RESEARCH ARTICLE

DynaFace: Discrimination between Obligatory and Non-obligatory Protein-Protein Interactions Based on the Complex's Dynamics

Seren Soner[1], Pemra Ozbek[2], Jose Ignacio Garzon[3], Nir Ben-Tal[4], Turkan Haliloglu[5]*

1 Department of Computer Engineering and Polymer Research Center, Bogazici University, Istanbul, Turkey, 2 Department of Bioengineering, Marmara University, Istanbul, Turkey, 3 Departments of Biochemistry and Molecular Biophysics and Systems Biology and Howard Hughes Medical Institute, Columbia University, New York, New York, United States of America, 4 Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, 5 Department of Chemical Engineering and Polymer Research Center, Bogazici University, Istanbul, Turkey

* halilogt@boun.edu.tr

Abstract

Protein-protein interfaces have been evolutionarily-designed to enable transduction between the interacting proteins. Thus, we hypothesize that analysis of the dynamics of the complex can reveal details about the nature of the interaction, and in particular whether it is obligatory, i.e., persists throughout the entire lifetime of the proteins, or not. Indeed, normal mode analysis, using the Gaussian network model, shows that for the most part obligatory and non-obligatory complexes differ in their decomposition into dynamic domains, i.e., the mobile elements of the protein complex. The dynamic domains of obligatory complexes often mix segments from the interacting chains, and the hinges between them do not overlap with the interface between the chains. In contrast, in non-obligatory complexes the interface often hinges between dynamic domains, held together through few anchor residues on one side of the interface that interact with their counterpart grooves in the other end. In automatic analysis, 117 of 139 obligatory (84.2%) and 203 of 246 non-obligatory (82.5%) complexes are correctly classified by our method: DynaFace. We further use DynaFace to predict obligatory and non-obligatory interactions among a set of 300 putative protein complexes. DynaFace is available at: http://safir.prc.boun.edu.tr/dynaface.

Author Summary

Protein-protein interactions mediate, in essence, all inter- and intra-cellular processes. Thus, understanding their molecular mechanism is of utmost importance. Here we focus on one mechanistic aspect: differentiation between obligatory interactions, which persist throughout the entire lifetime of the protein complex, and non-obligatory, which do not. For proper function, a protein complex should facilitate transduction between the

interacting proteins. Therefore the complex's dynamics should reveal whether it is obligatory or non-obligatory. Indeed, normal mode analysis shows that the dynamic domains of obligatory complexes often mix segments from the interacting chains. In contrast, in non-obligatory complexes the inter-chain interface often hinges between dynamic domains, held together through few anchor residues. An automated methodology based on these observations correctly classifies over 80% of the interfaces in a test set. We use it also to predict obligatory and non-obligatory interactions among putative protein complexes. DynaFace, a web-server implementation of the methodology, is available at: http://safir.prc.boun.edu.tr/dynaface.

"This is a *PLOS Computational Biology* Methods paper"

Introduction

Inter-protein interactions mediate a wide range of cellular and biochemical processes [1, 2]. The Protein Data Bank (PDB; [3]) includes a wealth of information on these important interactions, observed in X-ray crystal, NMR, and cryo-electron microscopy structures. Indeed, this large resource has been exploited to deduce interactions [4, 5], and infer interactions based on similarity in sequence and/or structure [6–8]. It is noteworthy, however, that the PDB includes various types of interactions, many of which are physiologically irrelevant and reflect crystal packing [9–11]. Here we focus on the rest, i.e., physiologically realistic interactions, and differentiate between these that are obligatory and non-obligatory.

Protein chains engaged in an obligatory complex are found only in association with their partner chains and bind throughout their functional lifetime, for example, because they are unstable on their own [12]. The most popular example here is the interaction between the beta and gamma subunits of G-proteins, which remain intact throughout their lifetime. In contrast, non-obligatory complexes, such as the interaction between the beta-gamma complex and the alpha subunit of the G-protein, form and dissociate in response to environmental changes. These proteins are stable both in their bound and unbound states, although their conformation may change upon binding. They are abundant, for example, in signal transduction, antibody-antigen interactions, and enzyme-inhibitor complexes. The obligatory-vs.-not classification is not always straightforward. For example, at least theoretically, one chain might be unstable alone and gain stability only upon association with another chain, which in turn is stable also alone. Furthermore, a chain with two domains might be engaged in obligatory interaction through the first domain and a non-obligatory interaction through the second, which again complicates the classification [13]. Additionally, the nature of the interaction may alter in response to external changes, such as pH, temperature, interaction with ligand, etc.

Non-obligatory interactions could be further classified as transient or permanent, depending on their lifetimes, providing a kinetic dimension of the association. In addition, changes in pH, ionic strength or concentrations of the interacting chains can shift the dynamic equilibrium towards or away from association. Thus, transient complexes can be further classified as "weak" or "strong", which are generally found in their unbound and bound states, respectively [12].

Obligatory interfaces compared to non-obligatory interfaces are larger, flatter and more evolutionarily conserved [14, 15]. Residues involved in obligatory interactions evolve at a slower rate, while residues in non-obligatory interactions exhibit an increased rate of

substitution for faster adaptation required. Obligatory interfaces have higher shape complementary consisting primarily of side-chain contacts and larger interface-to-surface ratio [14]. On the other hand, non-obligatory interfaces are usually smaller and more polar (except enzyme-inhibitor complexes) with lower geometric complementary and weaker association where backbone plays an important role [16]. The latter features should provide the interfaces with optimum topological means for the required functional motion and interaction.

Previous studies have focused mostly in the analysis and prediction of protein surfaces that mediate protein-protein interactions rather than in differentiation between the interaction types. Evolutionary conservation profiles of the amino acid positions have appeared as a valuable source of information for the success of sequence-based methods in binding site predictions [17–20]. Structure-based methods, on the other hand, could make use of some additional properties, such as solvent accessible surface area and shape complementarity methods for more successful predictions [21–23]. Combined with structural information, the patches of conserved amino acids on protein surfaces were shown to have functional importance [24–26]. Recently developed machine learning methods using several attributes in the latter provide algorithms with plausible performances [22, 27, 28].

The prediction of protein interaction sites and/or interaction type based on a single property is challenging [22]. The existing web-servers use various sequence and structure properties alone and combined. For example, web-servers such as Promate [29], PPI-Pred [30, 31], Con-PPISP [32], meta-PPISP [30], PRISM [33], SPPIDER [34], IBIS [35] and metaPIS [36] focus on binding site predictions. Promate [29] uses surface properties with various physicochemical properties to predict transient interactions. PPI-Pred [31], Cons-PPISP [32], PRISM [33, 37] and SPPIDER [34] predict interfaces that may include both obligatory and non-obligatory interactions. PPI-Pred uses a support vector machine method in combination with surface patch analysis, and Cons-PPISP is a structure-based neural network method mainly using sequence profiles and solvent accessibilities. PRISM predicts interfaces by structural matching. SPPIDER is mainly based on solvent accessibility. IBIS uses conservation of sequence and structure for binding site predictions and metaPIS is mainly based on protein sequences. PrePPI combines structural modeling with other genomic, evolutionary and functional clues for the prediction of binary interactions [6]. For a query protein pair, representative structures of the subunits are first searched in the PDB and in homology model databases and then a search for structural neighbors follows. If two neighbors of each subunit are found in a complex in the PDB, then this complex is used as template. As a scoring value, individual subunits are superposed on the template complex and a likelihood ratio is calculated in combination with the non-structural naïve Bayesian classifier.

The NOXclass web-server [38] is unique in providing automatic classification of the interaction/interface types of query protein complexes. NOXclass is a support vector machine classifier making use of properties such as interface area and interface/surface area ratio, amino acid composition, shape complementarity and residue conservation for the interaction types. Alternatively, SCOPPI (Structural Classification of Protein-Protein Interfaces) [39] is a database that classifies the interface type by using knowledge on protein domain-domain interactions with known structures. The domain interactions are determined by a distance-based criterion and the domain definitions are obtained from SCOP [40], where proteins are classified based on both structural and evolutionary relatedness. The BindML+ web-server [41] predicts the interface type in a single protein as transient or permanent without the knowledge of its interacting partner. Here the definitions of non-obligatory but permanent might have been confused with obligatory interactions. PiType [42] is a downloadable program classifying protein interactions into simultaneously possible and mutually exclusive as well as into obligate and non-obligate based on the sequence and functional properties of the binding partners and their network context based on amino acid sequence and functional similarity.

