM90M) models of the target proteins to nine structure-based web servers.

Table 2 shows that all predictors give the best performance when experimental structures are used. The prediction accuracy of most algorithms significantly decreases from crystal structures to protein models. Interestingly, the ranking of web servers based on MCC is similar for all three BM90 data sets, except for eFindSite^{PPI}, which outperforms PredUs for BM90H and BM90M. For BM90H, the ranking is eFindSite^{PPI}, PredUs, cons-PPISP, SPPIDER, PIER, ProMate, WHISCY, PSIVER, VORFFIP and InterProSurf. Using high-quality models, eFindSite^{PPI} yields the best results with ACC of 0.898 and MCC of 0.340, thus its performance only slightly deteriorates with respect to the BM90C data set. PredUs is also fairly insensitive to small distortions in the input structures and still gives relatively high ACC of 0.827 and MCC of 0.309, in contrast to the remaining web servers; see Figure 1B.

For moderate-quality structures from the BM90M data set, the MCC-based ranking of web servers is eFindSite^{PPI}, PredUs, PSIVER, ProMate, SPPIDER, cons-PPISP, WHISCY, PIER, VORFFIP and InterProSurf. Notably, the performance of most web servers for the BM90M data set is significantly lower than for BM90C and BM90H, suggesting that these algorithms are sensitive to moderate distortions in the input structures. Also, while

post-processing enhances the performance across all target structures, the improvement for protein models is not as good as that obtained for crystal structures. Figure 1C demonstrates that eFindSite^{PPI} has the highest tolerance to structural deformations with ACC and MCC for the BM90M data set of 0.889 and 0.242, respectively. Similar to BM90H, PredUs is ranked second with ACC of 0.782 and MCC of 0.135. Note that the performance of sequence-based PSIVER is independent on the quality of input structures, thus remains constant across all BM90 data sets. For the BM90C and BM90H data sets, PSIVER is ranked ninth and eighth, respectively. Nonetheless, it is ranked as high as third on the BM90M data set, suggesting that the performance of most structure-based methods using moderate-quality structures is lower than that of sequence-based approaches. Among the algorithms tested here, eFindSite^{PPI} and PredUs are the only exceptions to this limitation.

We believe that the main reason for the high sensitivity to distortions in target structures of many structure-based approaches to PPI residue prediction is their strong dependence on fine atomic details. For instance, PIER uses local statistical properties of protein surface derived at the level of atomic groups; therefore, its high ACC of 0.906 for BM90C drops to 0.852 (0.714) for BM90H (BM90M). Similarly, ACC for SPPIDER, which uses atomic-level RSA-based fingerprints, drops by >7% (20%) when high- (moderate-) quality models are used instead of experimental structures. In contrast, eFindSite^{PPI} and PredUs use global structure alignments by Fr-TM-align and Ska, respectively, which make these predictors fairly insensitive to even moderate structural distortions in computer-generated models. Therefore, except for eFindSite^{PPI} and PredUs, most web servers require high-quality structural data to provide accurate PPI residue predictions.

Rationale for a meta-predictor

It has been reported that combining predictions by WHISCY and ProMate into an integrated approach called WHISCYMATE yields an improved accuracy of the identification of protein interface residues [46]. Another study demonstrated that meta-PPISP, a meta-predictor built on PINUP, cons-PPISP and ProMate, outperforms its component methods [92]. In the present study, we perform a similar analysis to determine whether combining 10 web servers improves the prediction accuracy over individual

algorithms using experimental and computer-generated structures. To address this issue, we first applied the Friedman test, a nonparametric alternative to the repeated measures analysis of variance [93], to MCC values calculated for web server predictions. *P*-values obtained for the BM90C, BM90H and BM90M data sets are 2.19×10^{-12} , 1.36×10^{-10} and 5.07×10^{-09} , respectively, indicating that individual algorithms produce statistically different results. Next, we selected the most accurate prediction for each target protein, referred to as a pseudo-meta approach. Note that this protocol is not a true meta-predictor; rather, it helps estimate the upper bound for the prediction accuracy given an optimal combination of individual algorithms. As presented in Figure 1 (black asterisks) and Table 2, the pseudo-meta approach systematically outperforms individual web servers with MCC for the BM90C, BM90H and BM90M data sets of 0.481, 0.443 and 0.290, respectively. The top three contributors to the best predictions are PredUs (38% for BM90C, 30% for BM90H and 13% for BM90M), eFindSite^{PPI} (29% for BM90C, 31% for BM90H and 36% for BM90M) and cons-PPISP (18% for BM90C, 8% for BM90H and 13% for BM90M). Lastly, we tested the differences between individual web servers and the pseudo-meta approach using the Wilcoxon signed-rank test, a nonparametric alternative to the paired Student's *t*-test [94]. In all cases, the pseudo-meta protocol outperforms web servers with statistically highly significant *p*-values of $<<0.01$.

