Exploiting Conformational Ensembles in Modeling Protein–Protein Interactions on the Proteome Scale

Guray Kuzu,[†] Attila Gursoy,^{*,†} Ruth Nussinov,^{‡,§} and Ozlem Keskin^{*,†}[†]Center for Computational Biology and Bioinformatics and College of Engineering, Koc University Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey[‡]Basic Science Program, SAIC-Frederick, Inc., National Cancer Institute, Center for Cancer Research Nanobiology Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States[§]Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Supporting Information

ABSTRACT: Cellular functions are performed through protein–protein interactions; therefore, identification of these interactions is crucial for understanding biological processes. Recent studies suggest that knowledge-based approaches are more useful than “blind” docking for modeling at large scales. However, a caveat of knowledge-based approaches is that they treat molecules as rigid structures. The Protein Data Bank (PDB) offers a wealth of conformations. Here, we exploited an ensemble of the conformations in predictions by a knowledge-based method, PRISM. We tested “difficult” cases in a docking-benchmark data set, where the unbound and bound protein forms are structurally different. Considering alternative conformations for each protein, the percentage of successfully predicted interactions increased from ~26 to 66%, and 57% of the interactions were successfully predicted in an “unbiased” scenario, in which data related to the bound forms were not utilized. If the appropriate conformation, or relevant template interface, is unavailable in the PDB, PRISM could not predict the interaction successfully. The pace of the growth of the PDB promises a rapid increase of ensemble conformations emphasizing the merit of such knowledge-based ensemble strategies for higher success rates in protein–protein interaction predictions on an interactome scale. We constructed the structural network of ERK interacting proteins as a case study.

KEYWORDS: protein–protein interaction prediction, PRISM, structural network, knowledge-based method, conformations, docking

■ INTRODUCTION

Biological processes take place through protein–protein interactions (PPIs). They are crucially important for cellular function, regulation and signaling. Following the Human Genome Project, high-throughput studies have become popular and PPIs have also been investigated on large scales. Proteome-level studies elucidate cellular functions,^{1,2} regulation,^{3,4} disease mechanisms,^{5,6} conservation through evolution,^{7–9} drug discovery¹⁰ and drug side-effects.^{11,12} A key first step in such studies is determining the PPIs. PPIs on large scales are identified by experimental techniques like the yeast two-hybrid system,¹³ phage display,¹⁴ protein arrays,¹⁵ and affinity purification.¹⁶ The output of these experiments can include false negatives and false positives.^{17–20} Experimental structural techniques offer more reliable data. The structures of interacting proteins illustrate not only that they interact, but also how they interact. X-ray crystallography,²¹ nuclear magnetic resonance (NMR) spectroscopy,²² cryo-electron microscopy (Cryo-EM)²³ and small-angle X-ray scattering (SAXS)²⁴ provide the 3-dimensional (3D) architectures of the

proteins and their interactions, and high resolution structures are available in the PDB.²⁵ The number of structures in the PDB grows exponentially; there are more than 89 000 structures as of March 2013, and more than a quarter of them have been added after 2009.

Computational approaches can assist experiments in verifying results and in predicting new interactions; in addition, they are cheaper and faster. They can be classified into two categories: “blind” docking and knowledge-based methods. Blind docking methods^{26–29} search for the “best” (i.e., native) bound state by considering a large number of possible structural combinations of query proteins. They are computationally expensive. Further, in the absence of biological, functional knowledge that the proteins interact and some data on the interaction site, their predictions may not be reliable. This is because there will always be some favorable modes for any two proteins to interact. Blind docking is

Received: January 3, 2013

Published: April 16, 2013

impractical for large-scale studies, particularly if new interactions are sought, where no data are available on whether the proteins interact. Knowledge-based methods^{30–35} may fare better: they are based on the notion that motifs recur in nature.^{35–39} They compare the surfaces of query proteins with known protein–protein structures of interacting protein pairs. Since there are fewer templates (derived from known protein interactions), the process is faster and more affordable and thus applicable on large-scale. Knowledge-based methods are appropriate techniques to construct interactome-scale structural networks. Even though the PDB covers a limited number of protein–protein interactions, a systematic large scale study⁴⁰ showed that templates are available to model complexes of structurally characterized proteins. However, on the down side, knowledge-based approaches typically consider the structures as rigid bodies, even though proteins are flexible^{41–44} and their preferred conformational states are expected to change with the environment. Neighboring molecules, atoms, or ions can redistribute their ensembles;^{45,46} this redistribution is dynamic^{47–51} and is reflected in the observed protein structures. Binding, post-translational modifications, changes in ligand concentration, pH and ionic strength of the medium will affect the conformational distribution. The ensembles confer on the protein the ability to function.^{49–51} The bound forms of the proteins can be obtained from the unbound forms in *in silico* experiments;⁵² however, protein flexibility and conformational changes upon binding are mostly ignored in large scale knowledge-based methods. This limits the capability and success in prediction and, therefore, the construction of the structural network.

In principle, different conformations of proteins can be considered by knowledge-based methods, and this can be expected to improve the predictions. The PDB offers many protein structures including different conformations of the same protein. These conformations may include bound, unbound or any alternative forms (e.g., following allosteric post-translational modifications, or bound to different ligands) of the proteins, which can be utilized. The collection of all available structures of a protein provides a subset of the repertoire of its conformations under different conditions and can constitute the input rather than a single structure. These conformations would help to more reliably figure out whether the proteins interact and how they interact, and identifying binary protein interactions is the first step in the construction of the structural network. We tested a motif-based protein interaction prediction tool, Protein Interactions by Structural Matching (PRISM),^{33,35,53} on a docking benchmark data set,⁵⁴ to see if providing different conformations of proteins would indeed increase PRISM's capability to detect interactions. As we have shown earlier, if the structures are available and the proteins are known to play a role in the same pathway, PRISM can successfully predict if two proteins in these interact (76 and 78% accuracies in the ubiquitination⁵⁵ and apoptosis⁵⁶ pathways, respectively). We also demonstrated that it can be used to construct structural networks.^{57,58} This knowledge-based tool can successfully predict the “easy” cases of the docking benchmark data set;⁵⁹ thus, in this study we targeted the “difficult” cases toward proteome applications. We found the alternative structures of a protein through sequence homology and structural alignment. If two PDB entries of the same protein have different structures, they are included as different molecules in the target set. The interactions between the two proteins are predicted and compared with the bound

state given in the benchmark data set. The success of the prediction is assessed by the global energy values of the predicted complex and by IS-score,⁶⁰ which shows the structural similarity between the predicted molecule and the bound state of the proteins. As a case study we constructed the structural network of the extracellular signal-regulated kinases (ERK) interactions and show that considering alternative conformations of the query proteins improve the modeled structural network.

MATERIALS AND METHODS

In this section, we first describe PRISM and then the method used to find different conformations of the proteins together with the data set.

