ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 6, Issue, 07, pp.7443-7447, July, 2014

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

ACCUMULATION OF INTERACTING PARTNERS AND OLIGOMERISATION STATUS FOR PROTEIN DOMAINS OF PASS2 SINGLE-MEMBERED AND TWO-MEMBERED SUPERFAMILIES

[1]Sachinkumar Kanti, [1]Shambhu, M. G. and *[2]Anshul Sukhwal

[1]The Oxford College of Engineering, Hosur Road, Bommanahalli, Bangalore – 560068, Karnataka, India
[2]National Centre for Biological Sciences, GKVK, Bellary Road, Bangalore – 560065, Karnataka, India

ARTICLE INFO

Article History:
Received 21st April, 2014
Received in revised form
20th May, 2014
Accepted 10th June, 2014
Published online 20th July, 2014

Key words:
Protein-Protein Interactions,
Oligomerisation State,
Molecular Docking,
Homology Modeling,
Protein Hotspots.

ABSTRACT

Protein-Protein interactions play a major role in carrying out biological functions including enzymatic catalysis, ion transportation and cellular regulation. Systems level understanding of cellular functions uncovers all functional interactions of proteins. It is, thus, important to have resources which make available structural details about these interactions. In present study, a new database called PPI database has been developed which contains information about oligomerisation status, interacting partners and closest homologues for all the structural domains of Single-Membered and Two-Membered Superfamilies of the already existing in-house database, PASS2. Residues present at the protein-protein interface have been analyzed and the ones which provide stability to the complex, i.e. hotspot residues, have been predicted and recorded.

INTRODUCTION

Proteins perform variety of functions such as enzymatic catalysis, ion transportation, cellular regulation and physiological activities (Richardson, 1981). The classic principle of protein folding is that all the information required for a protein to adopt the current three-dimensional conformation is provided by its amino acids sequence. Two or more proteins interact with each other forming protein-protein interactions (PPIs) to carry out biological functions such as muscle contraction and signal transduction (Pawson *et al*., 2000). PPIs may occur in different types such as homo-oligomers (same proteins) and Hetero-oligomers (different proteins), depending on the type of proteins interacting with one another. These can also be distinguished into transient and permanent interactions based on the time, the two proteins remain bound with one another. Domains of each protein that are grouped together at superfamily level are called as "protein domain super families" and can be classified as Single-Membered Super families (SMS), Two-Membered Superfamilies (TMS) and Multi-Membered Super families (MMS). This classification is based on the number of domains with <40% sequence identity with other domains of the superfamily (Gandhimathi *et al*., 2012). Proteins are said to exist in an oligomeric state if they are composed of multiple subunits of same or different proteins. We collected the oligomerisation status of proteins, since the formation of protein oligomeric state produces increased stability with improved function for the multi-enzyme complexes (Franziska and Susan, 2010). The true oligomerisation status of a protein helps to understand the protein's physiological function. An oligomeric protein represents a significant fraction of cellular proteins. When a protein molecule binds with another biological polymer (protein or nucleic acid) to form a complex, the subset of residues in the interface that account for most of a protein binding free energy are called binding '***hotspots***' (Yanay and Burkhard, 2007). Several experiments have shown that PPIs are critically dependent on just a few residues ("hotspots") at the interface. These are the essential residues that delay interactions if mutated (Mishra 2012). The study of protein-protein interactions has become important for biotechnological and therapeutic reasons. Only few of the residues in protein-protein interfaces (hotspots) are essential for the interaction and these can help us understand the PPIs and could also act as desired drug targets. Even though there are databases available with predicted and experimental results on PPIs, we do not have many databases mainly focusing on structural details of the interacting complexes, their oligomerisation state and homologues. In the present study, we collected the oligomerisation status, interacting partners, and closest homologues for all the structural domains of SMS and TMS domains, from the in-house database, PASS2 (Gandhimathi *et al.,* 2012). All these information have been recorded in a new database called as PPI database. This

Corresponding author: Anshul Sukhwal
National Centre for Biological Sciences, GKVK, Bellary Road, Bangalore – 560065, Karnataka, India.

database, thus, can be used as a resource for gaining structural insights about the proteins before carrying rational drug-design and functional/mutational experiments.