To facilitate biological functionality, protein-protein association should involve transduction, e.g., of signal. Thus, it is anticipated that protein complexes that differ in their functionalities would manifest different dynamic behavior. In other words, the dynamic infrastructure underlying non-obligatory and obligatory interactions should be in compliance with the required functional motion of the respective interface types. Indeed, a recent analysis of proteins with multiple conformations in PDB showed that, on average, transient associations involve smaller conformational changes than permanent associations [43]. Here again, non-obligatory but permanent interfaces might be confused with obligatory interfaces. We attempt to take a step forward and try to classify obligatory vs. non-obligatory interfaces based on the analysis of dynamic fluctuations starting from a single conformation and link this with the functionality at the level of interaction types on a dataset of obligatory and non-obligatory interfaces. The dynamic fluctuations are calculated using the Gaussian Network Model (GNM) [44, 45], and used to differentiate between obligatory and non-obligatory interactions. We further analyze a set of structural models predicted using PrePPI [6].

Methods

Datasets

The dataset used here is a compilation of two available datasets of protein-protein interface types [14, 31]. Both datasets were compiled from the PDB [46] and their interfaces were then manually curated with the existing literature. The PISCES server [47] was used to reduce redundancy by removing proteins with over 25% sequence identity, resolution of 3.0 Å or better, and R factor of 0.3 or better. While combining the two datasets, some proteins were removed due to the redundancy and high number of missing residues or changes with the updates. After preliminary application of DynaFace on the set we noted cases of disagreement between prediction and annotation. Literature survey showed that the annotation of some of these was erroneous, presumably because of studies published after the compilation of the original sets [62–86]. These were fixed. The final dataset, consisting of 139 obligatory and 246 non-obligatory complex structures, is provided in S1 Table. The corrections are marked. It is noteworthy that the set includes 84 multi-subunit complexes. For these, at least one of the interacting units includes more than one polypeptide chain.

Additionally, a dataset of predicted structural models [6] is used to make testable hypotheses. This dataset includes 85 template structures and three subsets of 100 predicted structural models based on these template structures for which the interactions are ranked as high, low and very low quality referring to the structures having the highest score, 50% probability, and 25% probability of existence, respectively. First, the consistency of the predictions has been tested between the template and structural models, and then the interaction type based on the dynamics has been assigned. The dataset is given in S2 Table.

Gaussian Network Model

GNM [44, 45] is the simplest elastic network model at residue level, where residue pairs with their alpha carbons located within a cut-off radius (r_c), are assumed to be connected by harmonic springs. The potential function for a protein structure of N residues in the elastic network description is given as

$$V_{GNM} = \frac{\gamma}{2} \left[\sum_{i,j}^{N} \Gamma_{ij} (\Delta \mathbf{R}_i - \Delta \mathbf{R}_j)^2 \right] \tag{1}$$

where γ is the spring constant and $\Delta \mathbf{R}$ refers to the fluctuation $\mathbf{R}$ vector of each residue at its

alpha carbon position. $\mathbf{\Gamma}$ is the Kirchhoff matrix defined as

$$\mathbf{\Gamma} = \begin{cases} -1 & if\ i \neq j\ and\ R_{ij} \leq r_c \\ 0 & if\ i \neq j\ and\ R_{ij} > r_c \\ -\sum_{i,j \neq i} \mathbf{\Gamma}_{ij} & if\ i = j \end{cases} \qquad (2)$$

R_{ij} is the distance between alpha carbon atoms i and j. r_c is taken as 7 Å. The diagonal Lambda_ii term is the degree of a node and a measure of the local packing density around a given residue.

The correlation between equilibrium position fluctuations, $\mathbf{\Delta R_i}$ and $\mathbf{\Delta R_j}$, of residues i and j forms the covariance matrix given as

$$\langle \mathbf{\Delta R_i} \cdot \mathbf{\Delta R_j} \rangle = \left(\frac{3k_B T}{\gamma}\right)[\mathbf{\Gamma}^{-1}]_{ij} = \left(\frac{3k_B T}{\gamma}\right)[\mathbf{U \Lambda}^{-1}\mathbf{U}^T]_{ij} = \left(\frac{3k_B T}{\gamma}\right)\sum_k [\lambda_k^{-1} \mathbf{u_k u_k^T}]_{ij} \qquad (3)$$

where $\mathbf{U}$ is an orthogonal matrix whose columns $\mathbf{u_i}$ are the eigenvectors of the Kirchhoff matrix and $\mathbf{\Lambda}$ is a diagonal matrix whose elements λ_i represent the eigenvalues, k_B is the Boltzmann constant, and T is the absolute temperature in Eq 3. The slow modes with lower eigenvalues contribute to global cooperative motions, while the fast modes with higher eigenvalues describe local fluctuations. The normalized values of the correlation between residue fluctuations range between +1 and -1.

Motion and interaction.   The covariance matrix (Eq 2) of a mode of motion divides the structure into dynamic domains: clusters of amino acids that are in close contact with each other and that move collectively in one direction [48]. The motion could be along the eigenvector, marked by correlation of +1, or in opposite direction, with correlation of –1. The sign changes mark the positions of the hinges at the interface between the dynamic domains. When two or more modes are superimposed, the fluctuations along the two eigenvectors are superimposed and the correlations can vary in the range +1 to -1, reflecting the average behavior. For example, the average of the two slowest modes amalgamates the two most global conformational changes accessible to a given complex structure. Incorporation of higher modes integrates other global, and subsequently also local, motions.

The conformational transitions of proteins are intimately related to their function. With the premise of the link between the dynamic infrastructure provided by the interacting chains and the interface type, we propose a dynamic measure based on GNM through the analysis of obligatory versus non-obligatory complex structures in the dataset: The dynamics are often dominated by the two slowest modes, yet further refined by the next slow modes; dynamic domains capture the global connectivity similarity for obligatory versus non-obligatory complex structures.

Server

Given a protein complex structure, the DynaFace web-server builds the covariance matrix and uses it to classify the inter-protein interfaces as obligatory vs. non-obligatory. Various combinations of different sets of modes are used in order to best exploit the dynamic correlation patterns. For the most part, the two slowest modes capture the pattern of the dynamic domains with respect to the interaction type. The ten slowest modes contribution is significant, but higher modes lead to fine-tuning. As a result, DynaFace uses the ten slowest modes (+ all modes) to increase the prediction accuracy. The addition of the all-modes terms increase the performance in comparison to using only the slowest 10 modes, as well as in comparison to using the 10 slowest plus modes 11 and higher.

The underlying scoring function of DynaFace basically reflects the pattern of the cross-correlation map. The following seven attributes are calculated: The average of all (A), negative (N) and positive (P) correlations between residue fluctuations of two interacting subunits in the average ten slowest (s) and all (a) modes of motion, and the number of associating regions between two subunits in all modes (AR), defined in the following paragraph. For each attribute, a knowledge-based decision parameter (threshold) is set based on the corresponding correlation values of 139 obligatory and 246 non-obligatory complex structures in the whole dataset; see S3 Table for the individual performance of each attribute. The *subunit* refers to the interacting partner unit; a single chain or more than one chain. A decision outcome D is calculated with respect to an obligatory interaction as follows

$$D = (A_a > \hat{A}_a) + (N_a > \hat{N}_a) + (P_a < \hat{P}_a) + (A_s > \hat{A}_s) + (N_s > \hat{N}_s) + (P_s < \hat{P}_s) + (AR_a > \hat{AR}_a) \tag{4}$$

Where, the hat over each attribute refers to a pre-determined threshold value. If the majority, i.e., four or more of the seven binary attributes satisfy the criteria for the obligatory interaction ($D \geq 4$), the decision is *obligatory*; otherwise it is *non-obligatory*. The threshold value for each attribute is given in S4 Table. Note that lower P with higher A and N values in an obligatory interfaces is the result of the manifestation of larger number of weak to strong positively correlated and weak negatively correlated residue pairs A large value of AR means a large number of associating regions between two subunits, which is an indicator of obligatory interface.

Associating regions (AR, the last attribute in Eq 4) refer to the interface residues that transduce the motion between the two subunits. Residues that correlate dynamically more strongly to the juxtaposed subunit than to their own subunit are defined as anchors. AR is the ratio of the total number of anchor residues from both chains to the sum of amino acids in the whole complex. Anchor residues are more abundant in obligatory compared to non-obligatory complexes.