Protein Interactions by Structural Matching (PRISM)

The flowchart of PRISM is illustrated in Figure 1 (flow through blue boxes; steps 1–4) and explained below. PRISM was

Figure 1. The flowchart of PRISM. Cyan box shows the flowchart of the PRISM algorithm. Template and target data sets are inputs, and 3-dimensional data and global energies of predicted interactions are outputs of the algorithm. Step 0: Template data set organization; step 1: surface extraction of target proteins; step 2: structural alignment of target surfaces with template interfaces; step 3: elimination of clashing structures; step 4: flexible refinement.

implemented in Python and runs in a UNIX environment. Details on how to run PRISM were given earlier.³⁵ The PRISM source codes and external programs can be downloadable at http://prism.ccbb.ku.edu.tr/prism_protocol/. First, external programs, FASTA version 35,⁶¹ NACCESS,⁶² MultiProt⁶³ and FiberDock,⁶⁴ need to be installed. The user enters target protein names in the file named “PDB.list” in “PRISM_protocol/0-SurfaceExtraction” directory, and the PDB files are downloaded from the PDB Web page (<http://www.pdb.org>). Surfaces of target proteins are extracted using the script, and output files are copied to the structural matching directory, “PRISM_protocol/1-Prediction”. There, the user selects template interfaces. We used all template interfaces in this study. Structural matching between target protein surfaces and template interfaces is done using the script. Transformation of the structures and filtering are done in “PRISM_protocol/2-DistanceCalculation” directory. At the last step, flexible refinement and energy calculation are done in directory “PRISM_

protocol/6-FiberDock". When PRISM run is complete, structures of predictions and their energies are reachable in "FIBER-DOCK_Structures" and "ENERGIES" directories, respectively.

Template Data Set Organization of PRISM

The template data set includes a nonredundant set of protein–protein interface structures.³⁸ Interfaces of all two pair interactions in the PDB were extracted (Figure 1, step 0). Interfaces consist of "contacting" and "nearby" residues. If two residues from each interacting protein are close enough, they are labeled as contacting residues. The cutoff value is the sum of the van der Waals radii of the heavy atoms plus 0.5 Å. The nearby residues constitute the scaffold of an interface. They are the neighboring residues whose C α atoms are at most 6.0 Å away from the C α atom of a contacting residue. Interfaces are structurally clustered to obtain a nonredundant set. Each cluster has members that are structurally similar to the representative interface of the cluster. Computational hotspots of representative interfaces are found via the Web server, HotPoint.⁶⁵

Target Data Set of PRISM

Query proteins among whose interactions are to be searched constitute the target set (Figure 1). These can be assembled by the user.

Prediction Algorithm of PRISM

First, the surfaces of the target proteins are extracted using the NACCESS program (Figure 1, step 1).⁶² Then, PRISM checks the structural similarity between target surfaces and template interfaces. Structural similarity is searched via alignment of target surfaces onto template interfaces using MultiProt⁶³ (Figure 1, step 2). If a target surface is structurally similar to one side of a template interface and another target surface is structurally similar to the complementary side, these two target structures may interact. To guarantee a proper match between a target surface and a template interface in the alignment, PRISM checks if the matched residues from the two sides are against each other and if at least one residue of the target surface matches with a hotspot of the template interface. The candidate protein complexes are next physically and biologically evaluated. Clashes between residues of the two structures are counted (Figure 1, step 3). If there are ≥ 5 clashes among alpha-carbons, the candidate complex is discarded. After that, flexible refinement and energy calculation are done using FiberDock⁶⁴ (Figure 1, step 4). Side-chains of the structures are oriented to have a more favorable state. Hydrogen atoms are also considered in this process. The backbones of the structures can be slightly reoriented in this refinement step. Finally, the global energy of the candidate complex is calculated. Complexes with energy lower than the threshold value are considered as biologically meaningful, i.e., that these proteins interact. PRISM gives the atomic coordinates of the potentially interacting proteins.

Enlarging the Target Data Set Using Different Conformations

Different conformations of target proteins are found in the PDB. First, chains of structures which have the same sequence as the query protein are detected (Figure 2B). 100% (and then 95%) FASTA sequence homology between the molecules is considered. Then, the molecular structures are compared using MultiProt⁶³ (Figure 2C). If MultiProt matches the candidate structure with less than 90% of the query structure and the root-mean-square-deviation (RMSD) value between the matched residues of the two structures is more than 2.0 Å,

Figure 2. Exploiting different conformations to enlarge target set. (A) The standard target set includes one conformation of each protein. (B) The PDB structures of the standard target proteins are found using sequence homology. (C) Conformations are clustered using structural alignment. Structurally similar conformations are given in the same column. Representatives of the clusters are alternative conformations of the standard target proteins. (D) The target data set is enlarged with the different conformations of the standard target proteins.

the candidate structure is considered as a different conformation of the query protein. Different conformations of the target proteins are added to the target set (Figure 2D), and PRISM is run with the enlarged target set to test if the prediction improves.

Docking Benchmark Data Set and Prediction with PRISM

We tested the method on difficult cases of a docking benchmark data set.⁵⁴ The benchmark set provides the bound state of the two given molecules in each case. In some cases, structures are multimeric and there is more than one binary interaction. In these cases (cases 1, 2, 11, 12 and 18), each binary interaction was considered separately. A list of all interactions studied for 30 difficult cases is given in Table S1 (Supporting Information). The structural difference between the bound and unbound forms determines the case-type of the interaction: difficult cases of the benchmark data set include structures with interface root-mean-square distance (iRMSD) larger than 2.2 Å.⁵⁴ Figures 3 and 4 provide examples of allostery⁶⁶ and the conformational changes upon binding. First, PRISM was run with the target set containing only the two molecules from the benchmark to predict their bound state (Figure 5A). Second, different conformations of the two molecules were found (Table S2, Supporting Information; more than one chain name indicates that the molecule is similar to each chain) and added to the target set (Figure 2). Then, PRISM was run with the enlarged target set (where this is repeated for each case in the docking benchmark) (Figure 5B). The energy value cutoff, -10 kJ/mol, was used to determine favorable predictions. Table S3 (Supporting Information) provides the FiberDock energies of the bound forms for the benchmark data set. All interfaces except 2hmiCB and 1h1vAG (28 out of 30 interactions) have low energies (below -10 kJ/mol). On the basis of our previous studies,^{35,55,56,58,67} we set the cutoff value as -10 kJ/mol. The success of a prediction was evaluated on the basis of IS-score,⁶⁰ which is a metric to evaluate protein–protein interaction predictions. The IS-score does not consider the equivalence of target and template

Figure 3. An example of allosteric effect on binding: Case 17. (A) Unbound form of heparin cofactor II (blue, 1jmjA). (B) Allosteric effect of glycosaminoglycan (GAG) leads to conformational change in heparin cofactor II. Unbound (blue) and bound (cyan) forms of heparin cofactor II are given together. (C) Bound forms of heparin cofactor II (cyan: 1jmoA) and thrombin (orange: heavy chain, 1jmoH; green: light chain, 1jmoL) are shown separately. (D) Interaction of heparin cofactor II and thrombin.

