International Journal for Computational Biology (IJCB)
Vol.1, No.1, 2012,Page. 37-42
ISSN: 2278-8115

iFace: A Bioinformatics Tool for the Analysis of Protein-Protein Interface

Ganesan Pugalenthi[1*], Govindaraju Archunan[2], Thennarasu Sugumar[1]

[1]Bioscience Core Lab, King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia
[2]Center for Pheromone Technology, Department of Animal Science, Bharathidasan University, Tiruchirappalli, India.

Article Info

Article history:

Received Jan 1st, 2012
Revised Feb 5th, 2012
Accepted Feb 8th, 2012

Keyword:

Protein-Protein Interactions
Protein-Protein Complex
iFace
Structural Information

ABSTRACT

Detailed knowledge of protein-protein interaction is essential to understand various biochemical and biological functions. In this paper, we present a bioinformatics tool to analyze the protein-protein interfaces using three-dimensional structural information. iFace identifies protein-protein interaction sites and various interactions that contribute to the specificity and strength of the protein complex.

Corresponding Author:

Ganesan Pugalenthi,
Bioscience Core Lab, King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia
pugalenthi.ganesan@kaust.edu.sa

How to Cite:

Pugalenthi Ganesan *et. al.* iFace: A Bioinformatics Tool for the Analysis of Protein-Protein Interface. IJCB. 2012; Volume 1 (Issue 1): Page 37-42.

1. INTRODUCTION

Communication between proteins is vital for most biological process such as signaling pathways, respiration, enzyme regulation, receptor binding, the immune response etc. [1-3]. Protein-protein interaction is one of the important mechanisms that establish protein-protein communication in cellular responses at various levels. Therefore, detailed knowledge of protein-protein interaction is essential for the better understanding of biological pathways and their impact on various diseases [4-6].

Protein-protein interaction is governed by various intermolecular forces including hydrogen bonds, disulfide bonds, hydrophobic interaction, Vander Waals forces etc. [7]. Detailed knowledge of interface residues and intermolecular interactions is required to understand protein function, protein stability and protein evolution [8,9]. Several algorithms and databases have been reported for the analysis of protein interfaces. Some of the widely used online resources are Protein Interface Calculator (PIC), InterProSurf, hotPoint, KFC server, ProtorP, NCI, DIAL, PISA, HORI etc [10-24].

Although there have been many detailed studies of protein-protein interaction, it still remains as one of the most challenging problem in bioinformatics and molecular biology. Most of the available tools are restricted to the identification of interface residues and possible interactions at the interface. The detailed knowledge of interaction strength and the role of neighboring residues of each interface residue is important to understand the protein-protein mechanism. In this paper, we present a computational tool, iFace, which identifies interface residues, various interactions at the interface, estimated strength of the interaction and spatial neighbors of each interface residue.

2. MATERIALS AND METHODS

2.1 Steps involved in iFace analysis

Given the coordinate set of 3D structure of a protein complex, iFace identifies protein-protein interaction sites, spatial neighbors of interface residues and computes various interactions such as hydrophobic interaction, hydrogen bonds, disulfide bonds and salt bridges. Various steps involved in iFace analysis are shown in Figure 1.