Two different sets of thresholds are used depending on complex size: One set when both subunits are >65 residues, and another when the complex includes a subunit of less than 65 residues. The threshold values were also tested over sub-datasets produced by randomly dividing the original dataset into five groups.

In test trials of DynaFace using only the two and only the ten slowest modes, three attributes corresponding to A, N and P were calculated. If two out of three decision variables satisfy the criteria for obligatory interaction in these cases, the decision outcome D is considered *obligatory* otherwise *non-obligatory*. The resulting predictions are presented in *Results and Discussion*.

A more formal description of the calculations is provided in *Supplementary Data*. The flow chart of the algorithm is given in Fig 1. The confidence levels for obligatory and non-obligatory complex structures in the dataset are given in S1 Fig.

The DynaFace input can be a PDB ID or an uploaded coordinate file in standard PDB format. The output includes: predicted interface type, dynamic structural domains, cross-correlation maps and their projections on the structure's ribbon diagram. In the case of non-obligatory interactions, anchors are listed with anchoring groove residues, which are the residues in the juxtaposed subunit that display the highest correlations (top three) in all modes with the anchor residue across the interface.

Results and Discussion

The PDB includes many interfaces, some obligatory, some non-obligatory, and some artifacts due to crystal-packing. There are methods to filter out crystal-packing interfaces, which work
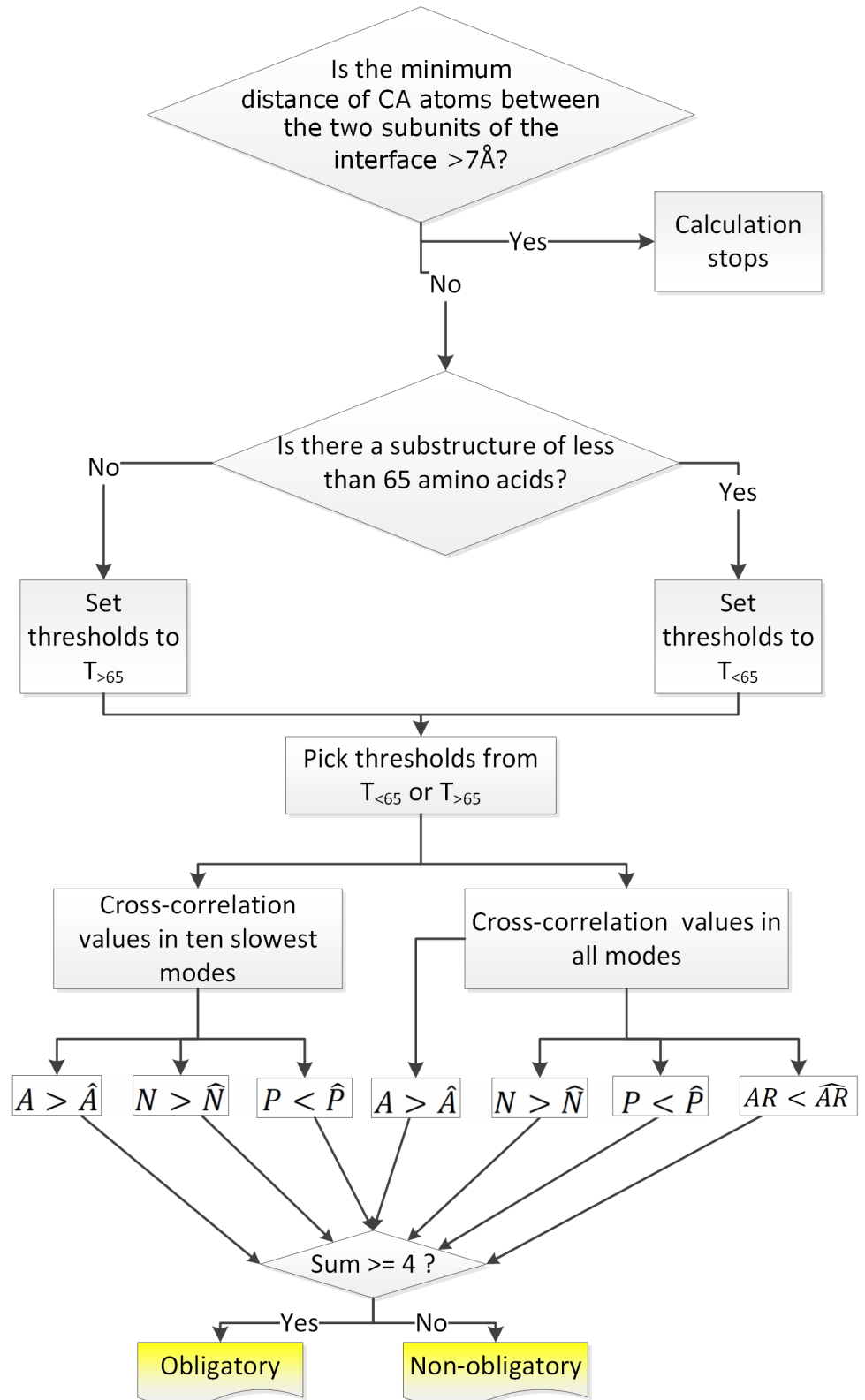
Fig 1. Flowchart of the DynaFace web server. A is the average of all cross-correlation values between different subunits, N is the average of negative cross-correlation values between different subunits, P is the

average of positive cross-correlation values between different subunits, AR is the ratio of associating regions in the protein, and superscript ^ stands for the pre-determined threshold for that parameter.

at least to some extent. We offer a complementary method that can differentiate between obligatory and non-obligatory interfaces based on the dynamics of the structural complex. For completeness, and because some crystal-packing artifacts might 'sneak in', we also check whether they confuse our method. The results show that DynaFace might mistakenly assign crystal-packing artifacts to be non-obligatory interfaces. We also use our methodology on a dataset of predicted complex structures as well as their corresponding template structures [6] to suggest the interaction type for each.

## Motion and interaction: Global dynamics

Equilibrium residue fluctuations define dynamic domains, where the cooperative modes of motion are the main determinants. The collective dynamics facilitates biological function(s), and correlated motions within the structure could reveal functional predispositions. Thus, the expression of the dynamic couplings across the inter-subunit interface could reveal whether the complex is obligatory or non-obligatory. To examine this we start with two exemplary dimeric complexes. The first is the obligatory homodimer 1QU7 (Cytoplasmic domain of a serine chemotaxis receptor) [49], and the second is the non-obligatory heterodimer 2SIC (SubtilisinBPN' in complex with Streptomyces subtilisin inhibitor) [50].

Cross-correlations of residue fluctuations in the two slowest modes for these dimers are presented in Fig 2 and Fig 3, respectively. Reassuringly, the positioning of the dynamic domains relative to the inter-subunits interfaces differs between the obligatory (Fig 2) and non-obligatory (Fig 3) interactions. The obligatory complex features mostly dynamic domains that share segments between the two subunits, as in the two slowest modes (Fig 2); indeed, the complex's dynamics is dominated by such modes (S2A Fig, S2B Fig and S2C Fig). The dynamic domains of some of the modes of motion of the non-obligatory complex also share segments across the subunits interface (e.g., the second slowest mode, Fig 3). However, the complex's dynamics is dominated by modes with correlations only within the individual subunits, such as the slowest mode (Fig 3C, S2D Fig, S2E Fig and S2F Fig). Interestingly, the slowest mode of the non-obligatory complex also manifests 'anchor' behavior, where an anchor in the inhibitor (residues MET70-VAL74) is dynamically correlated with the enzyme. The anchor MET70-VAL74 correlates with the majority of the enzyme residues in the average over the two slowest modes (S2D Fig). A subset of residues (GLY100, SER125, THR220) correlate with the anchor also in the average over all modes. These are referred to as 'anchoring groove' residues (S2F Fig). This behavior is typical for non-obligatory complexes; the anchors are lesser in number compared to obligatory complexes but provide a means for transduction between the intact chains. This has led us to add the last term in the formula used to discriminate between the interface types (Eq 4). More detailed description of the dynamics of the two complexes is provided in the following two sections.