Figure 4. An example of conformational change upon binding: Case 12ii. (A) Unbound forms of soluble tissue factor (purple, 1tfhB) and blood coagulation factor VIIA (blue: 1qfkL). (B) Conformational changes are shown, unbound and bound forms of the soluble tissue factor are given in purple and yellow, and unbound and bound forms of the soluble tissue factor are given in blue and green, respectively. Large conformational change of the blood coagulation factor VIIA is indicated via an arrow. (C) Bound forms are shown separately. (D) Interaction of soluble tissue factor (1fakT) and blood coagulation factor VIIA (1fakL).

residues,⁶⁰ unlike the critical assessment of predicted interactions (CAPRI) criteria,⁶⁸ which are based on RMSD and native contact fractions. These could present problems because PDB structures may lack residues or domains (because of, e.g., conformational disorder), and alternative conformations of the query proteins may differ in length. Predicted models are considered as “near native” if the IS-score is >0.17 , “acceptable” if the IS-score is between 0.12 and 0.17, and “incorrect” if the IS-score is <0.12 . We also presented the native (f_{nat}) and the

Figure 5. PRISM predictions with standard and enlarged target data sets. (A) PRISM is run using standard target data set. No favorable interaction can be found for the given conformations of the proteins. (B) PRISM is run using enlarged target data set. A favorable interaction is found with one conformation of each protein.

non-native ($f_{\text{non-nat}}$) contact fractions: f_{nat} is the number of contacts correctly predicted divided by the number of the native contacts, and $f_{\text{non-nat}}$ is the number of contacts incorrectly predicted divided by the number of the predicted contacts. IS-score detects substructure similarity and characterizes the similarity of the interfaces of the complex. The alignment exploits iAlign,³¹ and IS-score compares only matched residues. Zero contact value indicates the model and the native complex have no good match and the model is incorrect. In addition, we present the performance of PRISM as the ratio of successful predictions, which were evaluated by IS-score (IS-score >0.12), to the energetically favorable PRISM predictions (energy <-10 kJ/mol). The percentage of favorable predictions of PRISM that successfully match the bound forms is given for the benchmark data set.

Eliminating Bias by Discarding Information on the Bound Form

Bound forms of the proteins exist in the PDB. These forms can appear as alternative structures of the unbound forms or the template interface. For example, in interaction 14, the bound form given in the benchmark data set is 1ibrAB. 1ibrA and 1ibrC are found as alternative structures of the unbound form 1qg4A, and 1ibrB is for 1f59A (Table S2, Supporting Information). Moreover, the template 1ibrAB is used to predict this interaction. To eliminate bias, the bound forms were discarded from the target and template data sets (Figure 6). First, PDB entries of the bound forms were discarded from the PDB list; then alternative forms were found. These new alternative target structures are given in Table S4 (Supporting Information). Second, if a template interface was the interface of a bound form, another interface from the same template cluster was chosen as the representative. If the template cluster had no other member, this template was eliminated. The changes in the template organization are shown in Table S5 (Supporting Information). Since the bound forms are given also as the unbound forms in cases 2i (molecule: 2hmiC) and 2ii (molecule: 2hmiD) in the benchmark data set, these cases were not considered, and the remaining 28 interactions were processed under this scenario.

Figure 6. Prediction using or not using the structures of bound forms. The benchmark data set gives the unbound and the bound forms of the molecules. Different color tones represent different conformations of proteins shown in green and blue. The green-cyan template represents the interface of the bound forms. We followed two scenarios in PRISM predictions. (A) First, we found alternative conformations from all PDB structures and run PRISM using all templates. (B) Second, we eliminated bound forms from the PDB structures and then found alternative conformations. In addition, if a template interface was the interface of a bound form, another interface from the same template cluster was chosen as the representative. If the template cluster had no other member, this template was eliminated.

RESULTS AND DISCUSSION

We obtained two sets of predictions. First, PRISM was run with a “standard” target set of given unbound structures (Figure 5A), one conformation for each molecule. Then, the target set was

enlarged with the different conformations of the unbound forms of each protein available in the PDB (Table S2, Supporting Information, Figures 2 and 5B). The first target set had two conformations (a size of 2) in all cases, and on average the enlarged target sets had a size of 8.87, and the median value was 6.50, which did not affect considerably the computation process. Using standard target sets, 9 out of 30 difficult interactions were predicted as energetically favorable (-10 kJ/mol cutoff value; Table S6, Supporting Information). In the second set, when the target sets were enlarged with the different conformations of the given unbound structures, 7 of these 9 interactions were better predicted, with lower global energy values (interactions 2i, 11i, 11ii, 12i, 16, 18ii and 19). Energetically favorable predictions (an increase from 9 to 24) were obtained for additional interactions when enlarged target sets were used (Table S6, Supporting Information). We note that energetically favorable models do not necessarily correspond to the native complex structures of the benchmark. Energetically favorable interactions together with their templates are given in Table S7a,b (Supporting Information). If the chain is specified, the interacting molecule is the monomer; otherwise, it is the whole structure. If more than one chain is listed, the chains are structurally similar. IS-scores and contact fractions of the predictions obtained using alternative conformations are provided in Table 1 (Table S8, Supporting Information, shows the results of predictions obtained using the standard target set). As to IS-scores, among the predictions obtained using standard target sets, 7 cases were near native, 1 case was acceptable, and 1 case was incorrect (Table S8, Supporting Information), and among predictions obtained using enlarged target sets, 18 cases were near native and 6 cases were incorrect (Table 1). When we eliminated incorrect predictions, we had 8 successful predictions with the standard

