ORIGINAL RESEARCH PAPER

Life Science

Volume - 7 | Issue - 3 | March - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

Computational study of protein-protein interaction (PPI) interfaces in Universal stress protein from bacteria

KEYWORDS

Universal stress protein, Protein-protein interaction, PPI interface, Hot spots, Pyruvate kinase, Stress

Alok Jha

Assistant Professor, Department of Biotechnology and Life Sciences, Institute of Bio Medical Education & Research. Mangalayatan University, Beswan, Aligarh, UP, India. Pin: 202145.

ABSTRACT Protein-protein interactions (PPIs) are of importance in biological processes because these interactions play important role in understanding the protein functions and different pathways. This is especially relevant in disease conditions caused by the loss of specific protein-protein interactions in the organism. Establishing protein interaction networks to get information about cellular operations by using computational approaches have attracted tremendous attention. Computational methods are used because of the ability to predict protein-protein interactions and validate the obtained results. In this study protein-protein interactions of Universal stress proteins UspC (E. coli) and USP (Corynebacteriales) and Pyruvate kinase (E. coli. and Bacillus sp.) complexes have been established and the PPI interfaces have been analyzed. Indeed initial analyses of these networks have revealed interesting properties of such protein-protein interactions which may have practical consequences for designing new drugs.

Introduction:

Universal stress proteins (USPs) are present in a diverse set of organisms from archaea and bacteria to fungi and plants. This evolutionary abundance shows their importance for all domains of the tree of life taxonomy. In stress conditions such as the presence of oxidants, uncouplers, DNA damaging agents, nutrient starvation, heat shock or other stress agents that could cause arrest of cell growth, USP acts as a precursor in constituting a natural biological defense mechanism. Under stress, USPs are overproduced and through various mechanisms facilitate the survival of the organisms in such uncomfortable conditions. USPs also help in the persistence and intracellular survival of pathogens including bacterial growth, stress and virulence. Most organisms have multiple paralogs of USPs, where the number of copies depends on the organism. In E. coli, there are six USPs (UspA, UspC (yecG), UspD (yiiT), UspE (ydaA), UspF (ynaF), and UspG (ybdQ), where UspE is a fusion protein composed of two USP units E1 and E2).[9,10] In Arabidopsis thaliana, for example, there are four copies of usp genes. The exact function of USPs is not well understood and there are very little details about their role in aforementioned cellular processes. Thus it would be beneficial to predict the type of process they can be associated with. Such assignments suggest that Universal stress proteins occur both as single domain proteins and fusions with extra domains, where the extra domain may be an additional USP domain in Pseudomonas, a protein kinase domain in plants or an amino acid permease followed by two USP domains in some archaea.[9,10] The Usp proteins are further classified on the basis of their different substrate specificities. The three dimensional structures of these USPs provide valuable information for the understanding of their potential biochemical mechanisms, although the precise biological functions of these proteins are yet to be deciphered. The molecular basis of cellular operations is largely sustained by different type of interactions among proteins. However, it has become possible to combine the traditional study of proteins as independent entities with the analysis of large protein interaction networks.[13] This is of particular interest as many of the properties of complex systems seem to be more closely determined by their interactions than by the characteristics of their individual components. The study of protein interactions is important not only from a theoretical perspective but also in terms of potential practical implications because it might enable new drugs to be developed that can specifically disrupt or modulate protein interactions, instead of simply targeting a given protein's complete set of functions.[1-3] Noncovalent contacts between the residue side chains are the basis for protein folding, protein assembly and PPIs. These contacts induce a variety of associations and interactions among the proteins. Based on their contrasting structural and functional characteristics, PPIs can be classified in many ways.[1-3] On the basis of their interaction surface, they may be homo- or heterooligomeric; as defined by their stability; they may be obligate or nonobligate; as measured by their persistence; they may be transient or permanent. A given protein-protein interface may be a combination of these there specific pairs.[1-3,8] The transient interaction would form signaling pathways while permanent interactions will form a stable protein complex. The region where two protein chains come into contact is the binding site; or for both sides, an interface. In order to identify interface residues and regions that line the protein surfaces, it is essential to know the structures of the proteins. In order to understand binding principles, properties that distinguish interfaces (or, binding sites) from the rest of the protein surface need to be characterized.[1-3,8]