Obligatory interaction: The cytoplasmic domain of a serine chemotaxis receptor (1QU7). The cytoplasmic domain of the serine chemotaxis receptor of Escherichia coli, is a homodimeric complex with two protein chains of 227 amino acids, marked as A and B [49]. Each subunit is an elongated helical hairpin and the subunits intertwine to form a four helix bundle. In contrast to many other receptors, this receptor is dimeric regardless of ligand binding, and its signaling is independent of the monomer-dimer equilibrium [51].
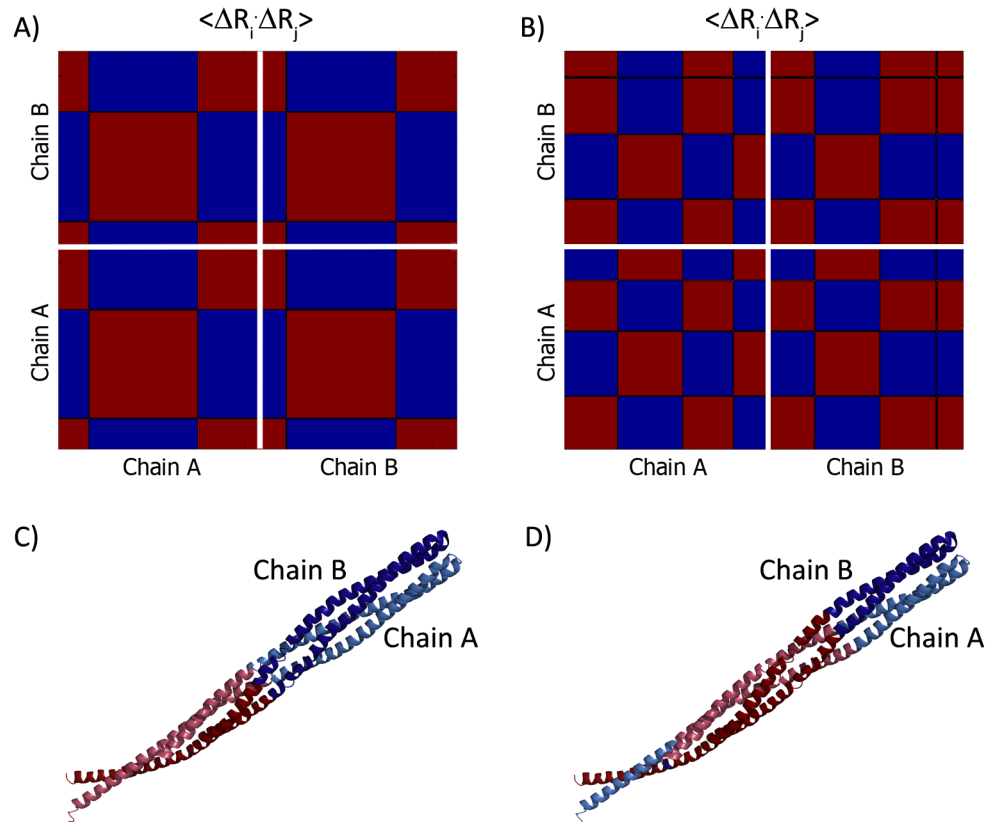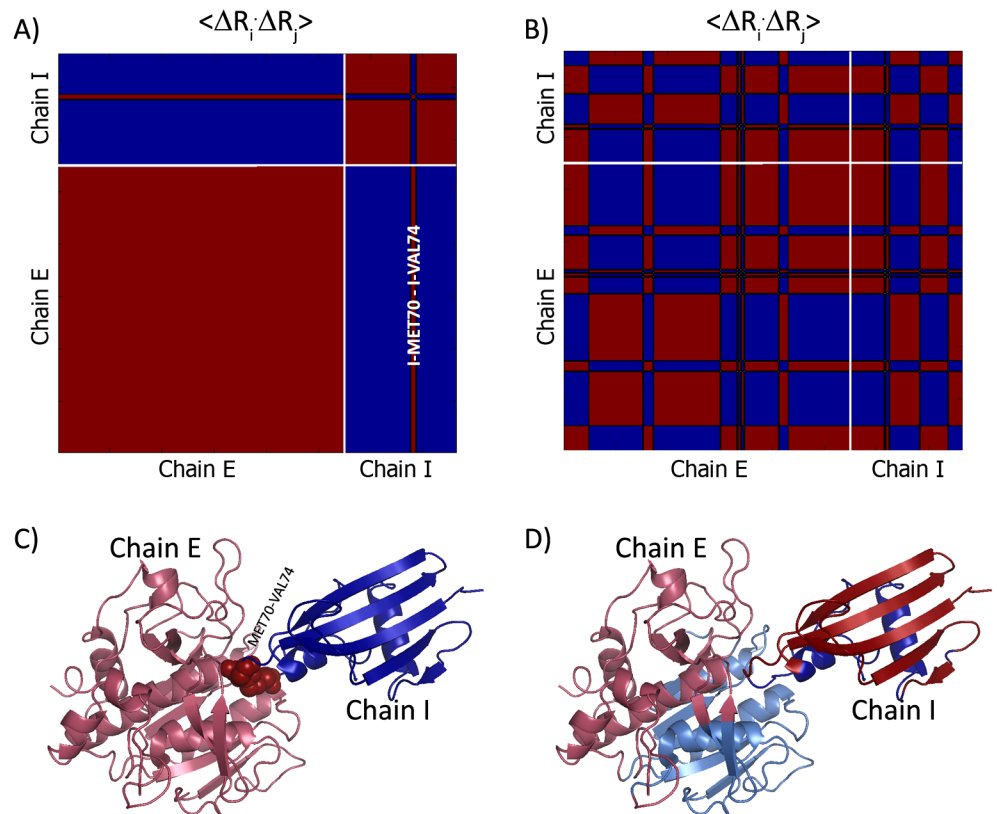
Fig 2. Dynamics of an obligatory complex. Slow modes in the cytoplasmic domain of the homodimeric serine chemotaxis receptor (1QU7 [49]). The upper panels show the matrices of correlations between residue fluctuations, $\langle \Delta R_i \cdot \Delta R_j \rangle$, of mode 1 (A) and mode 2 (B), with negative and positive marked as blue and red, respectively. The boundaries between the subunits are marked in white. The lower panels show projection of the correlations on the 3D-structure: C- mode 1 and D- mode 2. The subunits are shown on the PyMOL [61] figures in lighter and darker versions of the same colors.

doi:10.1371/journal.pcbi.1004461.g002

According to the slowest mode, the dimeric structure has four hinges, two in each subunit; A: HIS328/LEU329, A: THR450/ARG451, B: HIS328/LEU329, and B: ARG451/VAL452 ([Fig 2A](), [S5 Table]()). The hinges align along a plane that passes through the geometrical center of the complex. The hinge plan separates the dimer into two dynamic domains, each of which includes segments from the two subunits ([S5 Table]()). The second slowest mode describes another type of motion where the underlying dynamic domains ([S5 Table]()) are coordinated by hinge residues ASP353/ILE354, GLY426/LYS427, VAL483/THR484 in chain A, and ASP353/ILE354, GLY426/LYS427 in chain B. These dynamic domains also mix segments from both subunits. The slowest mode dominates the dynamic fluctuations of the dimer. It dominates the average over the two slowest modes ([S2A Fig]()), as well as the averages over the ten slowest ([S2B Fig]()) and all ([S2C Fig]()) modes, which are used to discriminate between interface types. [S4 Table]() shows the threshold values and the attributes of each decision variable for this case as given in [Eq 4](). The complex satisfies four of seven criteria, and is classified as obligatory, as it should.

Non-obligatory interaction: Subtilisin BPN' in complex with streptomyces subtilisin inhibitor (2SIC). The complex between subtilisin BPN' (chain E, residues 1–275) with its streptomyces inhibitor (chain I, residues 7–113) is an example for non-obligatory interaction (PDB ID: 2SIC; [52]). The slowest mode of motion divides the structure into two dynamic domains with a hinge plane at the enzyme-inhibitor interface ([Fig 3A and 3C](), [S6 Table]()). It is

Fig 3. Dynamics of a non-obligatory complex. Slow modes of motion in the complex between subtilisin BPN' and proteinaceous inhibitor from Streptomyces (2SIC [52]). The upper panels show the matrices of correlations between residue fluctuations, $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle$, of mode 1 (A) and mode 2 (B) with negative and positive marked as blue and red, respectively. The boundaries between the subunits are marked in white. The lower panels show projection of the correlations on the 3D-structure: C- mode 1 and D- mode 2. The subunits are shown on the PyMOL [61] figures in lighter and darker versions of the same colors.

doi:10.1371/journal.pcbi.1004461.g003

noteworthy that here each subunit is, in essence, an autonomous dynamic domain, which is often the case for non-obligatory interactions. The slowest mode also predisposes an anchor (MET70-VAL74) for an anchor-and-groove behavior, which is a hallmark for non-obligatory interactions.