Table 1. Energies, IS-Scores and Contact Fractions of PRISM Predictions Obtained Using Standard Template Set and Enlarged Target Set

difficult case	energy (kJ/mol)	IS-score	rating	real contacts	model contacts	common contacts	f_{nat}	$f_{\text{non-nat}}$
1i	-21.85	0.2366	near native	17	27	16	0.94	0.41
2i	-21.69	0.0648	incorrect	0	39	0	0.00	1.00
2ii	-63.70	0.0751	incorrect	0	72	0	0.00	1.00
4	-68.75	0.4041	near native	21	38	18	0.86	0.53
5	-74.40	0.6640	near native	65	74	55	0.85	0.26
8	-46.33	0.3775	near native	38	46	21	0.55	0.54
9	-114.68	0.6956	near native	86	113	80	0.93	0.29
11i	-63.43	0.6691	near native	47	57	41	0.87	0.28
11ii	-64.65	0.6036	near native	38	46	34	0.89	0.26
12i	-57.47	0.4789	near native	15	29	15	1.00	0.48
12ii	-102.61	0.5916	near native	53	76	47	0.89	0.38
13	-106.01	0.5890	near native	46	58	41	0.89	0.29
14	-144.40	0.7976	near native	79	85	68	0.86	0.20
15	-63.74	0.6283	near native	53	71	48	0.91	0.32
16	-49.32	0.6967	near native	62	78	56	0.90	0.28
17	-21.97	0.0659	incorrect	0	32	0	0.00	1.00
18i	-20.13	0.4864	near native	19	23	18	0.95	0.22
18ii	-50.84	0.4942	near native	35	38	26	0.74	0.32
19	-140.42	0.9363	near native	93	100	93	1.00	0.07
20	-58.42	0.6201	near native	38	44	29	0.76	0.34
21	-14.77	0.1770	near native	10	21	6	0.60	0.71
22	-40.35	0.0696	incorrect	0	50	0	0.00	1.00
23	-42.70	0.0679	incorrect	0	36	0	0.00	1.00
24	-43.80	0.0700	incorrect	9	43	0	0.00	1.00

target set and 18 with the enlarged set. In addition, if a different prediction was obtained when the enlarged target set was used, the prediction had a higher IS-score.

PRISM was also run without using the data of the bound forms in the benchmark data set to eliminate any bias. The results of this scenario (see Materials and Methods, “Eliminating Bias by Discarding Information on the Bound Form”) are given in Table S9 (Supporting Information). Enlarging the target set by including alternative conformations increased the number of energetically favorable predictions from 6 to 21 out of 28 interactions (5 of these 6 were predicted better, with lower global energy), with the cutoff energy set at -10 kJ/mol. Favorable interactions together with their templates and global energy values are listed in Table S10a,b (Supporting Information). IS-scores and contact fractions are provided in Table S11a,b (Supporting Information). According to the IS-scores classifications, all 6 predictions obtained using standard target sets were near native, and among predictions using enlarged target sets, 14 cases were near native and 7 cases were incorrect. When we eliminated incorrect predictions, there were 6 successful predictions with the standard target set and 14 with the enlarged set. In addition, higher IS-scores were obtained (except for case 16 with better energy but lower IS-score) for predictions with the enlarged target set.

95% Sequence Homology Was Used to Enlarge the Target Set More

Up to here, we considered 100% sequence homology for alternative conformations from the PDB. Analysis of the sequence homology between unbound and bound forms given in the benchmark data set illustrates that some are not 100% similar (Table S12, Supporting Information). However, reducing sequence homology much below 100% can lead to picking different proteins. We reduced this value to 95% for cases that we could not obtain successful predictions (for 16 interactions: 1ii, 2i, 2ii, 3, 4, 6, 7, 10, 14, 17, 18i, 20, 22, 23, 24, and 25; for cases 4, 14, 18i and 20, unsuccessful results were obtained when bound data was eliminated from the template and target sets; therefore, we added these to this list) to enlarge the target set. We used the same criteria to identify structurally different conformations as described in Materials and Methods, “Enlarging the Target Data Set Using Different Conformations”. These additional alternative conformations are given in Table S13 (Supporting Information) (more than one chain name indicates that the molecule is similar to each chain). No additional conformations were detected for the unbound forms of cases 7, 10, 22, 23, and 24. Although additional favorable predictions were obtained for cases 1ii, 2i, 2ii, 3, 4, 17, 20, and 25 by using these conformations (Table S14, Supporting Information), only successful results were obtained for cases 1ii, 3 and 14 (IS-score >0.12). Energy values, IS-scores and contact fractions are given in Table S15 (Supporting Information).

To enlarge the target sets, we also tested 90 and 85% sequence homology. Lower sequence homology yields more structures; however, it increases the risk of picking a different protein. For target sets enlarged with 90% sequence homology, we tested all templates. Because of computational cost, for target sets enlarged with 85% sequence homology, we tested templates that successfully matched target proteins found with higher sequence homology. At 95% sequence homology, successful results were obtained for 3 cases; the 90 and 85% sequence homology cases failed to give additional successful predictions (data not shown). 95% sequence homology covers

sequence difference between unbound and bound forms in the benchmark data set, where the lowest sequence similarity is 98.8% (Table S12, Supporting Information). Therefore, we consider 95% sequence homology as appropriate to further enlarge the target set.

When we could use the structures of bound forms, the success was 18 out of 30 difficult interactions at 100% sequence homology, and considering additional successful results of cases 1ii and 3 increased the number to 20. When we did not use the data of the bound forms (see Materials and Methods, “Eliminating Bias by Discarding Information on the Bound Form”), the success was 14 out of 28 difficult interactions at 100% sequence homology, and considering additional successful results of cases 1ii and 14, it increased to 16. The final success rates are given in Table 2. We also present the

Table 2. Success of PRISM Predictions

cases	total interaction	predicted interactions using standard target set		predicted interactions using enlarged target set	
	number	number	percentage	number	percentage
difficult cases (also using data of bound forms)	30	8	26.7	20	66.7
difficult cases (not using data of bound forms)	28	6	21.4	16	57.1

ratio of successfully predicted results vs favorable predictions in Table 3, which indicates the percentage of predictions that

Table 3. Successful PRISM Predictions in Energetically Favorable PRISM Predictions

predictions	favorable PRISM predictions (energy < -10 kJ/mol)	successful PRISM predictions (IS-score >0.12)	percentage of favorable PRISM results matching bound forms
standard target set, standard template set	9	8	88.9
enlarged target set, standard template set	26	20	76.9
standard target set, unbiased template set	6	6	100
enlarged target set, unbiased template set	21	16	76.2

successfully match the bound forms in the benchmark data set. When the standard target set was used together with the standard or modified template set, almost all predictions (except one obtained with the standard target and template sets) matched with the bound forms in the benchmark data set. When the target set was enlarged with alternative conformations, more than 76% of the favorable predictions matched the bound forms following both scenarios. Figure 7 presents the distribution of the predictions based on their energies and IS-scores. Predictions with relatively lower energies can have low or high IS-scores; predictions with relatively higher energies have high IS-scores. Predictions with favorable energies but low IS-scores (red triangles) might correspond to alternative binding modes of the proteins. The benchmark offers favorable

Figure 7. Distribution of predictions based on energy and IS-score. Red triangles show incorrect predictions (IS-score < 0.12). Green square represents acceptable prediction (0.12 < IS-score < 0.17). Blue diamonds show near native predictions (IS-score > 0.17). They are distributed in a triangle profile where predictions with higher energies do not have low IS-scores. We should note that the predictions with favorable energies but low IS-scores (red triangles) might correspond to alternative binding modes of the proteins.

interactions of protein pairs, yet other conformations of the proteins may also interact in a different way than the bound forms given in the benchmark data set, and we may capture such interactions. We checked if those interactions can be biological using machine learning tools, NOXclass,⁶⁹ DiMoVo⁷⁰ and EPPIC⁷¹ (Table S16, Supporting Information). Default parameters, interface area, interface area ratio and area-based amino acid composition, were used in NOXclass evaluations; cutoff value was chosen as 0.5 in the DiMoVo as indicated there; the EPPIC server directly indicates if the interface is biological or the outcome of crystal packing. NOXclass evaluated 13 interactions, DiMoVo evaluated 2 interactions, and EPPIC evaluated 5 out of 18 interactions as biological. Two interactions were evaluated as biological by all three tools and 13 interactions by at least one tool. Machine learning tools depend on their training data sets; however, evaluation of some interactions as biological indicates that these interactions might be biological, although structurally they are not similar to the bound forms given in the benchmark data set.