Materials and Methods:

Sequence based prediction approaches:

Predictions of PPIs have been carried out by integrating evidence of known interactions with information regarding sequential homology. This approach is based on the concept that an interaction found in one species can be used to infer the interaction in other species. However, there are two different methods under sequence based criterion. We performed sequence based prediction of PPIs within bacteria using BIPS-BIANA Interolog prediction server that is based on the known interactions of the orthologous genes of other organisms (interologs).[4]

Sequence similarity measures: Sequence similarity between proteins relies on basic local alignment search tool (BLAST) alignments.[14] The query protein universal stress protein UspC from E. coli and USP from Corynebacteriales were aligned against all sequences with known interactions stored in the BIANA MySQL data base.[4] The alignments provide a similarity measure based on the percentage of identical residues aligned and the percentage of sequence length of the queries and templates covered by the alignment (query and template coverage respectively). The threshold of 90% of template coverage has been used to ensure that the prediction is not inferred from local regions of the template interaction. Also, the geometric mean of individual identities (joint identities) and the geometric mean of individual BLAST E values (joint E values) are considered, i.e. E value $\leq 10^{-5}$, Similarity $\geq 30\%$, and alignment coverage $\geq 60\%$.[4]

Domain interactions: Based on the hypothesis that universal stress proteins UspC (E.coli) and USP (Corynebacteriales) have domains interacting with domains present in Pyruvate kinase (E.coli) and Pyruvate kinase (Bacillus sp.) respectively. Also, these domains could be interacting domains in the iPfam or the 3DID databases. The BIPS-BIANA server measured the similarity of the target sequences with Pfam domains and HMMER program as a function of the E-value cut

Volume - 7 | Issue - 3 | March - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

off of 10^{-5} in the pFam A database. [4]

Further, we used STRING server version 10.0 as the STRING database aims to provide a critical assessment and integration of protein-protein interactions, including direct (physical) as well as indirect (functional) associations. [5] It has scalable algorithms for transferring interaction information between organisms. For this purpose, hierarchical and self consistent orthology annotations have been introduced for universal stress proteins and other interacting proteins, grouping the proteins into families at various levels of phylogenetic resolution. The STRING server version 10.0 [5] includes a completely redesigned prediction pipeline for inferring protein–protein associations from co-expression data, an API interface for the *R* computing environment and thus statistical analysis for Universal stress proteins UspC and USP from both gram positive and negative bacteria were performed.

Phylogenetic Tree:
We generated Phylogenetic tree for UspC (E.coli.) and USP (Corynebacteriales) and Pyruvate Kinase (E. coli and Bacillus sp.) by using Clustal Omega. [15] The underlying principle behind this method is that the coevolution between the interacting proteins can be reflected from the degree of similarity from the distance matrices of corresponding phylogenetic trees of the interacting proteins. The set of organisms common to the two proteins are selected from the multiple sequence alignments (MSA) and the results are used to construct the corresponding distance matrix for each protein. The BLAST score [14] could also be used to fill the matrices. The linear correlation is calculated among these distance matrices.

PPI Identification:
We used KFC2 server to identify the protein-protein interaction interface.[6] The KFC model is a machine learning approach for predicting binding hot spots within protein-protein interactions. The KFC model is comprised of two decision tree based classifiers: K-FADE (based on shape specificity features calculated by the Fast Atomic Density Evaluator or FADE) or K-CON (based on biochemical contact features). Each decision tree that provides a set of hierarchical rules for hot spot classification is trained by a supervised learning process to recognize the local structural environments that are indicative of hot spots. [7] Those residues were classified as hot spots if their mutation to alanine resulted in a change of binding energy ($\Delta\Delta G$) greater than 2 kcal/mol. K-FADE predicts hot spots using the size of the residue and the radial distribution of shape specificity and interface points. K-CON predicts hot spots in terms of a residue's intermolecular atomic contacts, hydrogen bonds, interface points and chemical type. [6]