The second slowest mode with hinge sites at SER24/ASN25, VAL150/ALA151 of the enzyme (chain E) and HIS43/PRO44, PRO72/MET73, ASN99/GLU100 of the inhibitor (chain I) also separates the structure into two dynamic domains (Fig 3B and 3D, S6 Table). This mode resembles modes that are observed in obligatory complexes in that the dynamic domains mix segments from both subunits.

The two slowest modes dominate the average dynamic behavior of all modes (S2D–S2F Fig), which displays that the dynamic linkage between the two proteins is mostly through a groove, comprising residues GLY100, SER125 and THR220 of the enzyme (chain E), surrounding an anchor, comprising residues MET70-VAL74 of the inhibitor (chain I, S2F Fig). The residues of the anchor are dynamically coupled more strongly to the amino acids of the enzyme than the amino acids of their own chain, i.e., the inhibitor. The anchor-and-groove maintains the dynamic coupling between the subunits, presumably also the functional interaction. S4 Table shows the threshold values and the attributes of each decision variable for this case as

Table 1. The success rate of DynaFace.

	Obligatory Interfaces	Non-obligatory Interfaces	Overall
Two slowest modes	62.6%	64.6%	63.9%
Ten slowest modes	69.1%	72.8%	71.4%
Ten slowest modes combined with all modes	84.2%	82.5%	83.1%

The success rate of DynaFace to reproduce the interface annotations of S1 Table based on various normal mode combinations.

doi:10.1371/journal.pcbi.1004461.t001

given in Eq 4. Four of the seven attributes (AR, A_a, N_a, A_s) suggest that this protein is non-obligatory, as it really is.

Motion and interaction: Large-scale evaluation

The examples above suggest that it could be possible to discriminate between obligatory (1QU7) and non-obligatory (2SIC) interactions based on their dynamic footprints (Fig 2 and Fig 3). To examine how feasible it is we conduct GNM calculations of the 139 obligatory and 246 non-obligatory protein complexes in our dataset and analyze dynamic correlation maps as described in the flow chart of Fig 1. Preliminary tests are conducted using the fluctuations of the two slowest, ten slowest, and all modes of motion. The sense and magnitude of the correlations between intra- and inter-chain residue fluctuations and the number of anchor regions are analyzed for a best scoring metric to predict the interface type.

The results, summarized in Table 1, show that even estimates based only on the two slowest modes yield a significant success rate for the prediction: 62.6% (87 out of 139) and 64.6% (159 out of 246) for obligatory and non-obligatory interfaces, respectively. The success rates increase to 69.1% (96 out of 139) and 72.8% (179 out of 246), respectively, when the ten slowest modes are considered. Furthermore, combining the ten slowest and all modes, the success rates are as high as 84.2% (117 out of 139) and 82.5% (203 out of 246), respectively. To examine the stability of the results obtained with the latter setting we divide the dataset randomly into five subsets. Reassuringly, the success rates of the five sets are 84.6 ± 5.8% and 82.6 ± 4.6% for obligatory and non-obligatory complexes, respectively. With the latter results, the server now considers the ten slowest and all modes of motion in the background analysis for the best decisive outcome. These results suggest that the dynamics is able to capture largely the global topological similarity of obligatory and non-obligatory complex structures and provides a plausible measure for the interface type prediction. We should keep in mind, though, that there is continuous spectrum of dynamic behaviors from non-obligatory to obligatory interactions, rather than a clear separation into two clusters. Our choice of thresholds is optimal, albeit arbitrary.

As a reference we examine the success rates of NOXclass [38] on the same dataset (Table 2). NOXclass fails to process the query for PDB entries of more than two chains. Thus, only the binary complexes in the original set are used. The success rates are 74.2% (89 of 120) for obligatory complexes, and 66.5% (107 of 161) for non-obligatory, respectively. DynaFace's success rates on the same set are significantly higher: 81.7% (98 out of 120 cases) and 83.2% (134 out of 161 cases), respectively. It is noteworthy that the two approaches are complementary to each

Table 2. The success rate of NOXclass and DynaFace for the binary complexes in the set of S1 Table.

	Obligatory Interfaces	Non-obligatory Interfaces	Overall
NOXclass	74.2%	66.5%	69.8%
Dynaface	81.7%	83.2%	82.6%

doi:10.1371/journal.pcbi.1004461.t002

Table 3. The success rate of PQS, PINS, NOXclass, PISA and DynaFace in the detection of the crystal packing interfaces of S7 Table.

PQS	PINS	NOXclass	PISA	DynaFace
61.9%	53.9%	76.2%	55.5%	88.9%

other in that they are based on orthogonal properties; NOXclass uses various local sequence- and structure-based properties, and DynaFace is based on the fluctuation dynamics as a single and mainly global property. Thus, it could be possible to combine DynaFace and NOXclass to improve the performance further.

A similar comparison has been made with PiType. PiType, which uses Uniprot IDs, makes use of already existing interaction networks in its database, and most of the cases in our dataset are not included in their database. We thus made a partial comparison on 62 proteins (45 non-obligatory, 17 obligatory) that exist both in our dataset and PiType's. DynaFace successfully reproduced annotation of 85.4% (82.3% of obligatory and 86.7% of non-obligatory complexes) of the complexes, while PiType reproduced only 80.6% (70.6% of obligatory and 84.4% of non-obligatory complexes) of these complexes, respectively.

DynaFace classifies crystal packing as non-obligatory interfaces

Due to the close packing of protein molecules in crystals, some interactions observed in X-ray crystal structures are non-biological. PDB entries are thus known to contain artifacts of crystallization that do not have any biological relevance [10]. It is of interest to examine how these are classified by DynaFace.

Because a reliable gold-standard of non-biological interfaces is not available, we compiled one based on previous studies. The starting point was a dataset of 63 large crystal packing interfaces with buried surface area of 800 Å^2 or more, where the monomeric state was confirmed by biochemical or biophysical studies [53]. DynaFace calculations failed for four of the complexes due to the appearance of a singular connectivity matrix due to the topology of the complex structure where the minimum distance of alpha carbon atoms between the two subunits of the interface is > 7 Å. These complexes have been removed from the dataset. Of the rest of the 59, three were classified as obligatory and the rest as non-obligatory. Next, we applied existing computational tools for the detection of crystal packing interfaces; PQS [54], PITA [55], PISA [56] and PInS [57] and NOXclass [38]. The DIMOVO tool [58] was excluded because it was trained on the present dataset. Table 3 summarizes the success rate of each server and S7 Table shows the detailed results. Each and every method classifies many interfaces as biologically meaningful, rather than being crystal artefacts, suggesting that the methods or dataset are imperfect.

To be on the safe side, we filtered the dataset and applied a particularly restrictive criterion, keeping only 13 crystal-packing interfaces with consensus prediction (Highlighted structures in S7 Table). DynaFace classifies all the remaining interfaces as non-obligatory. Interestingly, ten of them do not display anchors, the key indication of non-obligatory interfaces. The outliers are 1a7v, 2atj and 3ng1, which display anchoring behavior between subunits. The dynamic behavior of crystal and biological non-obligatory complex structures needs to be further investigated on a larger clean dataset of crystal complex structures.

Prediction with models of protein complexes (PrePPI)

Having validated DynaFace on documented cases, we used it also to analyze interfaces of 300 model structures, predicted by the PrePPI tool [6]. The structural models were constructed using a combination of structural and non-structural interaction clues and assigned confidence

levels "high", "low" or "very low". For completion we also analyzed the 85 real dimeric structures used as the templates for the models, assuming that agreement between the classification of the interface of the model and template is indicative of the prediction accuracy. The results are provided in S2 Table. Reassuringly, the agreement between DynaFace predictions of model and template correlated with the confidence of the model according to the PrePPI estimate: 86.0% agreement for the 100 models of the 'high' confidence, 72.6% for the 100 with 'low' confidence, and 57.0% for the 100 with "very low" confidence.