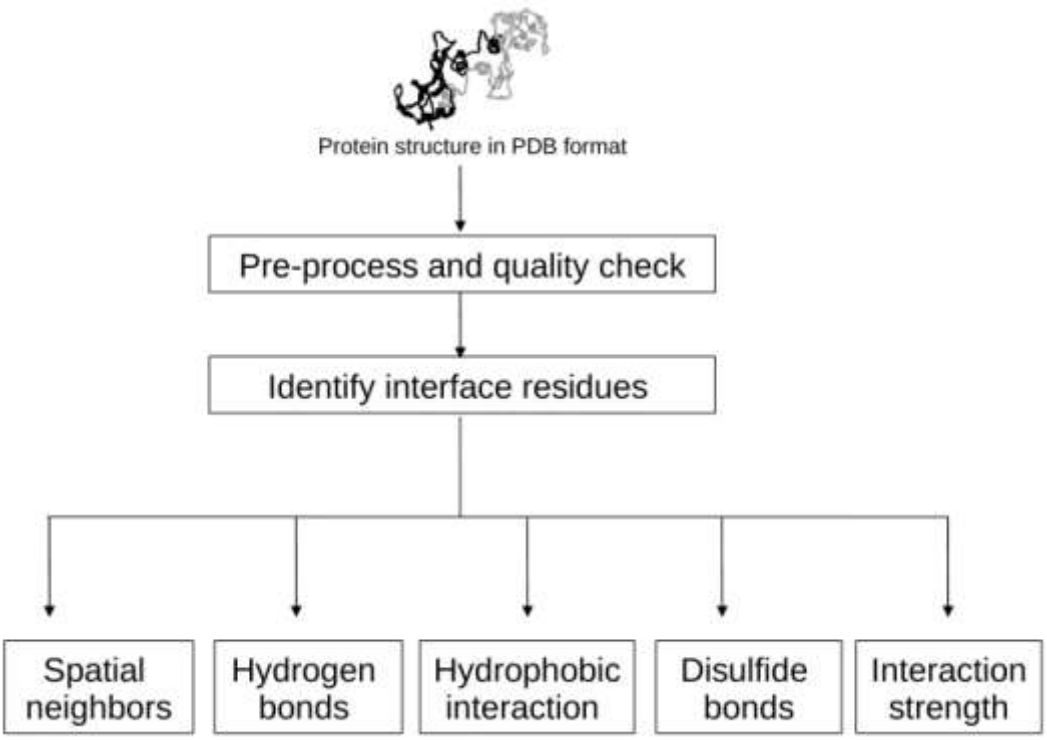

Fig.1: Steps to iFace Analysis

2.2 Identification of interface residues

For each protein complex, we identify interface residues and their spatial neighbors. Interface residues are defined as in Tsai *et al.* [25]. Briefly, two residues are considered to be in contact across the interface if there is at least a pair of atoms, one from each residue, at a distance smaller than the sum of their vander Waals radii plus a threshold of 0.5 Angstrom

2.3 Identification of spatially neighboring residues

Spatially neighboring residues serve both structural and functional roles in proteins. Spatially neighboring residues were shown to have positive influence in identification of critical sites in proteins [26-28]. iFace provides a list of spatially neighboring residues for each interface residue. The residues whose Cβ atoms are found within 5 Å distance from the Cβ of an interface residue are considered as spatial neighbors of the interface residue.

2.4 Identification of interactions at the interface

iFace program identifies the following interactions at the interface: hydrophobic interactions, hydrogen bonds, disulfide bonds and salt bridges. Interaction between hydrophobic side chains are identified using a distance cutoff of 5 Angstrom between apolar groups in the apolar side chains [18]. The hydrogen bonds formed between subunits are identified using HBOND program which is a part of JOY suite [29]. The hydrogen bonds are categorized into four classes: (i) main chain to main chain (MM); (ii) side chain to main chain amide (SN);

(iii) side chain to main chain carbonyl (SO); (iv) side chain to side chain (SS). Disulphide bonds are recognized using the distance criteria employed originally in the MODIP program [30].

2.5 Strength of interaction

The strength of interaction is estimated as in *Biro 2005* [31]. *Biro* constructed amino acid interaction matrices to characterize the expected strength of interaction of two amino acids using three major physico-chemical properties: size, charge and hydrophobicity [31]. Each matrix contains 20 × 20 values for 20 amino acids and each value ranges from 1 to 20, where 1 is the lowest and 20 is the highest probability that two amino acids will interact with each other on the basis of a given physico-chemical property. We used hydropathy compatibility index (HCI) and charge compatibility index (CCI) to characterize the strength of interaction between two interface residues. HCI and CCI are calculated using the following formulas.

$$HCI = 20 - | [HM(A) - HM(B)] \times 19/10.6] |$$

where HM(A) and HM(B) are the hydrophobic moments of the amino acids A and B and HM(Arg)-HM (Ile) = 10.6. This formula gives the maximal index (20) for identical amino acids (closest hydrophobicity) and the minimal value (1) for the two hydrophobically most distant amino acids (Arg and Ile). The "|" indicate absolute values.

$$CCI = 11 - [pI(A)-7] [pI(B)-7] \times 19/33.8$$

where pI(A) and pI(B) are the isoelectric points of the amino acids A and B. This formula gives an index between 1 and 20. The lowest index indicates the lowest possible attraction between amino acids (Asp-Asp) while the highest index indicates the highest possible attraction between amino acids (Arg-Asp).