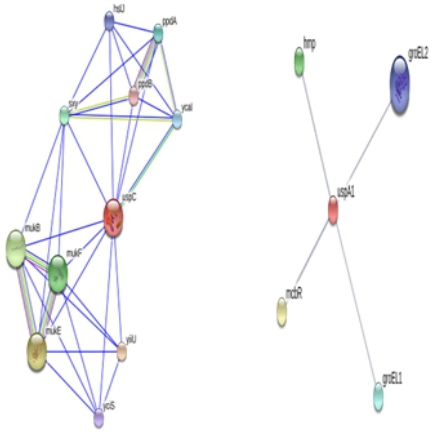
PPI analyses:
We analyzed our proteins, Universal stress proteins, UspC (E. coli) and USP (Corynebacteriales) and Pyruvate Kinase (from E.coli and Bacillus sp.) complexes for protein-protein interaction (PPI) interface by using PDBePISA [12] and PPCheck [11] servers. The stability of the complexes was calculated based on the following physiochemical criteria:

1) free energy of formation.
2) solvation energy gain.
3) interface area.
4) hydrogen bonds.
5) saltbridges across the interface.
6) hydrophobic specificity.

The oligomeric state and symmetry/space group are calculated. Each assembly was assigned a complexation significance score (CSS) that gives information about the importance of the interface in complex formation. Moreover, PPCheck server [11] provided information about all possible types of interactions and energy in the protein- protein complex, including Electrostatic energy, Hydrogen bond energy, Van der wall energy, Total stabilizing energy and Normalized energy per residue etc. [11]

Results and discussion:
The protein-protein interactions for universal stress protein, UspC (E.coli.) and USP (Corynebacteriales) were established by a sequence based approach with the assumption that homologous proteins would have similar behavior. The approach uses sequence similarity between proteins based on the sequence alignment. The alignment of UspC and USP sequences with predicted partner proteins, Pyruvate Kinase (From E.coli and Bacillus sp.) based on 60% identity and 70% of the total length of the target protein and 90% of the template, where, template coverage is fixed to 90%. In a second approach the similarity of the target sequence with Pfam domains as a function of e-value was calculated. This results in the assignation of one or several Pfam domains to the query and target sequences. Then the Interolog prediction server BIPS-BIANA based on interolog information compared the iPfam and 3DiD databases for domain-domain interactions. [4] The homology conditions were maintained with a joint e-value of 1e-10 and joint identity of 80%. The predictions were filtered assuming the traditional definition of interologs that both proteins suppose to interact if they are orthologous with proteins that interact. Both proteins were clustered by comparing with cluster of orthologous genes (COG) database and selected using GO annotations as they share the largest number of similar GO terms. [4]

We further analyzed the PPI network of UspC and USP on the basis of active interaction sources e.g. co-expression, co-occurrence, gene fusion, neighborhood, experiments and databases by using a deterministic network drawing algorithm (STRING Server) that uses a spring model to generate the network images. [5] Nodes are modeled as masses and edges as springs; the final position of the nodes in image is computed by minimizing the 'energy' of the system. The network nodes are proteins and the edges represent the predicted functional associations. The high confidence edges are given a higher 'spring strength' so that they will reach an optimal position before lower confidence edges. The UspC interaction shows 11 nodes, 25 edges, with average node degree 4.55, and clustering coefficient of 0.876.(Fig.1.) The PPI enrichment p-value is 0.000292 and the functional enrichments in the network shows assigned GO annotations for different biological processes with GO accession id GO:0030261, chromosome condensation (3 genes) with a false discovery rate of 0.000391; GO:0007059; chromosome segregation (3 genes) with a false discovery rate of 0.00586; GO:0007049, cell cycle (4 genes) with a false discovery rate of 0.0142; GO:0051301, cell division (4 genes) with a false discovery rate of 0.0142. The cellular component has been assigned a GO accession GO: 0009295, nucleoid (3 genes) with a false discovery rate of 0.00216. In confidence mode the thickness of the line indicate the degree of confidence prediction of the interaction. Action mode shows additional information about the binding, activation etc. [5]



(a) (b)

Volume - 7 | Issue - 3 | March - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

Fig 1. Protein-protein interaction network for UspC (E.coli) and USP (Corynebacteriales). In evidence mode the edge lines predict the associations with different colored lines that are Red line- indicates the presence of fusion evidence, Green line-neighborhood evidence, Blue line- cooccurrence evidence, Purple line-experimental evidence, Yellow line-textmining evidence, Light blue line- database evidence, Black line- coexpression evidence.

