ORIGINAL RESEARCH PAPER

Volume-6 | Issue-10 | October-2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 78.46

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

COMPUTATIONAL STUDY OF PROTEIN-PROTEIN INTERACTION (PPI) INTERFACE OF BACTERIAL UNIVERSAL STRESS PROTEIN (USPC) AND DNA POLYMERASE III SUBUNIT DELTA COMPLEX

Biotechnology

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

The protein-protein interaction network is a useful tool for systematic investigation of the complex biological activities in the cell. With increasing attention on proteome wide interaction networks, computational approaches can provide useful insights into the molecular basis of diseases by studying disease related subnetworks. Here, we present PPI interface study of universal stress protein (UspC) and DNA polymerase III subunit delta complex from E. coli. This study tries to integrate Universal stress protein (UspC), which has a functional role in providing resistance to DNA damaging agents and the structure of DNA polymerase III subunit delta that suggests a model for interaction of Pol III with the sliding clamp and DNA. These macromolecular structures engage in tasks critical for cell survival, such as regulation of DNA replication as well as a myriad of minor but important functions.

KEYWORDS

Universal stress protein, UspC, Protein-Protein Interaction, PPI interface, Hot spots, DNA polymerase

Introduction:

The evolutionary abundance of Universal stress proteins (USPs) in a diverse set of organisms from archaea and bacteria to fungi and plants shows their importance for all domains of the tree of life taxonomy. The universal stress protein appears to belong to all E.coli stress and starvation stimulons under the global control of gene expression.[14] Universal stress protein (USP) acts as a precursor in constituting a natural biological defense mechanism under 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. Under stress, USPs are overproduced and through various mechanisms facilitate the survival of the organisms in such uncomforting conditions. 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).[7,8] Usp genes including uspA, uspC, uspD and uspE respond to stress conditions causing growth arrest and under extreme conditions production of Usp proteins are repressed.[8,15,16] The bacterial usp genes usually encode either a small Usp protein (around 14-15 kD) that has one Usp domain, or a larger version (around 30 kD) consists of two Usp domains in tandem.[14] DNA polymerase III holoenzyme contains two DNA polymerases embedded in a particle with 9 other subunits. This multisubunit DNA polymerase is the Eschericia coli chromosomal replicase, and it has several special features that distinguish it as a replicating machine. For example, one of its subunits is a circular protein that slides along DNA while clamping the rest of the apparatus to the template. Other subunits act together as a matchmaker to assemble the ring onto DNA. Overall, E. coli DNA polymerase III holoenzyme is comparable in both structure and function to the chromosomal replicases of eukaryotes, from yeast to humans.[17] The core of the polymerase contains the catalytic polymerase subunit, α, the proofreading $3' \rightarrow 5'$ exonuclease, ε, and a subunit of unknown function, θ. The three subunits in the core polymerase, only α is required to form processive replication forks that move at high rates and that exhibit coupled leading- and lagging-strand synthesis in vitro. The primary determinant of replication fork processivity is the interaction between another holoenzyme subunit, τ, and the replication fork helicase, DnaB.[18] The gamma complex, an AAA+ ATPase, is the bacterial homolog of eukaryotic replication factor C (RFC) that loads the sliding clamp (beta, homologous to PCNA) onto DNA. The crystal structure of gamma complex reveals a pentameric arrangement of subunits, with stoichiometry delta': gamma (3): delta. The C-terminal domains of the subunits form a circular collar that supports an asymmetric arrangement of the N-terminal ATP binding domains of the gamma motor and the structurally related domains of the delta' stator and the delta wrench.[19] 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. Protein-protein interaction plays a central role in cellular function, improving the understanding of complex formation has various implications such 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. The implication of understanding PPI network includes rational design of new therapeutic agents targeting different signal transduction networks. [1-3,6,11]

Materials and Methods:

Sequence based PPI prediction:

The PPI prediction was made on the basis of sequence similarity and domain interactions. Sequence similarity between proteins relies on basic local alignment search tool (BLAST) alignments.[12] 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%. Again, the domains present on both proteins could be interacting domains in the iPfam or the 3DID databases. The similarity of the target sequences is measured with Pfam domains and HMMER program as a function of the E-value cut off of 10-5 in the Pfam A database. Further, we used STRING server version 10.0 as the STRING database aims to provide a critical assessment and integ ration of protein-protein interactions, including direct (physical) as well as indirect (functional) associations. [4]

3D Network prediction:

We used Interactome3D, a resource for the structural annotation and modeling of protein-protein interactions for PPI network prediction of universal stress protein UspC. Interactome3D through the integration of interaction data from the main pathway repositories provides structural details at atomic resolution.[20]

Phylogenetic Tree:

We generated Phylogenetic tree for UspC and DNA polymerase III subunit delta (E.coli) by using Clustal Omega.[13] 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 [12] could also be used to fill the matrices. The linear correlation is calculated among these distance matrices.