Among 100 model structures in the 'high' PrePPI category, 93 are predicted to be obligatory. Of the templates of these, 86 are also predicted to be obligatory. Only 7 of the model structures in the 'high' category are predicted to be non-obligatory, but their templates are not. Of the 100 PrePPI dimers in the "low" category, 22 are predicted to be obligatory, 73 to be non-obligatory, and 5 could not be processed because the subunits were over 7 Å apart from each other in the model. Consensus between DynaFace prediction for model and template is obtained for only one of the 22 obligatory dimers but for 68 of the 73 non-obligatory interfaces. Of the 100 Pre-PPI structures in the "very low" category, 27 and 73 structures are predicted respectively to be obligatory and non-obligatory; consensus with the prediction for the template is obtained for 10 of the 27 obligatory interfaces and 47 of the 73 non-obligatory ones.

In total there are 76 cases of disagreement in the prediction of the interface type between the model and template. Examination showed that in 52 cases the disagreement is related to large size differences between the polypeptide chains. That is, when a short chain in the template corresponds to a much larger chain in the model or vice versa. The size difference pronouncedly affects the global dynamics, which has significant contribution here to classify the interface type. As an example, the 8API template, predicted as non-obligatory, was used to model P29508_P01009, P35237_P01009, P48594_P01009, P50453_P01009, Q86WD7_P01009, where a short template chain of 36 amino acids corresponds to 369 in the model, which is predicted to be obligatory. We could not find an obvious reason for the observed disagreement in the rest of the cases (24, i.e., 8%) and we attribute it to parametric error in the twilight zone between obligatory and non-obligatory interfaces.

A case study with a modeled complex structure

The NMR structure of the transforming growth factor beta 1 (TGF-B1; 1KLD [59]) was used as a template for PrePPI modeling of ten dimer structures in the "high" confidence category and one in the "very low" category. TGF-B1 is a multifunctional cytokine with stimulatory and inhibitory effects. Its mature form is homodimeric. Indeed, the template is predicted to be obligatory as it should [59]), and so are all the PrePPI models assigned with high confidence level. In contrast, the one PrePPI model, assigned with very low confidence is predicted to be non-obligatory.

As a sample case, the underlying dynamics for the structural model P09529_P01137 is presented in S4 Fig and S8 Table. The RMSD between one of the structural models P09529_P01137 in the "high" category that was analyzed here and the template structure, chains A and B of 1KLD, is 1.6 Å. Both of these structures are found to be obligatory, in line with the experimental findings of this protein structure [60]. According to the slowest mode, the homodimeric structure has five hinges; three in chain A: GLU333/GLY334, CYS372/ILE373 and CYS404/GLY405, and two in chain B: GLY46/PRO47 and CYS77/CYS78. The hinges of both chains align to a single plane, dividing the structure into two dynamic domains. In the second slowest mode, the dynamic domains are defined by hinge residues PHE309/ILE310, TYR327/TYR328, THR379/MET380, ASP395/VAL396 in chain A and LEU20/TYR21, ALA41/ASN42, GLU84/PRO85, MET104/ILE105 in chain B. The two slowest modes dominate the overall dynamic

behavior in this case. Here, the intra- and inter-chain nature of the cooperativity is indistinguishable, assumed hallmark of obligatory interaction.

Conclusion

Protein-protein complexes are involved, in essence, in all intra- and inter-cellular processes, and the nature of the inter-protein interface determines mechanistic aspects of the interaction. The interfaces have been evolutionarily-designed to enable transduction between the proteins, suggesting that analysis of the dynamics of the complex can reveal the nature of the interaction. The global dynamics of protein complexes should be particularly crucial for function. Following this hypothesis, DynaFace exploits the dynamics of protein complexes in order to detect obligatory vs. non-obligatory interactions among the subunits. The global perspective of interactions across the subunits interface is described mainly by the dynamics of the structural complex, which is not easily accessible by studying only local sequence and structural properties. The dynamic domains, the motions of sub-structural units and how they cooperate with respect to the interacting chains, i.e. the dynamic infrastructure provided by the interacting chains, overall captures the global connectivity similarity for the obligatory and non-obligatory complex structures. To this end, the dynamics of protein complexes in terms of the motions of their dynamic domains and how they are dynamically coupled for their function is of importance for design and function modification of proteins. It is important to note that overall, the interaction spectrum is continuous and does not readily land itself to any discrete classifications. Thus, it is not surprising that DynaFace predictions are imperfect.

DynaFace could readily be embedded in other tools for predicting interface type. Such tools often use local characteristics, which are complementary to the global features used in DynaFace. Indeed, an approach that combines both local and global features would provide a means to discover further and more novel aspects of biological processes.

Supporting Information

S1 Fig. Obligatory and non-obligatory interactions often differ in their dynamic characteristics. The results obtained using the dataset of S1 Table, which includes a total of 139 obligatory and 246 non-obligatory complexes. The X-axis represents the value of the decision outcome D in Eq 4, ranging from 0 to 7; the Y-axis shows the fraction of the protein complexes having that decision outcome value D. In DynaFace, complexes assigned a D value greater than or equal to 4 is predicted to be obligatory; otherwise non-obligatory.
(TIF)

S2 Fig. The correlation between residue fluctuations of an obligatory vs. non-obligatory complex. The upper panels show the correlations $\langle \Delta R_i \cdot \Delta R_j \rangle$ for the average over the two slowest modes (A), the ten slow modes (B), and all modes (C) of motion for an obligatory interface, 1QU7 [49]. The lower panels (D, E, and F) show the respective correlations for a non-obligatory interface, 2SIC [52].
(TIF)

S3 Fig. The non-obligatory interaction between the enzyme and inhibitor in the 2SIC complex is mediated using anchor and groove. The anchor residues MAT70-VAL74 on the inhibitor (chain I) are shown as solid spheres, and the groove residues GLY100, SER125, THR220 on the enzyme (chain E) are shown as doted spheres. The figure was produced using PyMOL [61].
(TIF)

S4 Fig. Slow modes of motion in the structural model P09529_P01137 based on the template NMR structure of the transforming growth factor beta 1 (TGF-B1, PBD ID: 1KLD [59]). The top panels show the matrices of correlations between residue fluctuations, $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$, of mode 1 (A) and mode 2 (B) color-coded for negative (blue) and positive (red) correlations. The boundaries between the subunits are marked in white. The bottom panels show projection of the correlations on the 3D-structure: C- mode 1 and D- mode 2. The sub-units are shown on the PyMOL [61] figures in lighter and darker versions of the same colors; spheres are the hinge residues.
(TIF)

S1 Table. The dataset of 246 non-obligatory and 139 obligatory protein complexes [14, 31]. The original references as well as the references used to update the interaction type are given. DynaFace predictions are also included.
(DOCX)

S2 Table. Dataset of 85 template structures and the PrePPI structural models predicted based on these template structures [6] along with their DynaFace predictions. "Model ID"– the PrePPI index number of the predicted dimer model. "Probability"- the PrePPI prediction likelihood: high, low, very low. "Template ID"–the PDB accession number of the template Pre-PPI used to model the dimer. "Template Chain IDs"–the chains PrePPI used to model the structure. "Model"–DynaFace prediction of the interface type of the model structure: obligatory vs. non-obligatory. "Template"–DynaFace prediction of the interface of the template.
(DOCX)

S3 Table. Individual performance of each attribute for obligatory and non-obligatory complex structures in the dataset.
(DOCX)

S4 Table. The threshold value for the seven dynamic attributes used in a DynaFace calculations. The values of the attributes for two examples are presented, where attributes that exceed the threshold are highlighted in bold. PDB IDs 1QU7 and 2SIC are predicted to be obligatory and non-obligatory, respectively, as they should.
(DOCX)

S5 Table. The dynamic building units, structural units and hinge residues of an example obligomer: The homodimeric cytoplasmic domain of the serine chemotaxis receptor (1QU7 [49]).
(DOCX)

S6 Table. The dynamic building units, structural units and hinge residues of an example non-obligatory dimer: Subtilisin BPN' in complex with its streptomyces subtilisin inhibitor (2SIC [52]).
(DOCX)

S7 Table. The comparison of the predictions of the servers; PQS, PINS, NOXclass, PISA and DynaFace on the crystal dataset by [53].
(DOCX)

S8 Table. The dynamic building units, structural units and hinge residues of the structural model P09529_P01137 based on the template NMR structure TGF-B1 (1KLD[59]).
(DOCX)

S1 Method.
(DOCX)

Acknowledgments

We thank Dr. Emek Seyrek for her contribution in the initiation of this project. We thank Adam Ben-Shem, Sarel Fleishman, and Daved Fremont for valuable discussions.