3. RESULTS

3.1 Input and output

Running iFace program is a straight forward procedure. It accepts input as a protein complex in PDB format. Prior to prediction, structures undergo a set of quality checks. If atoms present alternative locations or rotamers, only the first occurring rotamer is kept. If the PDB structure is NMR structure, the first model is considered for the prediction. However, iFace provides option to analyze the interface for all the models.

iFace provides output in a convenient text format which can be parsed by simple scripts. For each complex, iFace provides the following information.

a) List of interface residues
b) List of spatially neighboring residues
c) Residues that participate in the interface interaction
d) Interaction type
e) Estimated strength of interaction in terms of HCI and CCI

3.2 An example: Bovine seminal ribonuclease

Table 1 shows an analysis result for bovine seminal ribonuclease (PDB code 11ba) [32]. This protein contains two identical subunits and each subunit has 124 amino acids. iFace identified 38 interface residues from chain A and 41 interface residues from chain B. As shown Table 1, the protein complex is stabilized by 8 hydrophobic interactions and 32 hydrogen bonds. In addition, iFace identified two disfulfide bonds at the interface.

Table 1.iFace analysis result for bovine seminal ribonuclease (PDB code 11ba)

Protein Chain	Residue Number	Residue	Protein Chain	Residue Number	Residue	Type of Interaction	HCI	CCI
A	8	PHE	B	108	VAL	Hydrophobic	19.80	10.16
A	9	GLU	B	33	ARG	Hbond-SO	20.00	19.12
A	10	ARG	B	33	ARG	Hbond-SO	20.00	2.88
A	11	GLN	B	44	ASN	Hbond-SO	20.00	9.83
A	12	HIS	B	45	THR	Hbond-SO	20.00	11.37
A	12	HIS	B	47	VAL	Hbond-MM	18.06	11.34
A	13	MET	B	33	ARG	Hbond-SO	18.85	13.78
A	13	MET	B	51	LEU	Hydrophobic	19.25	10.27
A	13	MET	B	54	VAL	Hydrophobic	19.21	10.27
A	14	ASP	B	25	TYR	Hbond-SS	19.53	7.93
A	14	ASP	B	47	VAL	Hbond-MM	18.06	8.64
A	14	ASP	B	48	HIS	Hbond-SO	20.00	12.42
A	15	SER	B	49	GLU	Hbond-SO	20.00	8.22
A	16	GLY	B	48	HIS	Hbond-MM	19.14	11.34
A	16	GLY	B	80	ARG	Hbond-SO	19.14	13.14
A	17	ASN	B	48	HIS	Hbond-SO	20.00	11.54
A	20	SER	B	101	GLN	Hbond-SS	20.00	10.05
A	25	TYR	B	14	ASP	Hbond-SS	19.53	7.93
A	28	LEU	B	28	LEU	Hydrophobic	20.00	10.44
A	29	MET	B	28	LEU	Hydrophobic	19.25	10.27
A	31	CYS	B	32	CYS	Disulfide	20.00	8.97
A	32	CYS	B	31	CYS	Disulfide	20.00	8.97
A	33	ARG	B	9	GLU	Hbond-SO	20.00	19.12
A	33	ARG	B	10	GLU	Hbond-SO	20.00	2.88
A	33	ARG	B	13	MET	Hbond-SO	18.85	13.78
A	44	ASN	B	11	GLN	Hbond-SO	20.00	9.83
A	45	THR	B	12	HIS	Hbond-SO	20.00	11.37
A	47	VAL	B	12	HIS	Hbond-MM	18.06	11.34
A	47	VAL	B	14	ASP	Hbond-MM	18.06	8.64
A	48	HIS	B	14	ASP	Hbond-SO	20.00	12.42
A	48	HIS	B	16	GLY	Hbond-MM	19.14	11.34
A	49	GLU	B	15	SER	Hbond-SO	20.00	8.22
A	51	LEU	B	13	MET	Hydrophobic	19.25	10.27
A	54	VAL	B	8	PHE	Hydrophobic	19.80	10.16
A	54	VAL	B	13	MET	Hydrophobic	19.21	10.27
A	80	ARG	B	16	GLY	Hbond-SO	19.14	13.14

Hbond-MM: main chain to main chain (MM); hbond-SN: side chain to main chain amide; hbond-SO: side chain to main chain carbonyl; hbond-SS: side chain to side chain; HCI- hydropathy compatibility index; CCI - charge compatibility index

4. CONCLUSION

Protein-protein interaction is an important problem due to its role in various pathways, disease studies, protein evolution, protein stability, rational drug development etc. In this paper, we present an efficient and user friendly tool, iFace, for the analysis of protein complexes. iFace program is freely available upon request.