Case Study: ERK Interactions

We tested this method on the extracellular signal-regulated kinases (ERK) interactions. In the KEGG database,⁷² interactions among the proteins ERK, MEK1, MEK2, MP1, RSK2 and Mnk1/2 are given as in Figure 8. In this “classical”

Figure 8. MAPK signaling pathway given in the KEGG database. Nodes are the proteins and edges are the interactions. Gray lines indicate the membrane. The data is taken from the KEGG database. The focus is on ERK interactions with MEK1, MEK2, MP1, Mnk1/2 and RSK2.

A) ERK interacting proteins given in the KEGG database

is the whole structure). In the first case, 5 predictions were obtained, and 4 of them were energetically favorable when the cutoff energy value was -10 kJ/mol (Figure 9B, Table S18a, Supporting Information). One of these was between MEK1 and MEK2, shown in Figure 10A. However, in the second case, 7

Some of the predicted interactions have been shown experimentally. MEK1 activates ERK1 via phosphorylation. MEK1 residues 33–393, which cover all interacting residues in our model, interact with ERK.⁷⁵ Deletion of ERK residues 241–272, which in our model contribute 10 out of the 13 interacting residues, abolishes the interaction. However, this deletion does not affect the ERK–Mnk association, and in our model the interaction is not through this portion of the surface. MP1 is a MEK1 scaffolding protein that regulates cell spreading. In the MEK1–MP1 interaction, binding is through MEK1 residues 220–393, which in our model include all interacting residues (without contribution from MEK1 residues 1–219).⁷⁶ No MEK2–MP1 interaction was detected in the two-hybrid system in agreement with our predictions. Kinase suppressor of Ras (KSR) is a scaffolding protein for the Raf–MEK–ERK complex. KSR2 promotes phosphorylation of MEK1, and the crystal structure of this heterodimer was obtained (PDB ID: 2y4i).⁷⁷ Since the KEGG database does not

Figure 11. Trimer structure of ERK, MEK1 and MEK2. Trimer structure was obtained via superimposing MEK1 protein (3mb1A) of MEK1–ERK (3mb1A–2ojgA) and MEK1–MEK2 (3mb1A–1s9iB) interactions predicted by PRISM. Blue: MEK1 protein; green: ERK protein; and orange: MEK2 protein.

Figure 12. Interacting conformations of ERK interacting proteins. Interactions are energetically most favorable PRISM predictions obtained by using alternative conformations. Different conformations of the same protein are given by different structures in the same color.

include KSR2 in the MAPK pathway, we also did not include this protein in our predictions. However, PRISM could predict the KSR2–MEK1 interaction (data not shown). MEK1–MEK2 interaction determines the strength and duration of the ERK signal⁷⁸ and the ERK–RSK interaction is observed in response to various stimuli, like growth factors, polypeptide hormones, neurotransmitters, and chemokines.⁷⁹ MEK1 and MEK2 signaling stimulate ribosomal S6 kinase (RSK).⁸⁰

Using a structural protein–protein interaction network, we are able to observe which interactions can (or cannot) occur simultaneously. If two proteins interact through different sites of a third protein and there is no spatial clash, the two interactions can occur simultaneously, forming a trimer. For

example, a trimer of ERK–MEK1–MEK2 can form, as shown in Figure 11. The energetically most favorable predictions were 3mb1A–2ojgA (PDB codes) for MEK1–ERK and 3e8nA–1s9iA for MEK1–MEK2. These formed via two different conformations of MEK1, 3mb1A and 3e8nA. In the construction of the trimer, we considered one of these MEK1 conformations, 3mb1A. The most favorable interaction of 3mb1A with a MEK2 was 3mb1A–1s9iB (energy: -62.75 kJ/ml, template: 1s9iAB). Since ERK and MEK2 bind through different sites of MEK1 and their residues did not clash, a trimer could be obtained based on MEK1–ERK and MEK1–MEK2 interactions. The trimer structure was modeled by superimposing MEK1 (3mb1A) of MEK1–ERK (3mb1A–

2ojgA) and MEK1–MEK2 (3mblA–1s9iB). Figure 12 presents the most favorable interacting conformations of ERK, obtained by using alternative conformations. Different conformations of the same protein are given by different structures in the same color. Collectively, these show which conformation of a protein is more favorable to interact with another protein and indicate that the protein switches its conformation while switching its interacting partner, information that is valuable for figuring out pathway regulation.

CONCLUSIONS

Identification of PPIs is a first step in the elucidation of the interactome-scale structural network. Toward this aim, a tool that can help in the modeling of the structural pathway can be very useful. Here, we investigated whether different conformations of the query proteins found from the PDB could increase success in PPI identification by a motif, knowledge-based protein–protein interaction prediction tool. Knowledge-based methods can predict successfully an interaction between proteins and its three-dimensional structure, if a physically and biologically meaningful interaction exists in the database. They neither consider the conformational space nor model conformational changes. Flexibility is considered only in the refinement process, which is the last step in the prediction, where side chain and slight backbone reorientation are carried out. In principle, the capability of knowledge-based methods, like PRISM, to consider protein flexibility is limited by the supply of different conformations to the target set. Adding different conformations to the target set and searching interactions among these can be expected to improve the predictions. In this study, we tested such a strategy. Possible conformations of the query proteins were extracted from the PDB and treated as individual structures in the predictions. These structures were guaranteed to belong to the same protein by requiring high percentage (first 100% and then 95%) sequence homology. Different conformations were identified by structural similarity to eliminate redundancy.

