Research Journal of Chemical Sciences _____ ISSN 2231-606X
Vol. **2(7),** 36-40, July **(2012)** *Res.J.Chem. Sci.*

Characterization of Protein Interfaces to Infer Protein-Protein Interaction

Mishra Subhra*
*Department of Chemistry, Alipurduar College, North Bengal University, Pin-736122, West Bengal, INDIA

Available online at: www.isca.in
Received 7th April 2012, revised 20th April 2012, accepted 23th April 2012

Abstract

Understanding of interaction of two key macromolecular species is one of the major problems in structural and molecular biology. An understanding of protein – protein interactions depend upon knowledge of both the three dimensional structural details of the interactions and the chemical dynamics of the systems. Here we present an analysis of several dimeric, trimeric and tetrameric obligatory complexes available in the PDB with homologous sequences filtered out at 70% sequence identity. In this study, oligomeric protein structures are viewed from a network perspective to obtain new insights into protein association. The aim of this paper is to describe the computational approach to design the strategies to recognize the protein–protein interfaces in an automated, generalizable fashion. The successes suggest that these computational methods can be used to modulate, reengineer and design protein–protein interaction networks in living cells.

Key words: Interface, complex, planarity, macromolecule, residue.

Introduction

Protein–protein interactions are central to many processes within cells and organisms, ranging from immune defense to cellular communication. For biological regulation, it is necessary to recognize their targets, and the networks responsible for interactions in macromolecular complexes[1]. Tools to alter and interfere with protein interactions offer great promise to help understand and delineate these networks. So, it is important to know the three dimensional structure of the protein molecules as well as the protein-protein interface. But, the limited nature of Protein Data Bank, and further limited number of X-ray crystallographic structures of high resolution has been a major constraint in the previous studies. Recently, however, there has been a large increase in the number of known three-dimensional structures that contain protein-protein recognition sites and more high-resolution structures have been solved. These structures cover a much broader range of activities than the earlier ones, which were almost exclusively protease inhibitor and antibody antigen complexes. The knowledge of those few structures guided us to determine the rules for general structural study.

The effect of various physical and chemical parameters on the strength of the interaction can be determined by finding their correlation with the energy of complexation[2]. So finding the linear correlation between the different structural and chemical parameters can lead to the determination of those parameters, which play an important role in the determining the strength of the interaction.

We first briefly outline general principles of computational design, with an emphasis on challenges encountered particularly in protein interfaces. We then describe certain features of new protein–protein complexes. These results highlight the features of molecular interactions that can and cannot be modeled using current computational approaches and illustrate the potential of the methodology for the redesign of protein interactions in the context of living cells.

Computational approach towards protein-protein docking: There are two parts to the docking problem: developing a scoring function/energy function that can discriminate correctly or near-correctly docked orientations from incorrectly docked ones, and developing a search method that will be able to `find' a near-correctly docked orientation with reasonable likelihood[3]. To use this, it is necessary to describe the surface shape of the protein. This may be done by discrediting the molecule onto a grid in space and considering which cells are occupied, or by using some sort of 'surfacing algorithm', which calculates the solvent-accessible or solvent-excluded surface, and a point set that triangulates it[4]. In carrying out this calculation, many special cases of geometry need to be considered.

The role of electrostatics in protein–protein interactions has been reviewed by Sheinerman[5], and was explored from a more physical point of view by Elcock[6]. To treat the desolation of charged groups in the interfaces accurately, it is necessary to solve the full Poisson–Boltzmann equation for each different orientation of the components that is to be examined.

In practice, then, the above considerations frequently lead to a two- or three-stage approach to docking, as outlined in. One begins by treating the proteins as rigid bodies, perhaps with some surface softness, searching the comparatively small (six-dimensional) space of relative protein orientations (translational and rotational) and identifying a set of candidate structures using some simple scoring function, with shape

complementarities playing a major role. Then these structures are re-scored using a more expensive energy function that is better at discriminating near-native orientations. In the third stage, we deal explicitly with a model in full atomic detail and allow movement of the side chains and possibly backbone, minimizing an energy function. The second and third stages may be combined. The energy/score landscape is rough and so it is clearly desirable to make the search as effective as possible by the use of efficient optimization algorithms[7]. If extra biological information about the location of the interface is available, it can also be used as early as possible to simplify the search. Many of these considerations apply to methods for docking small-molecule ligands to proteins and any developments will be mentioned if they may be relevant to protein–protein docking[8].