ACKNOWLEDGEMENTS

GP and TS acknowledge Bioscience Core Laboratory, King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia for infrastructural support. GA acknowledges Center for Pheromone Technology, Department of animal science, Bharathidasan University, Tiruchirappalli, Tamilnadu for infrastructural support.

REFERENCES

[1] Jones,S. and Thornton,J.M. (1995) Protein-protein interactions: a review of protein dimer structures. *Prog.Biophys. Mol. Biol.*, 63, 31-65.
[2] Nooren,I.M. and Thornton,J.M. (2003) Diversity of protein-protein interactions. *EMBO J.*, 22, 3486-3492.
[3] Ofran,Y. and Rost,B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, 325, 377-387.
[4] Segal, E. et al. (2003) Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*. 19 Suppl 1:i264-271.
[5] Kann, M.G (2007). Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief Bioinform*. 8(5), 333-346
[6] David, A. et al. (2012). Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum Mutat. 33(2), 359-363
[7] Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U S A.*, 93, 13–20.
[8] Kuttner, Y. Y. et al. (2011) Protein Hot Spots: The Islands of Stability. J Mol Biol. 415(2):419-428.
[9] Bonsor, D.A. and Sundberg, E.J. (2011) Dissecting protein-protein interactions using directed evolution. Biochemistry. 50(13), 2394-2402.
[10] Babu,M.M. (2003) NCI: A server to identify non-canonical interactions in protein structures. *Nucleic Acids Res.*, 31, 3345-3348.
[11] Finn, R.D. (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. Bioinformatics. 21(3):410-412.
[12] Pugalenthi, G. et al. (2005) DIAL: a web-based server for the automatic identification of structural domains in proteins. Nucleic Acids Res. 33(Web Server issue):W130-132.
[13] Murakami, Y. and Jones, S. (2006) SHARP2: protein-protein interaction predictions using patch analysis. Bioinformatics. 22(14), 1794-1795.
[14] Prieto. C. and De Las Rivas, J. (2006) APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res. 34(Web Server issue), W298-302.
[15] Negi,S.S. et al. (2007) InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics*, 23, 3397-3399.
[16] Tina,K.G. et al. (2007) PIC: Protein Interactions Calculator. *Nucleic Acids Res.*, 35, W473-476.
[17] Keskin, O. and Nussinov, R. (2008) Gursoy A. PRISM: protein-protein interaction prediction by structural matching. Methods Mol Biol. 484, 505-521.
[18] Reynolds,C. (2009) ProtorP: a protein-protein interaction analysis server. *Bioinformatics*, 25, 413-414.
[19] Tuncbag,N. et al. (2010) HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Res.*, 38, W402-406.
[20] Shameer, K et al. (2011) 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. Database (Oxford). 2011,bar042.
[21] Shin,Y.C. (2011) TRIP Database: a manually curated database of protein-protein interactions for mammalian TRP channels. Nucleic Acids Res. 39(Database issue), D356-361.
[22] Xiaolei,Z. and Julie,C.M. (2011) KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density and plasticity features. *Proteins*, 79, 2671-2683.
[23] Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol. 372(3), 774-797.
[24] Sundaramurthy, P. et al. (2010) HORI: aweb server to compute Higher Order Residue Interactions in protein structures. BMC Bioinformatics. 11 Suppl 1:S24
[25] Tsai, C.J. Et al. (1996) A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *J Mol Biol*., 260(4), 604-620.
[26] Pugalenthi,G. et al. (2008) Identification of catalytic residues from protein structure using support vector machine with sequence and structural features. *Biochem.Biophys. Res. Commun.* 367, 630-634.
[27] Pugalenthi,G. et al. (2009) Identification of structurally conserved residues of proteins in absence of structural homologs using neural network ensemble. *Bioinformatics*, 25, 204-210.

[28] Pugalenthi,G. et al. (2010) SMpred: a support vector machine approach to identify structural motifs in protein structure without using evolutionary information. *J. Biomol. Struct.Dyn*., 28, 405-414.
[29] Mizuguchi,K. et al. (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, 14, 617-623.
[30] Sowdhamini,R. et al. (1989) Stereochemical modelling of disulfide bridges: Criteria for introduction into proteins by site-directed mutagenesis. *Prot. Engng.*, 3, 95-103.
[31] Biro,J.C. (2006) Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theor. Biol. Med. Model*., 3, 15.
[32] Vitagliano,L. et al. (1998) Binding of a substrate analog to a domain swapping protein: X-ray structure of the complex of bovine seminal ribonuclease with uridylyl(2',5')adenosine. *Protein Sci.*, 7, 1691-1699.