MATERIALS AND METHODS

Oligomerisation status for each domain was identified using PDB (Bernstein *et al.,* 1977), PQS (Henrick *et al.,* 1998) and PISA (Krissinel *et al.,* 2007) web servers, in that order, and information about close homologues were obtained using PIQSI (Jensen *et al.,* 2009) database. When these tools failed to report interactions about a protein domain, STRING database was consulted to gather information about interacting partners with highest confidence results (score of 0.700 confidence and more). We also recorded oligomerisation status and interacting partner information of the closest homologue for all the PASS2 domains in the PPI database. If the interacting partner of a PASS2 SMS or a TMS domain, as obtained using STRING, does not have a structure in the PDB then its structure was modeled using homology modeling (Chothia and Lesk 1986). The two proteins (native PASS2 domain and its interacting partner) were then docked using FRODOCK (Garzon *et al*., 2009) and the best native-like pose was predicted using PPCheck. For the selected best pose, interface residues were identified using PPCheck and 'hotspot' prediction results were obtained using KFC2 server.

Development of PPI Database

The in-house domain database, PASS2 contains about 1500 protein domains, collectively in their SMS and TMS. In the present work, the database has been developed by collecting information such as protein/domain name, source organism, domain length, interacting partner, oligomerisation status and close homologues for all the 1500 domain entries of the PASS2 SMS and TMS.

Identification of interacting partners and close homologues

PDB, PQS and PISA servers have been used to collect the information on oligomeric state of all the domain superfamilies. Firstly, information about the oligomerisation status was obtained from PDB. When PDB failed to give information on interacting partners of a domain, we searched into PQS and PISA servers. STRING database was consulted when PQS and PISA failed to give oligomerisation status. Close homologues for all the interacting partners have been identified using PiQSi. Figure 1 depicts the pictorial representation of the proposed work where different bioinformatics tools have been used to develop PPI database.

Modeling of interacting partners

Interacting partners obtained from STRING were considered for homology modeling, using Modeller (Sali, 2009), if their structures were not available in PDB. For each target protein a suitable template was selected and five models with lowest DOPE score were generated. RAMPAGE, a tool based on Ramachandran analysis (Ramachandran *et al*., 1963) was used to validate all the five generated models. Based on DOPE score, lowest energy structure was selected as the best model. In general, a model is considered as a "good model" if the residues in favoured regions are greater than 90% and the number of outliers (residues in unfavourable regions) are less than 4%.

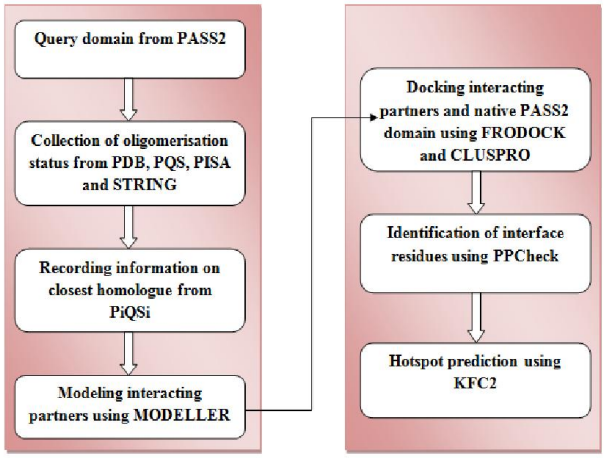

Figure 1. Block diagram describing various steps involved in collection of data from various sources to develop PPI Database

Generating the protein complexes using molecular docking

STRING results with 90% global sequence identity or more from SMS and TMS databases were directly considered for docking as they had a very good query coverage. 62 protein complexes were generated using FRODOCK, whereas 14 complexes were generated using CLUSPRO. In total, 76 docked complexes were generated. PPCheck was applied to obtain the best native-like conformation of the two docked proteins.

Finding the strength of interface residues and hotspot prediction

Best native-like docked complexes obtained from FRODOCK were subjected to PPCheck server (Sukhwal and Sowdhamini, 2013) for the identification of interface residues. These complexes were also subjected to KFC2 server for the prediction of 'hotspots'.