Author Contributions

Conceived and designed the experiments: SS PO NBT TH. Performed the experiments: SS PO. Analyzed the data: SS PO NBT TH. Contributed reagents/materials/analysis tools: SS PO JIG TH. Wrote the paper: SS PO NBT TH.

References

1. Russell R.B. and Aloy P., Targeting and tinkering with interaction networks. Nat Chem Biol, 2008. 4 (11): p. 666–73. doi: 10.1038/nchembio.119 PMID: 18936751

2. Aloy P., Pichaud M., and Russell R.B., Protein complexes: structure prediction challenges for the 21st century. Curr Opin Struct Biol, 2005. 15(1): p. 15–22. PMID: 15718128

3. Berman H.M., et al., The Protein Data Bank. Nucleic Acids Research, 2000. 28(1): p. 235–242. PMID: 10592235

4. Mosca R., et al., Towards a detailed atlas of protein–protein interactions. Current Opinion in Structural Biology, 2013. 23(6): p. 929–940. doi: 10.1016/j.sbi.2013.07.005 PMID: 23896349

5. Petrey D. and Honig B., Structural Bioinformatics of the Interactome. Annual review of biophysics, 2014. 43: p. 193–210. doi: 10.1146/annurev-biophys-051013-022726 PMID: 24895853

6. Zhang Q.C., et al., Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature, 2012. 490(7421): p. 556–60. doi: 10.1038/nature11503 PMID: 23023127

7. Hosur R., et al., A computational framework for boosting confidence in high-throughput protein-protein interaction datasets. Genome Biology, 2012. 13(8): p. R76. doi: 10.1186/gb-2012-13-8-r76 PMID: 22937800

8. Devos D. and Russell R.B., A more complete, complexed and structured interactome. Curr Opin Struct Biol, 2007. 17(3): p. 370–7. PMID: 17574831

9. Janin J. and Rodier F., Protein-protein interaction at crystal contacts. Proteins, 1995. 23: p. 580–587. PMID: 8749854

10. Janin J., Bahadur R.P., and Chakrabarti P., Protein-protein interaction and quaternary structure. Quarterly Reviews of Biophysics, 2008. 41: p. 133–180. doi: 10.1017/S0033583508004708 PMID: 18812015

11. Liu S., Li Q., and Lai L., A combinatorial score to distinguish biological and nonbiological protein-protein interfaces. Proteins, 2006. 64: p. 68–78. PMID: 16596649

12. Nooren I.M.A. and Thornton J.M., Diversity of protein-protein interactions. the The European Molecular Biology Organization Journal, 2003. 22: p. 3486–3492.

13. Wang R.-S., et al., Analysis on multi-domain cooperation for predicting protein-protein interactions. BMC Bioinformatics, 2007. 8(1): p. 391.

14. Mintseris J. and Weng Z., Structure, function, and evolution of transient and obligate protein–protein interactions. Proceedings of the National Academy of Sciences of the United States of America, 2005. 102: p. 10930–10935. PMID: 16043700

15. Kessel A. and Ben-Tal N., Introduction to proteins: structure, function and motion, 2010. CRC press, Taylor and Francis Group. Boca Raton.

16. De S., et al., Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. BMC Structural Biology, 2005. 5: p. 15. PMID: 16105176

17. Berezin C., et al., ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics, 2004. 20: p. 1322–4. PMID: 14871869

18. Guharoy M. and Chakrabarti P., Conservation and relative importance of residues across protein-protein interfaces. Proceedings of the National Academy of Sciences of the United States of America, 2005. 102: p. 15447–52. PMID: 16221766

19. Valdar W.S. and Thornton J.M., Protein-protein interfaces: analysis of amino acid conservation in homodimers. Proteins, 2001. 42: p. 108–124. PMID: 11093265

20. Ofran Y. and Rost B., ISIS: interaction sites identified from sequence. Bioinformatics, 2006. 23: p. e13–e16.

21. Wass M.N., David A., and Sternberg M.J., Challenges for the prediction of macromolecular interactions. Curr Opin Struct Biol, 2011. 21(3): p. 382–90. doi: 10.1016/j.sbi.2011.03.013 PMID: 21497504

22. Ezkurdia I., et al., Progress and challenges in predicting protein-protein interaction sites. Briefings in Bioinformatics, 2009. 10: p. 233–246. doi: 10.1093/bib/bbp021 PMID: 19346321

23. Tuncbag N., Gursoy A., and Keskin O., Prediction of protein-protein interactions: unifying evolution and structure at protein interfaces. Phys Biol, 2011. 8(3): p. 035006. doi: 10.1088/1478-3975/8/3/035006 PMID: 21572173

24. Casari G., Sander C., and Valencia A., A method to predict functional residues in proteins. Nature Structural Biology, 1995. 2: p. 171–178. PMID: 7749921

25. Nimrod G., et al., Detection of functionally important regions in "hypothetical proteins" of known structure. Structure, 2008. 16(12): p. 1755–63. doi: 10.1016/j.str.2008.10.017 PMID: 19081051

26. Nimrod G., et al., In silico identification of functional regions in proteins. Bioinformatics, 2005. 21 Suppl 1: p. i328–37. PMID: 15961475

27. Zhou H.-X. and Qin S., Interaction-site prediction for protein complexes: a critical assessment. Bioinformatics, 2007. 23: p. 2203–2209. PMID: 17586545

28. de Vries S.J. and Bonvin A.M., How proteins get in touch: interface prediction in the study of biomolecular complexes. Curr Protein Pept Sci, 2008. 9(4): p. 394–406. PMID: 18691126

29. Neuvirth H., Raz R., and Schreiber G., ProMate: a structure based prediction program to identify the location of protein-protein binding sites. Journal of Molecular Biology, 2004. 338: p. 181–199. PMID: 15050833

30. Qin S. and Zhou H.-X., meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics, 2007. 23: p. 3386–3387. PMID: 17895276

31. Bradford J.R. and Westhead D.R., Improved prediction of protein-protein binding sites using a support vector machines approach. Bioinformatics, 2005. 21: p. 1487–1494. PMID: 15613384

32. Chen H. and Zhou H.-X., Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. Proteins, 2005. 61: p. 21–35. PMID: 16080151

33. Ogmen U., et al., PRISM: protein interactions by structural matching. Nucleic Acids Research, 2005. 33: p. W331–W336. PMID: 15991339

34. Porollo A. and Meller J., Prediction-based fingerprints of protein-protein interactions. Proteins, 2007. 66: p. 630–645. PMID: 17152079

35. Tyagi M., et al., Homology inference of protein-protein interactions via conserved binding sites. PLoS One, 2012. 7(1): p. e28896. doi: 10.1371/journal.pone.0028896 PMID: 22303436

36. Huang J., et al., metaPIS: a sequence-based meta-server for protein interaction site prediction. Protein Pept Lett, 2013. 20(2): p. 218–30. PMID: 22894160

37. Aytuna A.S., Gursoy A., and Keskin O., Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics, 2005. 21: p. 2850–2855. PMID: 15855251

38. Zhu H., et al., NOXclass: prediction of protein-protein interaction types. BMC Bioinformatics, 2006. 7: p. 27. PMID: 16423290

39. Winter C., et al., SCOPPI: a structural classification of protein–protein interfaces. Nucleic Acids Research, 2006. 34: p. D310–D314. PMID: 16381874

40. Murzin A.G., et al., SCOP: a structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology, 1995. 247: p. 536–40. PMID: 7723011

41. La D., et al., Predicting permanent and transient protein-protein interfaces. Proteins, 2013. 81(5): p. 805–18. doi: 10.1002/prot.24235 PMID: 23239312

42. Goebels F. and Frishman D., Prediction of protein interaction types based on sequence and network features. BMC Syst Biol, 2013. 7 Suppl 6: p. S5. doi: 10.1186/1752-0509-7-S6-S5 PMID: 24564924

43. Bhardwaj N., et al., Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. Protein Sci, 2011. 20(10): p. 1745–54. doi: 10.1002/pro.710 PMID: 21826754

44. Haliloglu T., Bahar I., and Erman B., Gaussian Dynamics of Folded Proteins. Physical Review Letters, 1997. 79: p. 3090–3093.