PPI identification:

Protein-protein interactions are critically dependent on just a few 'hot spot' residues at the interface.[5] Hot spots make a dominant contribution to the free energy of binding and they can disrupt the interaction if mutated to alanine. We used HSPred [21], a support vector

machine (SVM)-based method to predict hot spot residues. HSPred uses basic energetic terms that contribute to hot spot interactions, i.e. van der Waals potentials, solvation energy, hydrogen bonds and Coulomb electrostatics, and treat them as input features of a Support Vector Machine (SVM) classifier. This method could predict hot spots with overall good accuracy, comparing favourably to other available approaches.[21]

PPI analyses:

We analyzed Universal stress proteins, UspC and DNA polymerase III subunit delta (E. coli) complexes for protein-protein interaction (PPI) interface by using PDBePISA [10] and PPCheck [9] servers. 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.[10] Moreover, PPCheck server [9] 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. [9]

Results and discussion:

The protein-protein interactions for universal stress protein, UspC (E.coli) and DNA polymerase III subunit delta were established by a sequence based approach and a detection method called two hybrid pooling approach with assumption that homologous proteins would have similar behavior. The approach uses sequence similarity between proteins based on the sequence alignment. This approach was supported by Interactome3D [20] predictions that use specific criteria to rank structures and models for single proteins; complete experimental structures are ranked in order of decreasing coverage. A structure is considered complete if it covers more than 80% of the protein sequence length; here it is 99% with SeqId 100%. A complete experimental structure always has the rank minor index equal to zero. Complete homology models are ranked just after complete experimental structures in decreasing order of sequence identity of the template to the target. A complete homology model always has the rank minor index equal to 0. Partial experimental structures and homology models are ranked after complete homology models. In this case Interactome3D tries to group together structures that cover, together, the largest portion of the protein sequence length. The different groups are generated by a greedy algorithm that takes into consideration the scoring function which takes into account the coverage of the structure and the sequence identity of the template to the target protein. Here, the interaction of UspC and DNA polymerase III subunit delta is complete experimental interaction. Structures for proteins and interactions are collected from the Protein Data Bank (PDB) while homology models for single proteins are collected from Modbase and homology models for interactions are generated using Modeller on structural templates extracted either from PDB biological unit files or domain-domain structural templates downloaded from 3did. [20]

Further, phylogenetic tree was generated on the basis of multiple sequence alignment of universal stress proteins UspC and DNA polymerase III subunit delta from E.coli using the Clustal W and Clustal Omega parameters.[13] The phylogenetic tree suggested a common evolutionary significance of both the protein with a distance of 0.74627 (Fig.1) The phylogenic tree generation method includes tree format (distance matrix) and clustering method (Neighbor Joining) with distance correlation for more divergent sequences. 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.

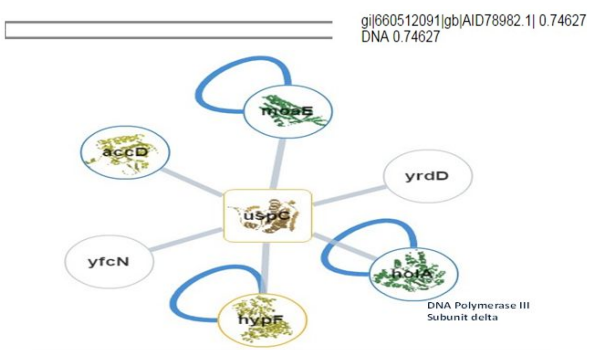

Fig 1. (a) Phylogenetic tree of universal stress protein UspC and DNA polymerase III subunit delta from *E. coli* (b) Interaction network for universal stress protein UspC.