Structural parameters characterizing a protein-protein interface: There are several parameters which can characterize a protein-protein interface like interface area, planarity, secondary structure, hydrogen bonds and hydrophobic and polar composition of the residues in the interface etc. The exposure of protein atoms to solvent can be obtained by calculating the surface area of atoms in contact with solvent molecule. The solvent accessible surface area is calculated by using Lee and Richards Algorithm[9]. The planarity of the interfaces is analysed by calculating the best fit plane through the 3-dimensional co-ordinates of the atoms in the interface using principal component analysis. The classification of secondary structure is based on the percentage frequency of alpha and beta secondary structures in the interface residues. The secondary structure composition of these segments was analyzed and was grouped into three different groups as: alpha (>80% alpha helix), beta (>80% beta sheet), coil (80 % coils) and alpha/beta.

Interface residues are defined as those residues that possess an accessible surface area (ASA) that decreases by >1 angstrom squared on complexation. The 1 angstrom squared was used to take account of the small errors in crystallographic coordinates and computational inaccuracies in the calculation of the ASAs. It has been often been assumed that proteins associate with their hydrophobic patches but polar interaction between the interface is also important[10]. It is therefore to explore the relative composition of polar and non polar residues on the interface. The interface residue propensity is an indication of a particular residue to be in an interface.

Material and Methods

Method uses the statistics of residue-residue contacts across the interfaces of complexes in the PDB, expressing how much more probable it was that residues would interact than would be expected merely from random contacts between residues with the observed global frequencies of occurrence.

The analysis has been carried out by selecting 86 dimeric, 17 trimeric and 52 tetrameric obligatory complexes available in the PDB with homologous sequences filtered out at 70% sequence identity**.** The *SEARCHFIELD* customizable form of the Brookhaven Protein Data Bank was used for the initial PDB mining. The proteins the structures of which have been predicted by methods other than X-Ray crystallography were filtered out. The filtering resolution of crystal structure was taken to be 2 A° for dimeric and tetrameric complexes and 2.5 A° for trimeric complexes.

Further selection has been done by selecting those protein complexes whose biologically active multimeric composition was similar to the multimeric composition present in the ASU (Assymetric Unit), which was given by the structure obtained from the PDB.

Results and Discussion

The accessible surface area (ΔASA) may be taken as the measure of the binding strength of two interfaces[11]. For this the linear correlation between interface area and energy of complexation was calculated for the entire data set.

For both dimers and tetramers the data set was wide spread with the ΔASAs ranging from 7160 A^2 to 14 A^2 incase of dimers, and from 3756 A^2 to 33.5 A^2 incase of tetramers. The set for trimers was more limited with ΔASAs ranging from 2046.55 A^2 to 518 A^2. The ΔASA may be taken as the measure of the binding strength of two interfaces.

The average interface area of dimers (single interface) came to be 2093.55 A^2. This was approximately twice that of the average interface area for trimers (double interface), which came out to be 997.8 A^2. This shows that the average area allotted for the interface in a protein surface has a tendency to remain a steady level of around 2000 A^2.

Planarity is taken as the measure of curvature of an interface. It has been a noted that as the interfaces grow larger the surfaces tend to become more curved, that is the value of this RMS deviation increases[12]**.**

This has been also verified from the fact that the average RMS deviation of atoms from the least-square plane incase of dimeric interfaces is higher (3.93) than those of trimeric interfaces (2.52) and tetrameric interfaces (1.64) as mentioned in table-1. This shows that dimeric interfaces are much more curved in nature than trimeric interfaces and tetrameric interfaces[13]. Thus as the surface area becomes larger the patch becomes more curved.

The average number of interacting segments per interface area for trimers (4.133) was lower than that of dimers (6.488) but the ratio of interacting segments in dimers and trimers was lower than that of the ratio of average area in the two**.** This showed that the trimers on an average have more residue segments per ASA than that of dimers.