The USP interaction shows 5 nodes, 4 edges, with average node degree 1.6 and clustering coefficient of 0.8. (Fig.1.) The PPI enrichment p-value is 0.649 and the functional enrichments in the network shows assigned KEGG pathway annotation for different biological processes with pathway id 03018, RNA degradation (2 genes) with a false discovery rate of 0.0182; PFAM protein domain pathway id PF00118, TCP-1/cpn60 chaperonin family (2 genes) with a false discovery rate of 0.00518; INTERPRO protein domain pathway id IPR001844, chaperonin Cpn60 (2 genes) with a false discovery rate of 0.00229; IPR002423, chaperonin Cpn60/TCP-1 family (2 genes) with a false discovery rate of 0.00229; IPR018370, chaperonin Cpn60 conserved site (2 genes) with a false discovery rate of 0.00229; IPR027409, GroEL-like apical domain (2 genes) with a false discovery rate of 0.00229; IPR027413, GroEL-like equatorial domain (2 genes) with a false discovery rate of 0.00229. [5] These pathway interactions could reveal that USP might play a role in such biological processes.

Further, phylogenetic tree was generated on the basis of multiple sequence alignment of universal stress proteins USPs (from E.coli and Corynebacteriales) and Pyruvate kinases (from E.coli and Bacillus sp.) using the Clustal W and Clustal Omega parameters. [15] The phylogenic tree generation method includes tree format (distance matrix) and clustering method (Neighbor Joining) with distance correlation for more divergent sequences.(Fig.2) The percent identity matrix was used with exclusion of gaps that is meant to forcing the alignment to use only positions where information can be included from all sequences.

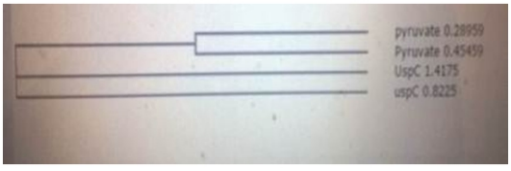
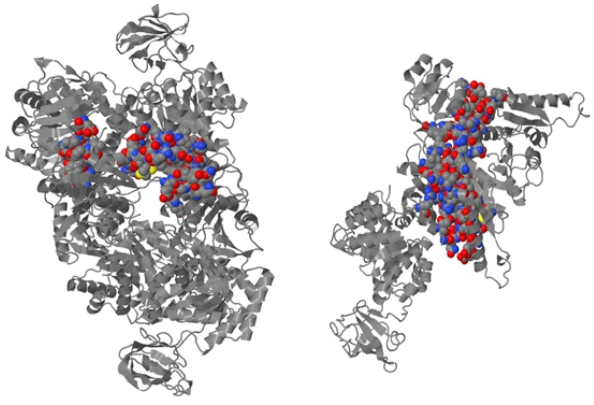
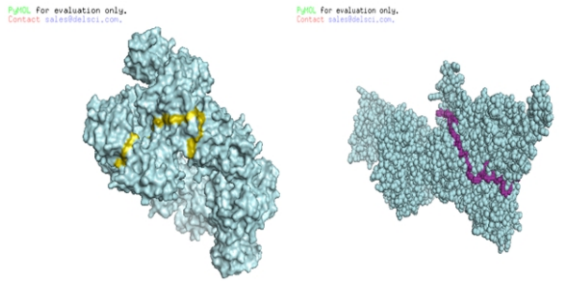


Fig. 2. Phylogenetic tree for UspC (E.coli), USP (Corynebacteriales), Pyruvate Kinase (E.coli) and Pyruvate Kinase (Bacillus sp.)