RESULTS

Collection of oligomerisation status

PASS2 has 864 SMS and we could successfully record information about interacting partners from various sources as mentioned above for 821 superfamilies. PASS2 contains 363 TMS wherein each superfamily has two domains, i.e., a total of 726 domain entries were present. Out of 726 entries, interacting partner information was obtained for 710 entries. Interacting partners from STRING were selected based on either 70% sequence identity or 60% identity and 80% similarity with the query domain. These sequence identities and similarities were obtained using EMBOSS Water Pairwise Sequence Alignment (PSA) search tool available at http://www.ebi.ac.uk/Tools/psa/emboss_water/. Out of 864

entries, information on 564 interacting partners were collected from PDB and PQS server and 148 partners from PISA webserver ("interface" module of PISA was consulted, it provides homodimeric coordinates of the protein). 101 interacting partners were obtained from STRING database where few domain entries resulted with two or more interacting partners for each single query domain. 269 close homologues having either more than 70% sequence identity or 60% - 70% identity and more than 80% similarity with the interacting partner were obtained using PiQSi and/or PDB search. Protein name, source organism and interactions for each close homologue were collected and listed. Out of 726 entries in TMS, 108 partners were obtained from STRING. 296 close homologues were also collected for all the interacting partners using PiQSi. Collection of oligomerisation status can be seen in Table-1, and the updated database with close homologues can be seen in Table-2. Figure-2 and Figure-3 gives graphical representation of the oligomerisation status and close homologues collected from PDB, PQS, PISA and STRING for SMS and TMS database, respectively.

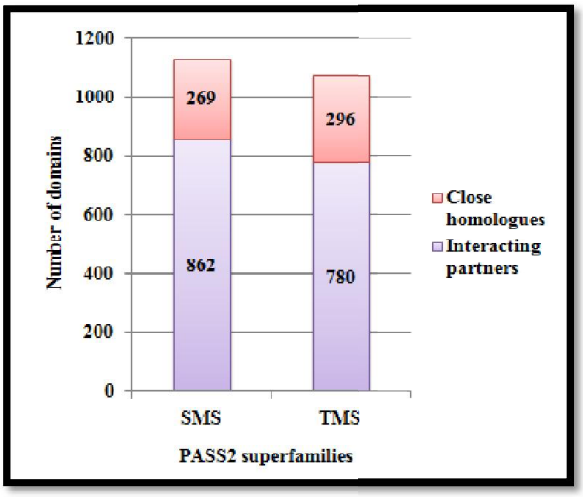

Figure 3. Graphical representation sowing number of interacting partners and close homologues being collected for SMS and TMS

Table 1. Information on oligomerisation status collection from PDB/PQS, PISA and STRING for some of the superfamily domains. Symbol "⇔" is used to represent interactions. Domains are represented by "4-letter PDB code_chain name"

S.No	PASS2 Superfamily	Domain Name	Protein name	Source Organism	Domain Length	Source Database	Oligomerisation Status	Interactions
1	57615	d1e8ra-	endo-1-4-beta-xylanase	Pseudomonas fluorescens	50	STRING	Heterodimer	(1E8R_A) A <=> B (1GQI_A)
2	90234	d1k18a-	dna polymerase alpha catalytic subunit	Synthetic construct	31	STRING	hetero tetrameric	(1K18_A) A <=> B and A <=>D (4BPU_A)
3	55159	d1d1ra-	hypothetical 11.4 kd protein ycih in pyrf-osmb intergenic region	Escherichia coli	83	STRING	Hetero tetrameric	(1D1R_A) A<=> B (1EIX_A)

Table 2. Information on close homologues and interacting partners for some of the superfamily domains. Domains are represented by "4-letter PDB code_chain name"

S.No	PASS2 Superfamily	Domain Name	Protein name	Source Organism	Interacting partner	Close Homologue
1	57615	d1e8ra-	endo-1-4-beta-xylanase	Pseudomonas fluorescens	1GQI_A	1GQJ_A
2	90234	d1k18a-	dna polymerase alpha catalytic subunit	Synthetic construct	4BPU_A	4BPW_A
3	55159	d1d1ra-	hypothetical 11.4 kd protein ycih in pyrf-osmb intergenic region	Escherichia coli	1EIX_A	1JJK_A

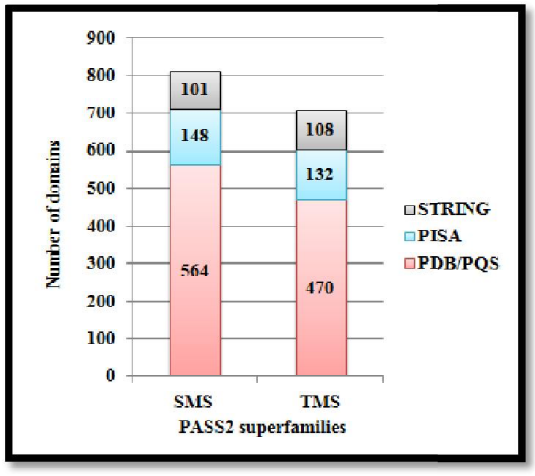

Figure 2. Graphical representation of oligomerisation status explaining the extent of data collection from various sources