Table-1
Structural parameter distribution in different multimeric complexes

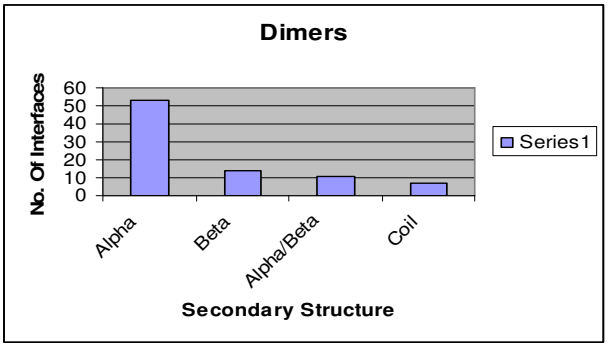
Structural Property		Dimers	Trimers	Tetramers
ASA(A^2)	Mean	2093.952	997.1803	759.998
	StDev.	1385.332	388.4883	893.1581
	Maximum	7167.160	2046.785	3796.69
	Minimum	14.115	518.545	33.5
Energy of complexation	Mean	-28.3375	-34.2333	-51.0533
	StDev.	22.82733	19.87355	41.67648
	Maximum	-149.1	-74.4	-143
	Minimum	10.8	-6.1	-15.8
Planarity	Mean	3.93	2.52	1.644221
	StDev.	0.9445	0.844131	0.791327
	Maximum	10.36	4.365	6.335
	Minimum	1.15	1.565	0.015
Polar Percentage	Mean	37.808	35.21193	37.74721
	StDev.	5.765	4.504961	4.766567
	Maximum	54.307	41.23285	59.34
	Minimum	25.140	28.21075	21.38
Non Polar Percentage	Mean	62.148	64.93424	62.24303
	StDev.	5.73	4.307215	4.782225
	Maximum	74.800	71.7291	78.56
	Minimum	45.653	58.7233	40.61
Hydrogen Bonds/100 A^2	Mean	0.905	0.983547	1.163922
	StDev.	0.450	0.376161	0.342583
	Maximum	2.347	1.670394	15.64
	Minimum	0	0	0

Figure-1 and 2 give the distribution of the secondary structure of the segments for the interfaces considered. It has been found in figure-3 that for homodimeric complex the higher percentage of α helices is present in the interacting zone[14]. It has been found that, most hydrophobic residues with exception of Ala have a high tendency to be in an interface[15]. Amino-acids with aromatic side chains like Phe, Tyr and Trp have a high propensity indicating that aromatic ring interactions may be playing a vital role in the formation of interfaces[16].

Polar amino acids as can be expected have a low interface propensity with the exception of Thr. Cys and Met have a high tendency to be in the interface. Pro showed a high propensity for trimeric than dimeric or tetrameric interfaces. Average percentage polar and nonpolar composition for the three kinds of multimeric complexes did not seems to vary much. Figure-4 and 5 show the variation of polar and nonpolar for the dimeric, trimeric and tetrameric interfaces.

It can also seen from Table 1 that as we go from dimeric to tetrameric interfaces the average number of H-Bonds per 100 A^2 increases from 0.905 to 1.16 as does the energy of complexation. This shows that there is a clear cut relationship between H-bonds and the strength of the interaction and H-bonds play a crucial role in the protein-protein interface contrary to the earlier belief that the protein-protein interactions are primarily driven by the coming together of the hydrophobic patches[17].Such studies on obligatory complexes can be helpful in not only getting about the properties characterizing a strong and permanent interface but, they can also be used to designing novel proteins which carry out their function in the multimeric state. Overall, through analysis of a large set of homomers, we have shown that the evolutionary pathway of a homomer can be inferred from its atomic structure morphology.

The construction and analysis of oligomeric protein structure networks and their comparison with monomeric protein structure networks provide insights into protein association[18]. We believe this analysis will significantly enhance our knowledge of the principles behind protein association and also aid in protein design.