The phylogenetic tree shows branch distances for Universal stress proteins and Pyruvate kinase with Pyruvate kinase (E.coli), value 0.28959, Pyruvate kinase (Bacillus sp.) value 0.45459, UspC (E.coli.) value 1.4175 and USP (Corynebacteriales) value 0.82250. (Fig.2) This shows that all proteins are evolutionary linked and possibly interact.



(a) (b)



(c) (d)

Fig.3. (a) Hot Spots for UspC-pk complex (E.coli) (b) Hot Spots for USP-pk complex (Corynebacteriales-Bacillus sp. respectively) (c) PPI Interface for UspC-pk (E.coli) (d) PPI Interface for USP-pk (Corynebacteriales-Bacillus sp. respectively)

We identified the PPI interface and hot spots residues that participate in the interaction. [6,7] Predicted hot spot residues based on K-FADE (Fast Atomic Density Evaluator) and K-CON (biochemical contact features) scores include Leu49 (K-FADE=0.52; K-CON=0.68), Arg58 (K-FADE=0.57; K-CON=0.72), Asn59 (K-FADE=0.55; K-CON=0.59), Val60 (K-FADE=0.79; K-CON=0.71) for UspC-pk (E.coli) complex. The hot spot residues for USP-pk (Corynebacteriales) complex include Asp3 (K-FADE= 0.72; K-CON= -0.50), Val6 (K-FADE= 0.49; K-CON= -0.28), Ala 31 (K-FADE= 0.34; K-CON= -0.70), Arg32 (K-FADE= 0.08; K-CON= 0.15), Asp33 (K-FADE= 1.24; K-CON=-0.22), Pro35 (K-FADE= 1.39; K-CON= -0.20), Val85 (K-FADE= 0.89; K-CON= -0.21), Leu87 (K-FADE= 1.69; K-CON= 0.19) , Gln89 (K-FADE= 0.35; K-CON=0.08, Gln90 (K-FADE=-0.30; K-CON=0.02), Tyr174 (K-FADE= 0.62; K-CON= -0.05), Asp178 (K-FADE=0.70; K-CON= -0.26), Ala179 (K-FADE= 0.54; K-CON= -0.46) for chain A. The Hot spot residues in chain B include Thr355 (K-FADE= 0.17; K-CON= -0.66), Ala360 (K-FADE= 0.79; K-CON= -0.50), Ile361 (K-FADE= 0.69; K-CON= 0.06), Ser364 (K-FADE= 0.68; K-CON= -0.46), His367 (K-FADE= 0.34; K-CON= -0.18), His368 (K-FADE= 0.58; K-CON= -0.17), Leu372 (K-FADE= 0.12; K-CON= -0.07), Asn467 (K-FADE= 0.15; K-CON= -0.33), Leu468 K-FADE= 0.14; K-CON= 0.01), Val471 (K-FADE= 1.13; K-CON= -0.15), Gly483(K-FADE= 0.26; K-CON= -0.44), Phe576 (K-FADE= 1.27; K-CON= 0.21); Leu587 (K-FADE= 0.31; K-CON= -0.17).[6,7] The best value for K-FADE is 1 and the worst is 0. The hot spot residues and the PPI interfaces for UspC-pk (E.coli) and USP-pk (Corynebacteriales-Bacillus sp.) are depicted in Fig 3. The PPI analyses from PDBePISA server [12] suggested that the UspC and Pyruvate Kinase (E.coli) complex has assemblies with multimeric state of 5 with formula AB4, composition ABCDE and the dissociation pattern is A+BCDE. The surface area of the complex that indicates the total solvent-accessible surface area of the assembly in $Å^2$ is 74016.2 and the buried area that indicates the total solvent accessible area of the assembly, buried upon formation of all assembly's interfaces in $Å^2$ is 10378.7. ΔG^{int} that indicates the solvation free energy gain upon formation of the assembly in kcal/mol. is -62.7 kcal/mol. The value of ΔG^{diss}, which indicates the free energy of assembly dissociation in kcal/mol is 4.5 kcal/mol.[12] The free energy of dissociation corresponds to the free energy difference between dissociated and associated states. Positive values of ΔG^{diss} indicate that an external driving force should be applied in order to dissociate the assembly, therefore, the assemblies with $\Delta G^{diss} > 0$ are thermodynamically stable. The rigid body entropy change at dissociation $T\Delta S^{diss}$ in kcal/mol is 12.2 kcal/mol. The symmetry number that indicates the number of different but equivalent orientations of the assembly, which can be obtained by rotation, its value is 1 in the UspC-pk complex. The surface of the UspC-pk complex contains 1825 atoms (52.7%) and 1792 atoms (55.3%) respectively and the interface contains 103 atoms (3%) and 100 atoms (3.1%) respectively. The complex surface contains 401 residues (86.4%) and 383 residues (88.5%) respectively and the interface