Homology modeling approaches

STRING results having 80% to 90% query coverage were considered for homology modeling. MODELLER with version 9.13 was used for modeling STRING results. 10 SMS and 4 TMS STRING results were modeled. In total, 16 STRING results were modeled. Interacting partner of each query domain was treated as template structures and all the STRING results were treated as query sequence individually. Figure-4 shows cartoon representation for few generated model structures as an example, whereas Table-3 shows a list of template and target sequences for modeling and the best model selected based on residues in favoured region and their number as outliers.

Docking studies

32 SMS and 30 TMS domains were docked with their interacting partners using FRODOCK. 14 modeled structures of interacting partners were also docked with their respective

PASS2 native domains. All together 76 domains were docked using FRODOCK. Figure-5 shows few docked complexes with two different protein chains represented in two different colors.

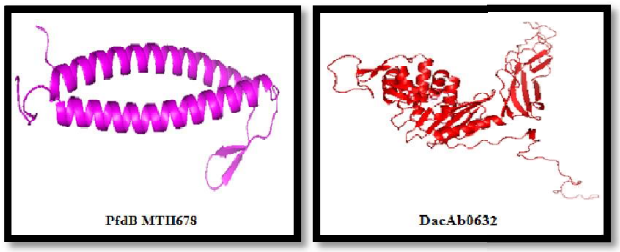

Figure 4. Modeled structures obtained for domains PfdB MTH678 and DacAb0632 as target sequences by considering 1FXK and 1HD8 as template structures

Table 3. Dope scores and RAMPAGE results for some of the modelled proteins. Template structures have been named based on "4-letter PDB code_chain name"

S.No.	Domain Name	Template structure	DOPE Score (kJ/mol)	Residues in Favored region (%)	Residues as outlier (%)
1	d1hy9a	1AX8_A	-16234.29492	94.50	1.20
2	d1rrza	2DHM_A	-8819.046875	97.10	1.00
3	d1hyka	1AX8_A	-17178.16602	93.90	1.80
4	d1pu1a	1FXK_A	-10706.35449	99.20	0.00
5	d1r5sa	1FMK_A	-52474.48828	96.10	1.70

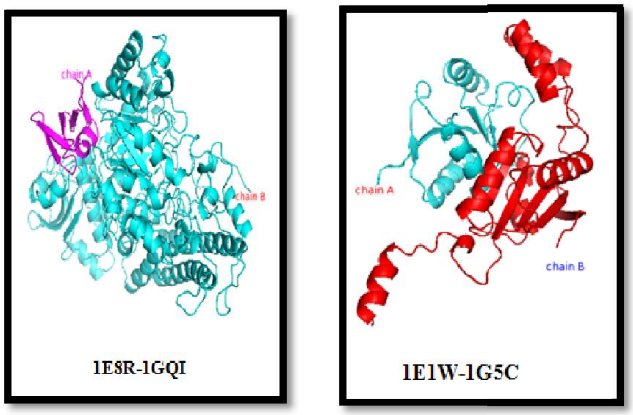

Figure 5. Protein complexes generated by the interaction between receptors 1E8R and 1E1W with the ligands 1GQI and 1G5C

Interface residues analysis

Docked complexes obtained from FRODOCK were analyzed using PPCheck, an in-house webserver for analysis of protein-protein interactions. PDB file with information about two chains of two same or different proteins is given as input to PPCheck. Charged residues, contributing towards electrostatic interactions are identified at distance threshold of 10 Å. van der Waals interactions were identified if atoms of residues from interacting proteins comes within a distance threshold of 7 Å. Hydrogen bonds were calculated using KABSCH and SANDERS equation as used in DSSP program (Kabsch and Sander, 1983). Based on conservation and solvent accessibility, best docking pose were selected for each of the protein complex. Interface residues between receptor-ligand interactions for SMS and TMS domain entries and docking pose were recorded in Table 4.