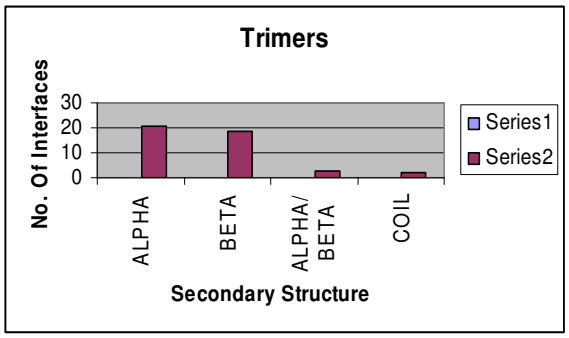


Figure-1
Secondary Structure distribution in Residue Segments in Dimers and Trimers

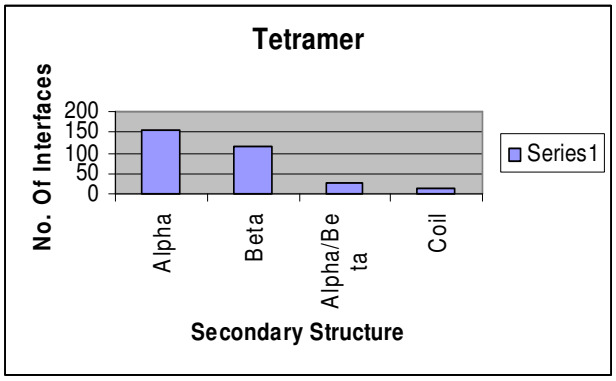


Figure-2
Secondary Structure distribution in Residue Segments in Tetramers

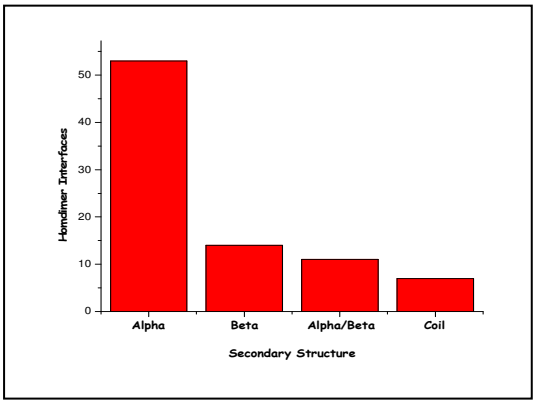
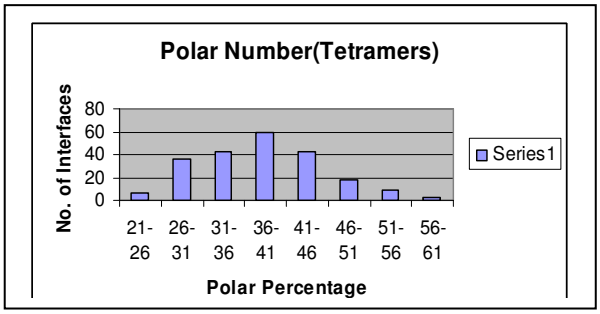


Figure-3
Distribution of secondary structure in the interface



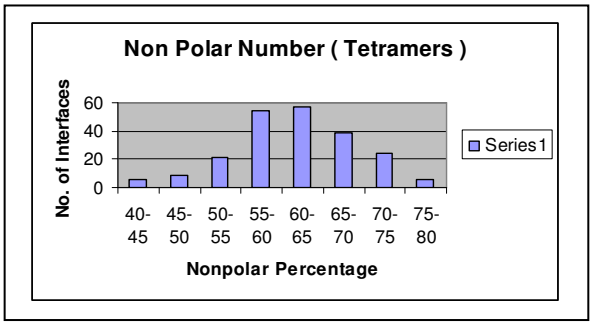
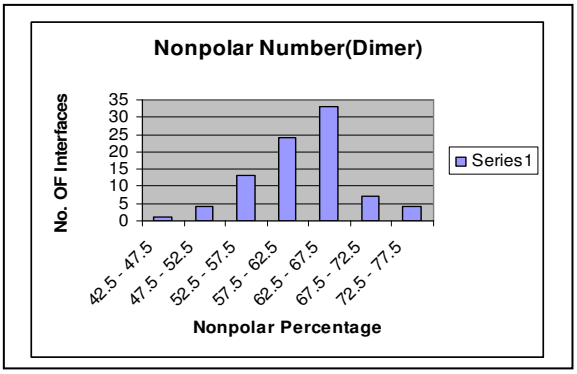