Adding sufficiently different conformations of proteins to the target set may allow predicting the interaction of two proteins in any form. However, two factors limit the capability of template-based methods to predict correctly the native interaction: (i) the template set organization and (ii) availability of the “right” conformations of the query proteins. For the first, it is important that the template structures represent the interactions in their corresponding clusters; otherwise, appropriate structural similarity cannot be found between the target and template structures. Second, in particular, a major limitation of the prediction success is the coverage of the PDB. If the appropriate conformation exists in the PDB, it can be found and the interaction can be predicted successfully; in its absence the prediction will fail. The exponentially increasing numbers of structures in the PDB promises a high likelihood of finding the “right” conformations of the query proteins. More successful predictions will be obtained as the PDB covers more structures. Even with these limitations, knowledge-based approaches offer successful predictions. A large scale, systematic study⁴⁰ indicates that knowledge-based approaches are more reliable than blind docking strategies. Appropriate templates can be extracted to model nearly all complexes of structurally characterized proteins, although there are currently a limited number of protein–protein complexes in the PDB.

In the specific example here with PRISM, successful prediction was achieved for almost all the easy cases (87 out

of 88) of the docking benchmark data set,⁵⁹ and considering other conformations in the PDB, difficult cases of this data set were predicted with 66.7% success rate. A success rate 57.1% was obtained in unbiased predictions. Previous studies predicted interactions in signaling pathways (76 and 78% accuracies in the ubiquitination⁵⁵ and apoptosis⁵⁶ pathways, respectively) and constructed structural networks.⁵⁷ Here, we modeled the structural network of ERK, first on the basis of one conformation of each protein; then considering alternative conformations in the PDB, an improved structural network was obtained. Given the limited number of architectural motifs in single chain proteins and as we have shown earlier in protein–protein interfaces,^{43,81–83} and the difficulties facing large-scale applications of blind docking methods, knowledge-based approaches are a reasonable venue for proteome-scale cellular network constructions. Fast incorporation of experiment-based flexibility into these is advantageous. While vastly incomplete, and with limited sampling, it is a reasonable and reliable strategy to adopt.

ASSOCIATED CONTENT

Supporting Information

Table S1. Difficult cases of the docking benchmark data set. Table S2. Other conformations of given molecules in difficult cases of the benchmark data set, found using 100% sequence homology. Table S3. FiberDock energies of the bound forms given in the benchmark data set. Table S4. Alternative structures as the substitute of the bound form. Table S5. Changes in template organization. Table S6. Energy values of PRISM results obtained using standard template set. Table S7a. Favorable PRISM predictions of difficult cases using standard template set and standard target set. Table S7b. Favorable PRISM predictions of difficult cases using standard template set and enlarged target set. Table S8. Energies, IS-scores and contact fractions of PRISM predictions obtained using standard template set and standard target set. Table S9. Energy values of PRISM results obtained using rearranged templates and target sets. Table S10a. Favorable PRISM predictions of difficult cases using rearranged templates and standard target set. Table S10b. Favorable PRISM predictions of difficult cases using rearranged templates and enlarged target set. Table S11a. Energies, IS-scores and contact fractions of PRISM predictions obtained using rearranged templates and standard target set. Table S11b. Energies, IS-scores and contact fractions of PRISM predictions obtained using rearranged templates and enlarged target set. Table S12. Sequence similarity of bound and unbound forms given in the benchmark data set. Table S13. Additional conformations of the query molecules found when the sequence homology is lowered from 100 to 95%. Table S14. Favorable PRISM predictions of difficult cases obtained using additional conformations found by 95% sequence homology. Table S15. Energies, IS-scores and contact fractions of PRISM predictions obtained using additional conformations found by 95% sequence homology. Table S16. Biological/nonbiological interaction test results of predictions with favorable energies but lower IS-scores. Table S17. Different conformations of ERK interacting proteins. Table S18a. Energetically most favorable PRISM predictions of the query ERK interacting proteins. Table S18b. Energetically most favorable PRISM predictions obtained using alternative conformations of the ERK interacting proteins. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: agursoy@ku.edu.tr; okeskin@ku.edu.tr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Guay Kuzu is supported by a TUBITAK (The Scientific and Technological Research Council of Turkey) fellowship. This work has been supported by TUBITAK, Research Grant Number: 109T343. Ozlem Keskin acknowledges Science Academy (of Turkey). It has also been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