Hotspot prediction

Best docking pose for each protein complex selected by PPCheck server was fed as input to KFC2 server to predict the hotspots. For each protein complex formed due to interaction between query domain and its interacting partner, two chains were selected. Selected chainset for both domain and its interacting partner were submitted to KFC2 sever to run the job. The results can be seen in Table 5 for some of the SMS and TMS.

DISCUSSION

This is a detailed study wherein a new database called PPI database has been developed which contains information on interacting partners, oligomerisation status and close homologues for all of the PASS2 domains from SMS and TMS. Although many databases are available which provides details about protein-protein interactions, a structural database was needed which would provide informatioin on the above mentioned structural aspects of the protein domains. PPI database also records information about all the residues present at the interface. It also records hotspot residues which provide stability to the complex and also add specificity to the binding sites. PPI database will be helpful to pharmaceutical industries to design drug targets based on the strength of interactions between two proteins. Information on 'hotspot' residues could be used by researchers to perform functional/mutational experiments to add/change the specificity of the binding site of

Table 4. Information about native PASS2 domain and its modelled interacting partner along with number of interface residues in the best docking pose as selected using PPCheck

S.No	Source Database	Domain Name (Receptor)	Interacting Partner (Ligands)	No. of interface residues in best docking pose
1	SMS	d1e8ra-	1GQI_A	116
2	SMS	d1k18a-	4BPU_A	77
3	SMS	d1v9va1	4CFE_A	122
4	TMS	d1d1ra-	1EIX_A	111
5	TMS	d1ny8a-	1A2N_A	139

Table 5. Hotspot prediction results for some selected PASS2 domains as obtained from KFC2 server

S. No	Source Database	Complex Name	Chain 1		Chain 2	
			Chain ID	No. of hotspots	Chain ID	No. of hotspots
1	SMS	1E8R_A-1GQI_A	A	11	B	8
2	SMS	1K18_A-4BPU_A	A	6	B	4
3	SMS	1V9V_A-4CFE_A	A	11	B	8
4	TMS	1D1R_A-1EIX_A	A	8	B	9
5	TMS	1NY8_A-1A2N_A	A	12	B	17

the proteins, thus disrupting the unwanted protein-protein interactions which could potentially lead to disease.

Acknowledgement

We are thankful to Professor R. Sowdhamini for useful discussions and giving us an opportunity to carry out this work in her lab at NCBS, Bangalore and The Oxford College of engineering for providing infrastructure.

REFERENCES

Bernstein *et al.* 1977. The Protein Data Bank: A computer based archival file for macromolecular structures. *Journal of Molecular Biology* 112: 535–542.

Chothia C. and Lesk A M. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO Journal* 5(4): 823–826.

Franziska B and Susan JB 2010. Ribonucleoprotein multimers and their functions. *Biochemistry and Molecular Biology* 45(5): 331–350.

Gandhimathi *et al.* 2012. PASS2 version 4: An update to the database of structure-based sequence alignments of structural domain superfamilies. *Nucleic Acids Research* 40: 531–534

Garzon JI *et al*. 2009. FRODOCK: a new approach for fast rotational protein-protein docking. *Bioinformatics* (25): 2544-2551.

Henrick *et al.* 1998. PQS: A Protein Quaternary Structure file server. *Trends in Biochemical Sciences* 23(9): 358–361.

Jensen *et al.* 2009. "STRING 8--a global view on proteins and their functional interactions in 630 organisms". *Nucleic Acids Research* 37: D412–D416.

Kabsch W and Sander C 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22 (12): 2577–637.

Krissinel *et al.* 2007. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 372: 774–797

Mishra S 2012. Computational prediction of protein-protein complexes. BMC Res Notes 5:495.

Pawson *et al*. 2000. "Protein-protein interactions define specificity in signal transduction". *Genes & development* 14 (9): 1027–47.

Ramachanran GN *et al.* 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular biology* 7: 95-99.

Richardson JS 1981. "The anatomy and taxonomy of protein structure". *Advances in Protein Chemistry* 34: 167–339.

Sali A. *et al*. 2009. Modeller: A Program for Protein Structure Modeling.

Sukhwal A and Sowdhamini R. 2013. oligomerisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. *Molecular BioSystems* 9: 1652-1661.

Yanay O and Burkhard R. 2007. Protein–Protein Interaction Hotspots Carved into Sequences. *Plos computational biology* 3(7): e119.

****