Figure-4
Average percentages of polar and non-polar compositions for tetramers



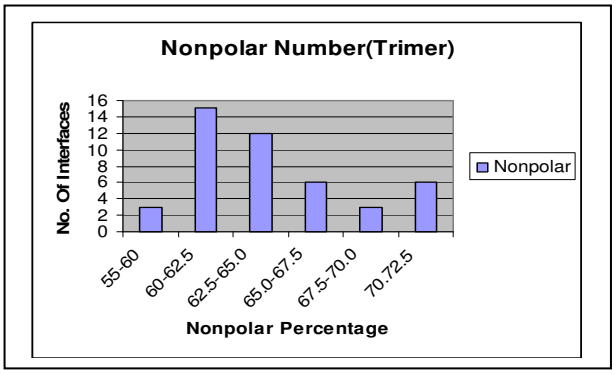


Figure-5
Average percentages of non-polar percentages for dimmers and trimers

Conclusion

Such studies on obligatory complexes can be helpful in not only getting about the properties characterizing a strong and permanent interface but, they can also be used to designing novel proteins which carry out their function in the multimeric state.

References

1. Brinda K.V. and Vishveshwara S., Oligomeric protein structure networks: insights into protein-protein interactions, *BMC Bioinformatics.*, **6**, 296 **(2005)**

2. Blundell T.L. and Srinivasan N., Symmetry, stability, and dynamics of multidomain and multicomponent protein systems, *Proc. Natl Acad. Sci.USA.,* **93,** 14243–14248 **(1996)**

3. Veith M., Hirst J.D., Koilinski A. and Brooks C.L., Assesing energy functios for flexible docking, *J. Comp Chem.,* **19(14)**, 1612-1622 **(1998)**

4. Nooren I.M.A. and Thornton J.M., Structural characterization and functional significance of transient protein-protein interactions, *J Mol Biol.,* **325**, 991–1018 **(2003)**

5. Sheinerman F.B., Norel R. and Honig B., Electrostatic aspects of protein-protein interactions, *Curr Opin Struct Biol.,* **10(2)**, 153-9 **(2000)**

6. Elcock A.H., Gabdoulline R.R., Wade R.C. and McCammon J.A., Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin, *J Mol Biol.*, **291(1)**, 149-62 **(1999)**

7. Bahadur R.P., Rodier F. and Janin J., A dissection of the protein-protein interfaces in icosahedral virus capsids. *J. Mol. Biol.,* **367**, 574–590 **(2007)**

8. Jones S. and Thornton J.M., Principles of protein–protein interactions, *Proc. Natl Acad. Sci*., **93**, 13–20 **(1996)**

9. Lee B. and Richards F.M., The interpretation of protein structures: Estimation of static accessibility, *J. Mol. Biol.,* **55**, 379-400 **(1971)**

10. Del Sol A., Fujihashi H. and O'meara P., Topology of small-world networks of protein-protein complex structures, *Bioinformatics.*, **21,** 1311–5 **(2005)**

11. Miller S., Lesk A.M., Janin J. and Chothia C., The accessible surface area and stability of oligomeric proteins. *Nature.,* **328**, 834–836 **(1987)**

12. Levy E.D., Pereira-Leal J.B., Chothia C. and Teichmann S.A., 3D complex: a structural classification of protein complexes, *PLoS Comput. Biol.,* **2**, 155 **(2006)**

13. Jones S. and Thornton J.M., Protein-Protein Interactions: A Review of Protein Dimer Structures, *In Progress in Biophysics and Molecular Biology*., **63**, 31-165 **(1995)**

14. Valdar W.S.J. and Thornton J.M., Protein-protein interfaces: analysis of amino acid conservation in homodimers, *Proteins: Struct Funct Genet*, **42**, 108–124 **(2001)**

15. Ying G., Wang R. and Lai L., Structure based method for analyzing protein-protein interfaces, *J Mol Model*., **10**, 44–54 **(2004)**

16. Lo Conte L., Chothia C. and Janin J., The atomic structure of protein-protein recognition sites, *J Mol Biol.,* **285**, 2177–2198 **(1999)**

17. Lukatsky D.B., Shakhnovich B.E., Mintseris J. and Shakhnovich E.I., Structural similarity enhances interaction propensity of proteins, *J. Mol. Biol.,* **365**, 1596–1606 **(2007)**

18. Bhatt T.K., In-Silico Structure Determination of Protein Falstatin from Malaria Parasite *Plasmodium Falciparum*, *Research Journal of Recent Sciences,* **1(4),** 68-71 **(2012)**