We identified the PPI interface and hot spots residues that participate in the interaction. 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. The predicted hot spot residues include Ala11, Val12, Thr13, Pro14, Glu15, Ser41, Pro43, Glu44, Met45, Tyr46, Leu49, Ala50, Ala51, Pro52, Met53, Asp56, Leu57, Val60, Met61, Glu64, Ser67, Phe68, Lys71, His111, Asn112, His113, Ser114, Phe115, Phe116, Ser117, Arg118 with score 1.0. The hot spot residues and the PPI interfaces for UspC-DNA polymerase III subunit delta are depicted in Fig.2. The PPI analyses from PDBePISA server [10] suggested that the UspC and DNA polymerase III subunit delta complex has assemblies with multimeric state of 2 with formula AB, composition AD and the dissociation pattern is A+D.

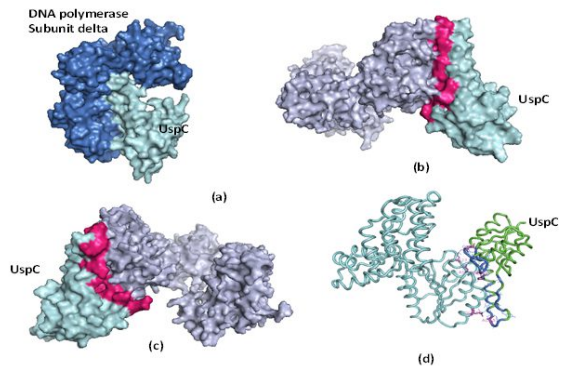

Fig.2. (a) UspC-DNA polymerase III subunit delta complex *(E.coli)* (b,c) UspC-DNA polymerase III subunit delta complex with Hot Spots (Magenta) (d) Hot Spot residues (blue) and hydrogen bonded residues (Magenta) in the UspC-DNA polymerase III subunit delta complex.

The surface area of the complex that indicates the total solvent-accessible surface area of the assembly in $Å^2$ is 25070.2 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 1908.9. ΔG^{int} that indicates the solvation free energy gain upon formation of the assembly in kcal/mol. is -18.0 kcal/mol. The value of ΔG^{diss}, which indicates the free energy of assembly dissociation in kcal/mol is 7.1 kcal/mol.[10] 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 12.3kcal/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-DNA polymerase III subunit delta complex. The engaged interfaces correspond to interfacing structures A+D with N_{occ} 1 that indicates the total number of the corresponding interfaces in the assembly. Also, the buried area is 954.4 $Å^2$ (50%) that indicates the interface area calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two. The solvation free energy gain upon formation of interface Δ^iG is -18.0 kcal/mol. The value is calculated as difference in total salvation energies of isolated and interfacing structures. Negative Δ^iG corresponds to hydrophobic interfaces, or positive protein affinity. This value does not include the effect of satisfied hydrogen bonds and salt bridges across the interface. The surface of the UspC-DNA polymerase III subunit delta complex contains 657 atoms (62.5%) and 1655 atoms (61.6%) respectively and the interface contains 91 atoms (8.7%) and 103 atoms (3.8%) respectively. The complex surface contains 127 residues (94.1%) and 318 residues (94.1%) respectively and the interface contains 30 residues (22.2%) and 27 residues (8.0%) respectively. [10] The interface solvent accessible area in Å are 999.5 (12.7%) and 909.4 (4.8%) respectively. The solvation energies of the isolated structures in the UspC-DNA polymerase III subunit delta complex in kcal/mol are -126.2 and -322.4 respectively. The gain on complex formation in kcal/mol are -11.6 (9.2%) and -6.5 (2.0%) respectively and the average gain accounts for -3.0 kcal/mol (2.4%) and -1.2 kcal/mol (0.4%)