Future work

Currently available web servers represent a diverse collection of algorithms for PPI residue prediction. Despite their relatively high accuracy obtained for experimentally solved target structures, using computer-generated models clearly yields less accurate predictions. Based on the results of our analysis, we suggest that post-processing protocols, which seem to quantitatively improve the results only for experimental structures, should be revisited and perhaps tuned up for the homology models of target proteins. Furthermore, meta-predictors should be systematically explored, for example, using techniques already extensively studied in protein threading [75, 95] and ligand binding site prediction [96, 97]. Here, we show that even a simple combination of outputs from various web servers gives a chance to outperform the best single method. More advanced meta-prediction techniques using nonlinear machine learning models are likely to further improve the accuracy of PPI residue prediction.

Key Points

- Easily accessible web servers provide the scientific community with convenient resources for protein-protein interaction site prediction.
- PredUs is the best-performing server for experimental target structures, but eFindSite^{PPI} gives the highest accuracy for computer-generated models.
- Post-processing procedures implemented in many web servers work well on experimental structures but need to be improved for protein models.
- Except for eFindSite^{PPI} and PredUs, structure-based methods are sensitive to moderate distortions in target structures.
- Meta-predictors will likely lead to significant improvements in interface residue prediction.

Acknowledgments

We thank Misagh Naderi who read the manuscript and provided critical comments.

Funding

This work was supported by the Louisiana Board of Regents through the Board of Regents Support Fund [contract LEQSF(2012-15)-RD-A-05].

References

1. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;**437**:1173–78.
2. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 2007;**450**:1001–9.
3. Jubb H, Higuero AP, Winter A, et al. Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol Sci* 2012;**33**:241–8.
4. Sowa ME, He W, Wensel TG, et al. A regulator of G protein signaling interaction surface linked to effector specificity. *Proc Natl Acad Sci USA*. 2000;**97**:1483–8.
5. Sowa ME, He W, Slep KC, et al. Prediction and confirmation of a site critical for effector regulation of RGS domain activity. *Nat Struct Biol* 2001;**8**:234–7.
6. Halperin I, Ma B, Wolfson H, et al. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 2002;**47**:409–43.
7. Chelliah V, Blundell TL, Fernández-Recio J. Efficient restraints for protein-protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment. *J Mol Biol* 2006;**357**:1669–82.
8. Li B, Kihara D. Protein docking prediction using predicted protein-protein interface. *BMC Bioinformatics* 2012;**13**:7.
9. Rigaut G, Shevchenko A, Rutz B, et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 1999;**17**:1030–2.
10. Sobott F, Robinson C V. Protein complexes gain momentum. *Curr Opin Struct Biol* 2002;**12**:729–34.
11. Yates JR. Mass spectrometry from genomics to proteomics. *Outlook* 2000;**16**:5–8.
12. Pellegrini M, Marcotte EM, Thompson MJ, et al. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999;**96**:4285–8.
13. Huynen MA, Snel B, Von Mering C, et al. Function prediction and protein networks. *Curr Opin Cell Biol*. 2003;**15**: 191–8.
14. Pazos F, Valencia A. In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins Struct Funct Genet* 2002;**47**:219–27.
15. Jones S, Thornton JM. Prediction of protein-protein interaction sites using patch analysis. *J Mol Biol* 1997;**272**:133–43.
16. Obenaus J, Yaffe M. Computational prediction of protein-protein interactions. *Methods Mol Biol* 2004;**261**:445–68.
17. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins* 2007;**64**:5:630–45.
18. Pitre S, Alamgir M, Green JR, et al. Computational methods for predicting protein-protein interactions. *Adv Biochem Eng Biotechnol* 2008;**110**:247–67.