Volume - 7 | Issue - 3 | March - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

contains 29 residues (6.2%) and 31 residues (7.2%) respectively. [12] The interface solvent accessible area in Å are 1074.5 (5.6%) and 1074.8 (5.6%) respectively. The solvation energies of the isolated structures in the UspC-pk complex in kcal/mol are -445.7 and -405.4 respectively. The gain on complex formation in kcal/mol are -6.3 (1.4%) and -6.3 (1.6%) respectively and the average gain accounts for -1.8 kcal/mol (0.4%) and -2.6 kcal/mol (0.6%) respectively. The P-values for the UspC-pk complex are assigned as 0.068 and 0.104 respectively. The interface residues involved in hydrogen bond formation in the Usp-pk complex include Arg334 with Glu254 and Glu255 (3.89, 2.82 and 2.77 Å), Arg292 with Met279 and Asp297 (3.73 and 2.81 Å, 3.05 Å), Lys261 with Asn300 (3.04 Å), Glu255 with Arg334 (2.72 and 2.69 Å), Met279 with Arg292 (3.53 Å), Asp297 with Arg292 (2.93 and 3.89 Å), Asn300 with Lys261 (3.17 Å). The interface residues involved in salt bridge formation include Arg334 with Glu255 (2.82, 3.86, 3.08 and 2.77 Å), Arg292 with Asp297 (3.05 and 3.80 Å), Glu255 with Arg334 (2.72, 3.92, 3.07 and 2.69 Å) and Asp297 with Arg292 (2.93, 3.89 and 3.89 Å). The interface in complex Formation Significance Score (CSS) assigned is zero. CSS ranges from 0 to 1 as interface relevance to complex formation increases. However, achieved CSS implies that the interface does not play any role in complex formation and seems to be a result of crystal packing only.[12]

The surface of the USP-pk complex contains 1487 atoms (68.4%) and 2321 atoms (53.2%) respectively and the interface contains 155 atoms (7.1%) and 140 atoms (3.2%) respectively. The surface contains 290 residues (99%) and 517 residues (88.1%) respectively and the interface contains 50 residues (17.1%) and 42 residues (7.2%) respectively. The interface in USP-pk complex has solvent accessible area 1529.3 Å (9.1%) and 1481.5 Å (5.9%) respectively. The solvation energies of isolated structures in kcal/mol are -244.6 and -573.1. The gains on complex formation are -5.4 kcal/mol (2.2%) and -10.4 kcal/mol (1.8%) respectively. [12] Moreover, the average gain are -6.5 kcal/mol (2.7%) and -1.8 kcal/mol (0.3%). The P-values for USP-pk complex have been assigned as 0.619 and 0.008 respectively. The solvation free energy gain upon formation of the interface $\Delta^{i}G$ in kcal/mol is -15.8 kcal/mol and the $\Delta^{i}G$ P-value has been assigned to the complex is 0.072. However, a negative $\Delta^{i}G$ corresponds to hydrophobic interfaces, or positive protein affinity. The interface in complex Formation Significance Score (CSS) assigned is zero. However, achieved CSS implies that the interface does not play any role in complex formation and seems to be a result of crystal packing only. [12] The interface residues involved in hydrogen bond formation include Asp3 with Tyr580 (3.35 Å), Glu61 with Gln354 (3.67 Å), Gln68 with Thr355 (2.88 Å), Glu84 with Asn371 (2.22 Å), Tyr174 with Gly483 (3.83 Å). The USP-pk complex has electrostatic energy of 8.91 kJ/mol, Van der Wall energy of -24.60kJ/mol, total stabilizing energy of -15.69kJ/mol and normalized energy per residue of -0.11kJ/mol. This indicates that the energetic signatures in the isolated proteins are retained in the bound forms that can help determining the binding orientation of proteins on complex formation. It also shows for protein surface networks characterized by strong couplings among constitutive residue pairs, as well as surface networks characterized by weak couplings among residue pairs. Strong couplings define residue networks important for structure stabilization and weak couplings (or low coupling) sub networks identify regions that are uncoordinated with the rest of the protein and are more likely to interact with other partners. The analysis of such energetic motifs could help identify binding sites and their orientations from the monomeric, isolated partners for a diverse ensemble of protein-protein assemblies.