respectively. The P-values for the UspC-DNA polymerase III subunit delta complex are assigned as 0.004 and 0.058 respectively. The interface residues involved in hydrogen bond formation in the UspC-DNA polymerase III subunit delta complex include Ala51 with Gln69 (3.67 Å), Phe115 with Asn94 (3.34 Å), Glu15 with Asn123 (3.41 Å). Potential hydrophobic interactions include Val12 and Ala131 (6.71 Å), Tyr46 and Phe65 (6.67 Å), Leu49 and Phe65 (6.62 Å), Ala51 and Leu73 (6.03 Å), Leu57 and Leu103 (6.49 Å), Val60 and Leu104 (6.59 Å), Phe 115 and Ala91, Ala124, Ala125 (5.29, 6.06 and 6.30 Å), Phe116 and Ala91 (5.32 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.[10] The UspC-DNA polymerase III subunit delta complex has Van der Wall energy of 66.54 kJ/mol, total stabilizing energy of 66.54 kJ/mol and normalized energy per residue of 0.75 kJ/mol. The mutation of Asn110 to Ala110 doesn't change total energy of the complex; also, it doesn't affect normalized energy per residue and residue interaction.[9] For the UspC-DNA polymerase III subunit delta complex the hotspot residues are predicted by a method that integrates structural and energetic features with two types of structural neighborhoods (Euclidian and Voronoi), and combines random forest and sequential backward elimination algorithms to select an optimal subset of features; that is an advantage to the approaches based on the recognition of differences in physicochemical properties, evolutionary conservation and solvent accessible area. Further, the hotspots are utilized for focusing structure based pharmacophore modeling. The analysis of energetic motifs, the change in solvent accessible surface area upon binding and gain in binding energy on complex formation and the identification of hotspot residues in PPIs is useful not only to provide insights into mechanism of protein-protein recognition but also to indicate the areas to be targeted with small molecules.[11] Designing drugs that target hot spots is a good strategy to break PPIs, since they have a major contribution in the stability. Moreover, the hotspots of PPIs help detecting the regions which a ligand may bind with high affinity. These approaches can accurately predict and characterize the PPI interfaces as well as modulation of hot spots could be an important strategy for blocking PPIs.[3,11]

Conclusion:

Although, there are various experimental methods to solve the structures of macromolecular complexes and investigate the mechanisms of PPI formation, the experimental detection of PPI is very difficult. Alternatively, PPIs can be predicted by computational methods, and can be sufficiently useful to swift functional hypotheses and direct experiments. The analysis of protein-protein interaction as described above shows that the crystal packing and specific interfaces have some geometric and physicochemical properties. Also, similar properties between different types of interfaces needs more attention and it would be relevant to find and specially characterize new determinants so that the performance of computational methods can be improved. Moreover, PPI networks can be used in biomedicine to unravel the molecular basis of disease by studying disease related subnetworks. PPI networks offer new opportunities for the development of new therapeutic interventions.

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. 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.
5. 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.
6. Valencia, A., Pazos. F. (2002) Computational methods for the prediction of protein interactions. Current Opinion in Structural Biology.12:368–373.
7. 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.
8. Siegele D.A. (2005) Universal stress proteins in Escherichia coli. J Bacteriol. 187, 18:6253–4.
9. Sukhwal, A., Sowdhamini, R. (2013) Oligomerisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. Mol. Biosyst. 9, 1652-1661.
10. Krissinel, E., Henrick, K. (2007). 'Inference of macromolecular assemblies from crystalline state.' J. Mol. Biol. 372, 774--797.
11. 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.
12. 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.
13. The EMBL-EBI bioinformatics web and programmatic tools framework. (2015) Nucleic acids research. 4:580-4.
14. Kvint, K., Nachin, L., Diez, A., Nystro¨m, T. (2003) The bacterial universal stress protein: function and regulation. Current Opinion in Microbiology. 6:140–145.
15. Nystro¨m T, Neidhardt FC. (1992) Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in Escherichia coli. Mol Microbiol. 6:3187-3198.
16. Gustavsson N, Diez A, Nystro¨m T. (2002) The universal stress protein paralogues of Escherichia coli are co-ordinately regulated and co-operate in the defence against DNA damage. Mol Microbiol. 43:107-117.
17. Kelman Z, O'Donnell M. (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annu Rev Biochem. 64:171-200.
18. Marians, KJ., Hiasa, H., Kim, DR., McHenry, CS. (1998) Role of the Core DNA Polymerase III Subunits at the Replication Fork α is the only subunit required for processive replication. The Journal of Biological Chemistry. 273, 2452-2457.
19. Lamers MH, Georgescu RE, Lee SG, O'Donnell M, Kuriyan J. (2006) Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III. Cell. 8; 126 (5): 881-92.
20. Mosca, R., Céol, A., Aloy, P. (2013) Interactome3D: adding structural details to protein networks. Nature Methods. 10, 47–53.
21. Lise, S., Buchan, D., Pontil, M., Jones, DT (2011) Predictions of Hot Spot Residues at Protein-Protein Interfaces Using Support Vector Machines PlosOne.. 6,2, e16774:1-7