REFERENCES

- (1) Song, J. M.; Singh, M. How and when should interactome-derived clusters be used to predict functional modules and protein function? *Bioinformatics* **2009**, *25* (23), 3143–3150.
- (2) Glatter, T.; Wepf, A.; Aebersold, R.; Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol. Syst. Biol.* **2009**, *5*.
- (3) Breitkreutz, A.; Choi, H.; Sharom, J. R.; Boucher, L.; Neduva, V.; Larsen, B.; Lin, Z. Y.; Breitkreutz, B. J.; Stark, C.; Liu, G. M.; Ahn, J.; Dewar-Darch, D.; Reguly, T.; Tang, X. J.; Almeida, R.; Qin, Z. S.; Pawson, T.; Gingras, A. C.; Nesvizhskii, A. I.; Tyers, M. A global protein kinase and phosphatase interaction network in yeast. *Science* **2010**, *328* (5981), 1043–1046.
- (4) Baker, C. L.; Kettenbach, A. N.; Loros, J. J.; Gerber, S. A.; Dunlap, J. C. Quantitative proteomics reveals a dynamic interactome and phase-specific phosphorylation in the *Neurospora* circadian clock. *Mol. Cell* **2009**, *34* (3), 354–363.
- (5) Nibbe, R. K.; Markowitz, S.; Myeroff, L.; Ewing, R.; Chance, M. R. Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer. *Mol. Cell. Proteomics* **2009**, *8* (4), 827–845.
- (6) Kreeger, P. K.; Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **2010**, *31* (1), 2–8.
- (7) Bell, R.; Hubbard, A.; Chettier, R.; Chen, D.; Miller, J. P.; Kapahi, P.; Tarnopolsky, M.; Sahasrabudhe, S.; Melov, S.; Hughes, R. E. A human protein interaction network shows conservation of aging processes between human and invertebrate species. *PLoS Genet.* **2009**, *5*, 3.
- (8) Fossum, E.; Friedel, C. C.; Rajagopala, S. V.; Titz, B.; Baiker, A.; Schmidt, T.; Kraus, T.; Stellberger, T.; Rutenberg, C.; Suthram, S.; Bandyopadhyay, S.; Rose, D.; von Brunn, A.; Uhlmann, M.; Zeretzke, C.; Dong, Y. A.; Boulet, H.; Koegl, M.; Bailer, S. M.; Koszinowski, U.; Ideker, T.; Uetz, P.; Zimmer, R.; Haas, J. Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog.* **2009**, *5*, 9.
- (9) Wiles, A. M.; Doderer, M.; Ruan, J. H.; Gu, T. T.; Ravi, D.; Blackman, B.; Bishop, A. J. R. Building and analyzing protein interactome networks by cross-species comparisons. *BMC Syst. Biol.* **2010**, *4*.
- (10) Chautard, E.; Thierry-Mieg, N.; Ricard-Blum, S. Interaction networks: From protein functions to drug discovery. A review. *Pathol. Biol.* **2009**, *57* (4), 324–333.
- (11) Barabasi, A. L.; Gulbahce, N.; Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **2011**, *12* (1), 56–68.
- (12) West, G. M.; Tucker, C. L.; Xu, T.; Park, S. K.; Han, X. M.; Yates, J. R.; Fitzgerald, M. C. Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (20), 9078–9082.
- (13) Bruckner, A.; Polge, C.; Lentze, N.; Auerbach, D.; Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **2009**, *10* (6), 2763–2788.
- (14) Pande, J.; Szczyzyk, M. M.; Grover, A. K. Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* **2010**, *28* (6), 849–858.
- (15) Katz, C.; Levy-Beladev, L.; Rotem-Bamberger, S.; Rito, T.; Rudiger, S. G. D.; Friedler, A. Studying protein–protein interactions using peptide arrays. *Chem. Soc. Rev.* **2011**, *40* (5), 2131–2145.
- (16) Kim, E. D. H.; Sabharwal, A.; Vetta, A. R.; Blanchette, M. Predicting direct protein interactions from affinity purification mass spectrometry data. *Algorithms Mol. Biol.* **2010**, *5*.
- (17) Chen, P. Y.; Deane, C. M.; Reinert, G. Predicting and validating protein interactions using network structure. *PLoS Comput. Biol.* **2008**, *4*, 7.
- (18) Hart, G. T.; Ramani, A. K.; Marcotte, E. M. How complete are current yeast and human protein-interaction networks? *Genome Biol.* **2006**, *7*, 11.
- (19) von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Fields, S.; Bork, P. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **2002**, *417* (6887), 399–403.
- (20) Deane, C. M.; Salwinski, L.; Xenarios, I.; Eisenberg, D. Protein interactions—Two methods for assessment of the reliability of high throughput observations. *Mol. Cell. Proteomics* **2002**, *1* (5), 349–356.
- (21) Fourme, R.; Girard, E.; Kahn, R.; Prange, T.; Dhaussy, A. C.; Mezouar, M.; Ascone, I. High-resolution structures and properties of biomolecules under high pressures probed by X-ray crystallography. *High Pressure Res.* **2010**, *30* (1), 100–103.
- (22) O'Connell, M. R.; Gamsjaeger, R.; Mackay, J. P. The structural analysis of protein–protein interactions by NMR spectroscopy. *Proteomics* **2009**, *9* (23), 5224–5232.
- (23) Zhou, Z. H. Atomic resolution cryo electron microscopy of macromolecular complexes. In *Advances in Protein Chemistry and Structural Biology*; Ludtke, S. J., Prasad, B. V. V., Eds.; Elsevier: Amsterdam, 2011; Vol 82: Recent Advances in Electron Cryomicroscopy, Pt B, pp 1–35.
- (24) Lipfert, J.; Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307–327.
- (25) 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* (1), 235–242.
- (26) De Vries, S. J.; van Dijk, M.; Bonvin, A. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **2010**, *5* (5), 883–897.
- (27) Garzon, J. I.; Lopez-Blanco, J. R.; Pons, C.; Kovacs, J.; Abagyan, R.; Fernandez-Recio, J.; Chacon, P. FRODOCK: a new approach for fast rotational protein–protein docking. *Bioinformatics* **2009**, *25* (19), 2544–2551.
- (28) Chen, R.; Li, L.; Weng, Z. P. ZDOCK: An initial-stage protein-docking algorithm. *Proteins: Struct., Funct., Genet.* **2003**, *52* (1), 80–87.
- (29) Li, L.; Guo, D. C.; Huang, Y. Y.; Liu, S. Y.; Xiao, Y. ASPDOCK: protein–protein docking algorithm using atomic solvation parameters model. *BMC Bioinf.* **2011**, *12*.
- (30) Chen, H. L.; Skolnick, J. M-TASSER: An algorithm for protein quaternary structure prediction. *Biophys. J.* **2008**, *94* (3), 918–928.
- (31) Gao, M.; Skolnick, J. iAlign: a method for the structural comparison of protein–protein interfaces. *Bioinformatics* **2010**, *26* (18), 2259–2265.