Conclusion:

The dominant philosophy in rational drug design i.e. the "one gene, one drug and one disease" paradigm focuses on the individual properties of a protein, for example, whether, it is essential for survival. Many effective drugs with different phenotypic effects can affect a group of molecular targets rather than a single protein. From a system's biology perspective, a protein's importance cannot be well defined by its individual biochemical function(s) but also its position in the protein-protein interaction (PPI) network i.e. its potential for interacting with other proteins. As the role of functional dysregulation of PPIs as the underlying cause of disease is well understood, network pharmacology that advocates combination therapies targeting multiple interconnected nodes in a PPI network represents a new setting for disease treatment.

References:

1. Rao, V. S., Srinivas, K., Sujini, G. N., Sunand Kumar, G. N. (2014) Protein-Protein Interaction Detection: Methods and Analysis. International Journal of Proteomics, 1-12.
2. Zhang, M., Su, S., Bhatnagar, R.K., Hassett, D. J., Lu. L. J. (2012) Prediction and Analysis of the Protein Interactome in Pseudomonas aeruginosa to Enable Network-Based Drug Target Selection. 7, 7: 1-13.
3. Tuncbag, N., Kar, G., Keskin, O., Gursoy, A., Nussinov, R. (2009) A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. Briefings in Bioinformatics. 1-16.
4. Garcia, J.G., Schleker, S., Seetharaman, J. K., Oliva, B. (2012) BIPS: BIANA Interolog Prediction Server. A tool for protein–protein interaction inference. Nucleic Acids Res. 40, 147-151.
5. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D. Cepas, J.H., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., Mering, C.V. (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43:447-452.
6. Darnell, S.J., LeGault, L. Mitchell, J.C. (2008) KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Research. 36: 265-269.
7. Kortemme, T., Baker, D. (2002) A simple physical model for binding energy hot spots in protein–protein complexes. Proceedings of the National Academy of Science. 99,22:14116-14121.
8. Valencia, A., Pazos. F. (2002) Computational methods for the prediction of protein interactions. Current Opinion in Structural Biology.12:368–373.
9. Tkaczuk K.L., Shumilin I.A., Chruszcz M., Evdokimova E., Savchenko A., Minor W. (2013) Structural and functional insight into the universal stress protein family. Evol Appl. 6, 3: 434–49.
10. Siegele D.A. (2005) Universal stress proteins in Escherichia coli. J Bacteriol. 187, 18:6253–4.
11. Sukhwal, A., Sowdhamini, R. (2013) Oligomerisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. Mol. Biosyst. 9, 1652-1661.
12. Krissinel, E., Henrick, K. (2007). 'Inference of macromolecular assemblies from crystalline state.'. J. Mol. Biol. 372, 774--797.
13. Laraia, L., McKenzie, G., Spring, D R., Venkitaraman, A R., Huggins, D J. (2015) Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. Chem Biol. 18; 22, 6: 689–703.
14. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.
15. The EMBL-EBI bioinformatics web and programmatic tools framework. (2015) Nucleic acids research. 4:580-4.