19. Wang B, Sun W, Zhang J, et al. Current status of machine learning-based methods for identifying protein-protein interaction sites. *Curr Bioinform* 2013;**8**:177–82.
20. Ofra Y, Rost B. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett* 2003;**544**:236–9.
21. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 1996;**257**:342–58.
22. Chen X, Jeong JC. Sequence-based prediction of protein interaction sites with an integrative method. *Bioinformatics* 2009;**25**:585–91.
23. Sikić M, Tomić S, Vlahovick K. Prediction of protein-protein interaction sites in sequences and 3D structures by random forests. *PLoS Comput Biol* 2009;**5**:e1000278.
24. Zhou H-X, Qin S. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics* 2007;**23**:2203–9.
25. de Vries SJ, Bonvin AMJJ. How proteins get in touch: interface prediction in the study of bio-molecular complexes. *Curr Protein Pept Sci* 2008;**9**:394–406.
26. Ezkurdia I, Bartoli L, Fariselli P, et al. Progress and challenges in predicting protein-protein interaction sites. *Brief Bioinform* 2009;**10**:233–46.
27. Caffrey DR, Somaroo S, Hughes JD, et al. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci* 2004;**13**:190–202.
28. Kufareva I, Budagyan L, Raush E, et al. PIER: protein interface recognition for structural proteomics. *Proteins* 2007;**41**:400–17.
29. Ofra Y, Rost B. ISIS: interaction sites identified from sequence. *Bioinformatics* 2007;**23**:e13–16.
30. Bordner AJ, Abagyan R. Statistical analysis and prediction of protein-protein interfaces. *Proteins* 2005;**60**:353–66.
31. Maheshwari S, Brylinski M. Prediction of protein-protein interaction sites from weakly homologous template structures using meta-threading and machine learning. *J Mol Recognit* 2015;**28**:35–48.
32. Chen H, Zhou H-X. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins* 2005;**61**:21–35.
33. Jones S, Thornton JM. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* 1997;**272**:121–32.
34. Jones S, Thornton JM. Review Principles of protein-protein interactions. *Proc Natl Acad Sci USA* 1996;**93**:13–20.
35. Fiorucci S, Zacharias M. Prediction of protein-protein interaction sites using electrostatic desolvation profiles. *Biophys J* 2010;**98**:1921–30.
36. Murakami Y, Mizuguchi K. Applying the Naïve Bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. *Bioinformatics* 2010;**26**:1841–8.
37. Negi SS, Schein CH, Oezguen N, et al. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics* 2007;**23**:3397–9.
38. Liang S, Zhang C, Liu S, et al. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res* 2006;**34**:3698–707.
39. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* 2004;**338**:181–99.
40. Segura J, Jones PF, Fernandez-Fuentes N. Improving the prediction of protein binding sites by combining heterogeneous data and Voronoi Diagrams. *BMC Bioinformatics* 2011;**12**:352.
41. Jordan RA, El-Manzalawy Y, Dobbs D, et al. Predicting protein-protein interface residues using local surface structural similarity. *BMC Bioinformatics* 2012;**13**:41.
42. Grimm V, Arakaki K, Skolnick J. Prediction of physical protein–protein interactions. *Phys Biol* 2005;**2**:S1–16.
43. Li JJ, Huang DS, Wang B, et al. Identifying protein-protein interfacial residues in heterocomplexes using residue conservation scores. *Int J Biol Macromol* 2006;**38**:241–7.
44. Wang B, Chen P, Huang DS, et al. Predicting protein interaction sites from residue spatial sequence profile and evolution rate. *FEBS Lett* 2006;**580**:380–4.
45. Wang B, Wong HS, Huang D-S. Inferring protein-protein interacting sites using residue conservation and evolutionary information. *Protein Pept Lett* 2006;**13**:999–1005.
46. de Vries SJ, Van Dijk ADJ, Bonvin AMJJ. WHISCY: what information does surface conservation yield? Application to Data-Driven Docking. *Proteins* 2006;**48**:479–89.
47. Zhou HX, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* 2001;**44**:336–43.
48. Koike A, Takagi T. Prediction of protein-protein interaction sites using support vector machines. *Protein Eng Des Sel* 2004;**17**:165–73.
49. Zhang QC, Petrey D, Norel R, et al. Protein interface conservation across structure space. *Proc Natl Acad Sci USA* 2010;**107**:10896–901.
50. Armon A, Graur D, Ben-Tal N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* 2001;**307**:447–63.
51. Martin J. Benchmarking protein-protein interface predictions: why you should care about protein size. *Proteins* 2014;**82**:1444–52.
52. Dunker AK, Lawson JD, Brown CJ, et al. Intrinsically disordered protein. *J Mol Graph Model* 2001;**19**:26–59.
53. Hsu WL, Oldfield CJ, Xue B, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci* 2013;**22**:258–73.
54. Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins Struct Funct Genet* 2007;**66**:761–5.
55. Mohan A, Oldfield CJ, Radivojac P, et al. Analysis of molecular recognition features (MoRFs). *J Mol Biol* 2006;**362**:1043–59.
56. Mooney C, Pollastri G, Shields DC, et al. Prediction of short linear protein binding regions. *J Mol Biol* 2012;**415**:193–204.
57. Disfani FM, Hsu WL, Mizianty MJ, et al. MoRFPred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 2012;**28**:75–83.
58. Dosztányi Z, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 2009;**25**:2745–6.
59. Fuxreiter M, Tóth-Petróczy Á, Kraut DA, et al. Disordered proteinaceous machines. *Chem Rev* 2014;**114**:6806–43.
60. Zhang QC, Deng L, Fisher M, et al. PredUs: a web server for predicting protein interfaces using structural neighbors. *Nucleic Acids Res* 2011;**39**:W283–7.
61. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.
62. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;**22**:2577–637.
63. Wagner M, Adamczak R, Porollo A, et al. Linear regression models for solvent accessibility prediction in proteins. *J Comput Biol* 2005;**12**:355–69.