- (32) Konc, J.; Janezic, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26* (9), 1160–1168.
- (33) Ogmen, U.; Keskin, O.; Aytuna, A. S.; Nussinov, R.; Gursoy, A. PRISM: protein interactions by structural matching. *Nucleic Acids Res.* **2005**, *33*, W331–W336.
- (34) Sinha, R.; Kundrotas, P. J.; Vakser, I. A. Docking by structural similarity at protein–protein interfaces. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (15), 3235–3241.
- (35) Tuncbag, N.; Gursoy, A.; Nussinov, R.; Keskin, O. Predicting protein–protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc.* **2011**, *6* (9), 1341–1354.
- (36) 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* (4), 604–620.
- (37) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Protein–protein interfaces: Architectures and interactions in protein–protein interfaces and in protein cores. Their similarities and differences. *Crit. Rev. Biochem. Mol. Biol.* **1996**, *31* (2), 127–152.
- (38) Tuncbag, N.; Gursoy, A.; Guney, E.; Nussinov, R.; Keskin, O. Architectures and functional coverage of protein–protein interfaces. *J. Mol. Biol.* **2008**, *381* (3), 785–802.
- (39) Keskin, O.; Nussinov, R. Similar binding sites and different partners: Implications to shared proteins in cellular pathways. *Structure* **2007**, *15* (3), 341–354.
- (40) Kundrotas, P. J.; Zhu, Z. W.; Janin, J.; Vakser, I. A. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (24), 9438–9441.
- (41) Bennett, W. S.; Huber, R. Structural and functional aspects of domain motions in proteins. *CRC Crit. Rev. Biochem.* **1984**, *15* (4), 291–384.
- (42) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins: Struct., Funct., Genet.* **2001**, *44* (2), 150–165.
- (43) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of protein–protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.* **2008**, *108* (4), 1225–1244.
- (44) Ma, B. Y.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Sci.* **2002**, *11* (2), 184–197.
- (45) Dill, K. A. Polymer principles and protein folding. *Protein Sci.* **1999**, *8* (6), 1166–1180.
- (46) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254* (5038), 1598–1603.
- (47) Ozbabacan, S. E. A.; Gursoy, A.; Keskin, O.; Nussinov, R. Conformational ensembles, signal transduction and residue hot spots: Application to drug discovery. *Curr. Opin. Drug Discovery Dev.* **2010**, *13* (5), 527–537.
- (48) Tsai, C. J.; Del Sol, A.; Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. BioSyst.* **2009**, *5* (3), 207–216.
- (49) Ma, B. Y.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **1999**, *12* (9), 713–720.
- (50) Tsai, C. J.; Kumar, S.; Ma, B. Y.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8* (6), 1181–1190.
- (51) Tsai, C. J.; Ma, B. Y.; Nussinov, R. Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 9970–9972.
- (52) Bakan, A.; Bahar, I. Computational generation inhibitor-bound conformers of p38 map kinase and comparison with experiments. *Pac. Symp. Biocomput.* **2011**, 181–192.
- (53) Aytuna, A. S.; Gursoy, A.; Keskin, O. Prediction of protein–protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics* **2005**, *21* (12), 2850–2855.
- (54) Hwang, H.; Vreven, T.; Janin, J.; Weng, Z. P. Protein–protein docking benchmark version 4.0. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (15), 3111–3114.
- (55) Kar, G.; Keskin, O.; Nussinov, R.; Gursoy, A. Human proteome-scale structural modeling of E2–E3 interactions exploiting interface motifs. *J. Proteome Res.* **2012**, *11* (2), 1196–1207.
- (56) Acuner Ozbabacan, S. E.; Keskin, O.; Nussinov, R.; Gursoy, A. Enriching the human apoptosis pathway by predicting the structures of protein–protein complexes. *J. Struct. Biol.* **2012**, *179* (3), 338–346.
- (57) Kuzu, G.; Keskin, O.; Gursoy, A.; Nussinov, R. Constructing structural networks of signaling pathways on the proteome scale. *Curr. Opin. Struct. Biol.* **2012**, *22* (3), 367–377.
- (58) Kar, G.; Gursoy, A.; Keskin, O. Human cancer protein–protein interaction network: a structural perspective. *PLoS Comput. Biol.* **2009**, *5*, 12.
- (59) Tuncbag, N.; Keskin, O.; Nussinov, R.; Gursoy, A. Fast and accurate modeling of protein–protein interactions by combining template-interface-based docking with flexible refinement. *Proteins* **2011**, *80* (4), 1239–1249.
- (60) Gao, M.; Skolnick, J. New benchmark metrics for protein–protein docking methods. *Proteins: Struct., Funct., Bioinf.* **2011**, *79* (5), 1623–1634.
- (61) Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (8), 2444–2448.
- (62) Hubbard, S. J.; Thornton, J. M. NACCESS; University College London: London, 1993.
- (63) Shatsky, M.; Nussinov, R.; Wolfson, H. J. A method for simultaneous alignment of multiple protein structures. *Proteins: Struct., Funct., Bioinf.* **2004**, *56* (1), 143–156.
- (64) Mashiah, E.; Nussinov, R.; Wolfson, H. J. FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. *Nucleic Acids Res.* **2010**, *38*, W457–W461.
- (65) Tuncbag, N.; Keskin, O.; Gursoy, A. HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Res.* **2010**, *38*, W402–W406.
- (66) Baglin, T. P.; Carrell, R. W.; Church, F. C.; Esmon, C. T.; Huntington, J. A. Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (17), 11079–11084.
- (67) Tuncbag, N.; Kar, G.; Gursoy, A.; Keskin, O.; Nussinov, R. Towards inferring time dimensionality in protein–protein interaction networks by integrating structures: the p53 example. *Mol. BioSyst.* **2009**, *5* (12), 1770–1778.
- (68) Lensink, M. F.; Mendez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins: Struct., Funct., Bioinf.* **2007**, *69* (4), 704–718.
- (69) Zhu, H. B.; Domingues, F. S.; Sommer, I.; Lengauer, T. NOXclass: prediction of protein–protein interaction types. *BMC Bioinf.* **2006**, *7*.
- (70) Bernauer, J.; Bahadur, R. P.; Rodier, F.; Janin, J.; Poupon, A. DiMoVo: a Voronoi tessellation-based method for discriminating crystallographic and biological protein–protein interactions. *Bioinformatics* **2008**, *24* (5), 652–658.
- (71) Duarte, J. M.; Srebnik, A.; Scharer, M. A.; Capitani, G. Protein interface classification by evolutionary analysis. *BMC Bioinf.* **2012**, *13*, 334.
- (72) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30.
- (73) Tyagi, M.; Hashimoto, K.; Shoemaker, B. A.; Wuchty, S.; Panchenko, A. R. Large-scale mapping of human protein interactome using structural complexes. *EMBO Rep.* **2012**, *13* (3), 266–271.
- (74) Gursoy, A.; Keskin, O.; Nussinov, R. Topological properties of protein interaction networks from a structural perspective. *Biochem. Soc. Trans.* **2008**, *36*, 1398–1403.
- (75) Robinson, F. L.; Whitehurst, A. W.; Raman, M.; Cobb, M. H. Identification of novel point mutations in ERK2 that selectively disrupt binding to MEK1. *J. Biol. Chem.* **2002**, *277* (17), 14844–14852.

- (76) Schaeffer, H. J.; Catling, A. D.; Eblen, S. T.; Collier, L. S.; Krauss, A.; Weber, M. J. MP1: A MEK binding partner that enhances enzymatic activation of the MAP kinase cascade. *Science* **1998**, *281* (5383), 1668–1671.
- (77) Brennan, D. F.; Dar, A. C.; Hertz, N. T.; Chao, W. C. H.; Burlingame, A. L.; Shokat, K. M.; Barford, D. A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. *Nature* **2011**, *472* (7343), 366–364.
- (78) Catalanotti, F.; Reyes, G.; Jesenberger, V.; Galabova-Kovacs, G.; Simoes, R. D.; Carugo, O.; Baccarini, M. A Mek1-Mek2 heterodimer determines the strength and duration of the Erk signal. *Nat. Struct. Mol. Biol.* **2009**, *16* (3), 294–303.
- (79) Hauge, C.; Frodin, M. RSK and MSK in MAP kinase signalling. *J. Cell Sci.* **2006**, *119* (15), 3021–3023.
- (80) Carriere, A.; Cargnello, M.; Julien, L. A.; Gao, H.; Bonneil, E.; Thibault, P.; Roux, P. P. Oncogenic MAPK signaling stimulates mTORC1 activity by promoting RSK-mediated Raptor phosphorylation. *Curr. Biol.* **2008**, *18* (17), 1269–1277.
- (81) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Protein–protein interfaces: architectures and interactions in protein–protein interfaces and in protein cores. Their similarities and differences. *Crit. Rev. Biochem. Mol. Biol.* **1996**, *31* (2), 127–152.
- (82) Keskin, O.; Nussinov, R. Favorable scaffolds: proteins with different sequence, structure and function may associate in similar ways. *Protein Eng., Des. Sel.* **2005**, *18* (1), 11–24.
- (83) Keskin, O.; Ma, B. Y.; Rogale, K.; Gunasekaran, K.; Nussinov, R. Protein–protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. *Phys. Biol.* **2005**, *2* (2), S24–S35.