45. Bahar I., Atilgan A.R., and Erman B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding design, 1997. 2: p. 173–181. PMID: 9218955

46. Berman H.M., et al., The Protein Data Bank. Nucleic Acids Research, 2000. 28: p. 235–242. PMID: 10592235

47. Wang G. and Dunbrack R.L., PISCES: a protein sequence culling server. Bioinformatics, 2003. 19: p. 1589–1591. PMID: 12912846

48. Emekli U., et al., HingeProt: automated prediction of hinges in protein structures. Proteins, 2008. 70: p. 1219–1227. PMID: 17847101

49. Kim K.K., Yokota H., and Kim S.H., Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. Nature, 1999. 400(6746): p. 787–92. PMID: 10466731

50. Takeuchi Y., et al., Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtili-sin inhibitor at 1.8 A resolution. Journal of Molecular Biology, 1991. 221: p. 323–327.

51. Coleman M.D., et al., Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on-off switching. Biochemistry, 2005. 44(21): p. 7687–95. PMID: 15909983

52. Takeuchi Y., et al., Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtili-sin inhibitor at 1.8 A resolution. J Mol Biol, 1991. 221(1): p. 309–25. PMID: 1920411

53. Bahadur R.P., et al., A dissection of specific and non-specific protein-protein interfaces. Journal of Molecular Biology, 2004. 336: p. 943–955. PMID: 15095871

54. Henrick K. and Thornton J.M., PQS: a protein quaternary structure file server. Trends in Biochemical Sciences, 1998. 23: p. 358–361. PMID: 9787643

55. Ponstingl H., Kabir T., and Thornton J.M., Automatic inference of protein quaternary structure from crys-tals. J Appl Cryst, 2003. 36: p. 1116–1122.

56. Krissinel E. and Henrick K., Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology, 2007. 372: p. 774–797. PMID: 17681537

57. Bordner A.J. and Gorin A.A., Comprehensive inventory of protein complexes in the Protein Data Bank from consistent classification of interfaces. BMC Bioinformatics, 2008. 9: p. 234. doi: 10.1186/1471-2105-9-234 PMID: 18474114

58. Bernauer J., et al., DiMoVo: a Voronoi tessellation-based method for discriminating crystallographic and biological protein-protein interactions. Bioinformatics, 2008. 24: p. 652–658. doi: 10.1093/bioinformatics/btn022 PMID: 18204058

59. Hinck A.P., et al., Transforming growth factor beta 1: three-dimensional structure in solution and com-parison with the X-ray structure of transforming growth factor beta 2. Biochemistry, 1996. 35(26): p. 8517–34. PMID: 8679613

60. Lichtenberger F.J., et al., NAC and DTT promote TGF-beta1 monomer formation: demonstration of competitive binding. J Inflamm (Lond), 2006. 3: p. 7.

61. Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

62. Gordon E.H.J., Steensma E., and Ferguson S.J., The Cytochrome c Domain of Dimeric Cytochrome cd1 of Paracoccus pantotrophus Can Be Produced at High Levels as a Monomeric Holoprotein Using an Improved c-Type Cytochrome Expression System in Escherichia coli. Biochemical and Biophysical Research Communications, 2001. 281(3): p. 788–794. PMID: 11237728

63. Shimba N., et al., Herpesvirus protease inhibition by dimer disruption. J Virol, 2004. 78(12): p. 6657–65. PMID: 15163756

64. Kissinger C.R., et al., Crystal structures of human calcineurin and the human FKBP12-FK506-calci-neurin complex. Nature, 1995. 378(6557): p. 641–644. PMID: 8524402

65. Perona J.J. and Martin A.M., Conformational transitions and structural deformability of EcoRV endonu-clease revealed by crystallographic analysis1. Journal of Molecular Biology, 1997. 273(1): p. 207–225. PMID: 9367757

66. Barkalow F.J., Barkalow K.L., and Mayadas T.N., Dimerization of P-selectin in platelets and endothelial cells. Blood, 2000. 96(9): p. 3070–7. PMID: 11049986

67. Wilson C.J., et al., The experimental folding landscape of monomeric lactose repressor, a large two-domain protein, involves two kinetic intermediates. Proc Natl Acad Sci U S A, 2005. 102(41): p. 14563–8. PMID: 16203983

68. Brito J.A., et al., Crystal structure of Archaeoglobus fulgidus CTP:inositol-1-phosphate cytidylyltransfer-ase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper)thermophiles. J Bacteriol, 2011. 193(9): p. 2177–85. doi: 10.1128/JB.01543-10 PMID: 21378188

69. Debreczeny M.P., et al., Monomeric C-Phycocyanin at Room-Temperature and 77-K—Resolution of the Absorption and Fluorescence-Spectra of the Individual Chromophores and the Energy-Transfer Rate Constants. Journal of Physical Chemistry, 1993. 97(38): p. 9852–9862.

70. Cowieson N.P., et al., Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Current Biology, 2000. 10(9): p. 517–525. PMID: 10801440

71. Hanzelmann P., et al., The effect of intracellular molybdenum in Hydrogenophaga pseudoflava on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase. Journal of Molecular Biology, 2000. 301(5): p. 1221–1235. PMID: 10966817

72. Douangamath A., et al., Structural evidence for ammonia tunneling across the (beta alpha)(8) barrel of the imidazole glycerol phosphate synthase bienzyme complex. Structure, 2002. 10(2): p. 185–193. PMID: 11839304

73. Bamford V.A., et al., Structural basis for the oxidation of thiosulfate by a sulfur cycle enzyme. Embo Journal, 2002. 21(21): p. 5599–5610. PMID: 12411478

74. Dey S., et al., The Subunit Interfaces of Weakly Associated Homodimeric Proteins. Journal of Molecular Biology, 2010. 398(1): p. 146–160. doi: 10.1016/j.jmb.2010.02.020 PMID: 20156457

75. Ponstingl H., Henrick K., and Thornton J.M., Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins-Structure Function and Bioinformatics, 2000. 41(1): p. 47–57.

76. Madden D.R., et al., The 3-Dimensional Structure of Hla-B27 at 2.1 Angstrom Resolution Suggests a General Mechanism for Tight Peptide Binding to Mhc. Cell, 1992. 70(6): p. 1035–1048. PMID: 1525820

77. Heikinheimo P., et al., Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. Proceedings of the National Academy of Sciences of the United States of America, 2001. 98(6): p. 3121–3126. PMID: 11248042

78. Niefind K., et al., Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. Embo Journal, 2001. 20(19): p. 5320–5331. PMID: 11574463

79. Robinson R.C., et al., Crystal structure of Arp2/3 complex. Science, 2001. 294(5547): p. 1679–84. PMID: 11721045

80. van den Akker F., Steensma E., and Hol W.G., Tumor marker disaccharide D-Gal-beta 1, 3-GalNAc complexed to heat-labile enterotoxin from Escherichia coli. Protein Sci, 1996. 5(6): p. 1184–8. PMID: 8762150

81. Athanasiadis A., et al., Crystal-Structure of Pvuii Endonuclease Reveals Extensive Structural Homologies to Ecorv. Nature Structural Biology, 1994. 1(7): p. 469–475. PMID: 7664066

82. Tedeschi G., et al., Purification and primary structure of a new bovine spermadhesin. European Journal of Biochemistry, 2000. 267(20): p. 6175–6179. PMID: 11012670

83. Chen F.E., et al., Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. Nature, 1998. 391(6665): p. 410–3. PMID: 9450761

84. Murphy O.J., et al., Hydrogen exchange reveals a stable and expandable core within the aspartate receptor cytoplasmic domain. Journal of Biological Chemistry, 2001. 276(46): p. 43262–43269. PMID: 11553619

85. Coleman R.A., et al., Dimerization of the Tata-Binding Protein. Journal of Biological Chemistry, 1995. 270(23): p. 13842–13849. PMID: 7775442

86. Gomez-Garcia M.R., Losada M., and Serrano A., A novel subfamily of monomeric inorganic pyrophosphatases in photosynthetic eukaryotes. Biochemical Journal, 2006. 395: p. 211–221. PMID: 16313235