64. Hwang H, Pierce B, Mintseris J, et al. Protein-protein docking benchmark version 3.0. *Proteins* 2009;**73**:705–9.
65. Negi SS, Braun W. Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. *J Mol Model* 2007;**13**:1157–67.
66. Fraczkiwicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* 1998;**19**:319–33.
67. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 2001;**17**:700–12.
68. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 1991;**9**:56–68.
69. De Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 2010;**5**:883–97.
70. Chen R, Mintseris J, Janin J, et al. A protein-protein docking benchmark. *Proteins Struct Funct Genet* 2003;**52**:88–91.
71. Fernandez-Recio J, Totrov M, Skorodumov C, et al. Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins* 2005;**58**:134–43.
72. Abagyan R, Totrov M, Kuznetsov D. ICM - a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994;**15**:488–506.
73. Abagyan RA, Batalov S. Do aligned sequences share the same fold? *J Mol Biol* 1997;**273**:355–68.
74. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 1983;**221**:709–13.
75. Brylinski M, Lingam D. eThread: a highly optimized machine learning-based approach to meta-threading and the modeling of protein tertiary structures. *PLoS One* 2012;**7**:e50200.
76. Pandit SB, Skolnick J. Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinformatics* 2008;**9**:531.
77. Zhang H. The optimality of naive bayes. *Mach Learn* 2004;**1**:3.
78. Engelen S, Trojan LA, Sacquin-Mora S, et al. Joint evolutionary trees: a large-scale method to predict protein interfaces based on sequence sampling. *PLoS Comput Biol* 2009;**5**:e1000267.
79. Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J Mol Biol* 2000;**301**:665–78.
80. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. *Trends Biochem Sci* 1998;**23**:358–61.
81. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–42.
82. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 1991;**11**:281–96.
83. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011;**2**:1–39.
84. Keskin O, GURSOY A, Ma B, et al. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem Rev* 2008;**108**:1225–44.
85. Gene T, Consortium O. Gene ontology: tool for the. *Nat Genet* 2000;**25**:25–9.
86. Hubbard TJP, Ailey B, Brenner SE, et al. SCOP: a structural classification of proteins database. *Nucleic Acids Res* 1999;**27**:254–6.
87. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic Acids Res* 2014;**42**:D222–30.
88. Mulder NJ, Apweiler R, Attwood TK, et al. InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief Bioinform* 2002;**3**:225–35.
89. Hwang H, Vreven T, Janin J, et al. Protein-protein docking benchmark version 4.0. *Proteins* 2010;**78**:3111–14.
90. Brylinski M, Feinstein WP. Setting up a meta-threading pipeline for high-throughput structural bioinformatics: eThread software distribution, walkthrough and resource profiling. *J Comput Sci Syst Biol* 2012;**6**:1–10.
91. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins* 2004;**57**:702–10.
92. Qin S, Zhou H-X. meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics* 2007;**23**:3386–7.
93. Friedman M. A comparison of alternative tests of significance for the problem of m rankings. *Ann Math. Statist* 1940;**11**:86–92.
94. Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull* 1945;**1**:80–3.
95. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 2007;**35**:3375–82.
96. Zhang Z, Li Y, Lin B, et al. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* 2011;**27**:2083–8.
97. Chen K, Mizianty MJ, Gao J, et al. A critical comparative assessment of predictions of protein-binding sites for biologically relevant organic compounds. *Structure* 2011;**19**:613–621.