

# PICI: A web server with a multi-parametric algorithm for identifying interaction sites within protein complexes

Arindam Atanu Das & Ramadas Krishna\*

Centre for Bioinformatics, Pondicherry University, Puducherry 605014, India; Ramadas Krishna – Email: ramadaskr@gmail.com; Phone: +91 9489434153; \*Corresponding author

Received February 28, 2016; Revised April 6, 2016; Accepted April 6, 2016; Published April 10, 2016

## Abstract:

Identifying interactions between proteins in a complex is essential to analyze their role in various cellular activities and structural stability. Therefore, effective algorithms and computer programs are required for the better understanding of various interactions exist at the protein-protein interface. The protein inter-chain interaction (PICI) server was developed to identify these interaction sites and various interactions that contribute to the specificity and strength of the protein complex by using Protein Interaction Identification Algorithm (PIIA). The multi-parametric approach of this algorithm involves the interface area between the subunits, the atomic coordinate distance, the linearity of bonds, the orientation of aromatic side-chains, and physicochemical properties of the residues. Particular advantages of PICI include its ability to identify various strong and weak interactions, including those which have not been considered before by any other server. Representation of interface data in text tables, 3D interaction visualizer, colored sequence viewer, and a dynamic colored graphical matrix display makes it distinct from similar web servers. Users can analyze interactions within multi-chain proteins by their PDBids or by uploading a structure atomic coordinate file and can download the result in plain text or XML format.

**Availability:** PICI is freely accessible at <http://pici.bicpu.edu.in/>.

## Background:

Analyzing interactions within protein complexes are crucial to gain a better understanding of their role in signaling pathways, enzyme regulation, receptor binding, antigen recognition, and several other cellular processes, hence are prime targets for drug design. Earlier studies have described the principles governing the interactions and factors that influence the formation of protein complexes [1]. Many papers have discussed different kinds of protein interactions such as disulphide bonds [2], interaction between hydrophobic residues [3], hydrogen bonds [4], electrostatic interactions [5], aromatic-aromatic [6], cation- $\pi$  [7], anion- $\pi$  [8], and atoms having lone pair electrons (N, O, S) to  $\pi$  interactions [9-10]. Along with these interactions several other hydrogen bond weaker than classical hydrogen bonds are also present in proteins, such as C-H $\cdots$ N, C-H $\cdots$ O, C-H $\cdots$ S, C-H $\cdots$  $\pi$ , N-H $\cdots$  $\pi$ , O-H $\cdots$  $\pi$ , and S-H $\cdots$  $\pi$  bonds [4, 11-13]. PPI is largely governed by numerous such weak interactions. Here we report a web server PICI (Protein Inter-Chain Interactions) version 3.2, which calculates and recognizes various kinds of interactions in structural coordinate files of proteins. PICI initially developed

as an integrated tool in PepBind database server [14] to aid analysis of protein-peptide interaction interface. The improved web server includes several programs and scripts to execute its function more efficiently across large protein-protein interfaces.

## Architecture and Data Processing

PICI computes all strong and weak interactions using atomic coordinate data structure files (PDB format) with various distances and angle parameters. The data flow and processing of the web server pass through four stages: (i) Input, (ii) Pre-processing, (iii) Processing with PIIA, and (iv) Output (Figure 1). For calculation of interface area, PICI processes each chain within the multi-chain protein files separately using NACCESS [15] program with 1.4Å probe (radius of a water molecule). The REDUCE [16] algorithm is used to precise the position of hydrogen atoms in the structure file. The hydrogen added file is then parsed for the atomic coordinate positions and processed in PIIA program to identify the interactions.

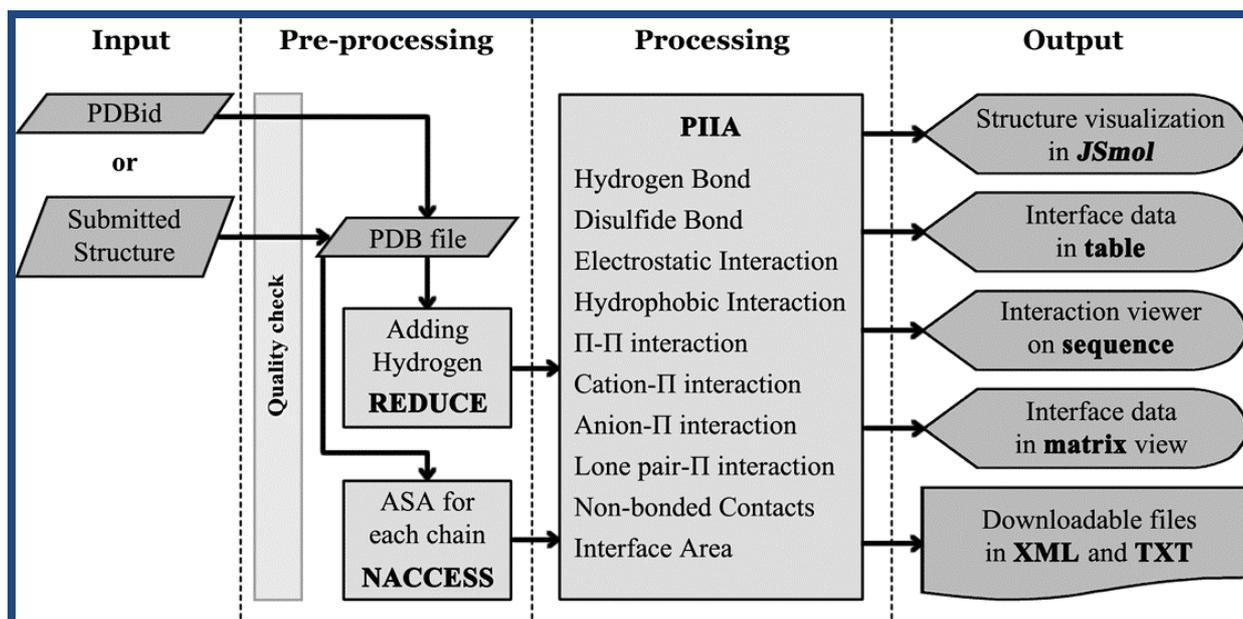


Figure 1: Dataflow diagram of PCI server.

Type	Chain Selected					Chain Other					Distance D <sub>d-a</sub> Å	Distance D <sub>h-a</sub> Å	Angle ∠ <sub>d-h...a</sub>	H-bond type
	ch	res	pos	d/a	atom	ch	res	pos	d/a	atom				
SS	X	THR	102	d	OG1	A	GLU	49	a	OE1	2.83	2.01	162.49	OH...O
SM	X	ARG	104	d	NH1	A	GLU	45	a	O	3.27	2.58	137.68	NH...O
SS	X	MET	109	a	SD	A	LYS	55	d	NZ	2.75	2.23	116.53	NH...S
SM	X	LYS	111	d	NZ	A	ALA	83	a	N	3.10	2.37	138.68	NH...N

Figure 2: Snapshots of the PCI server result viewer page depicting the interface between the two subunits in a protein (PDBid: 1DT7). The top panel shows buttons for all interaction types along with download links for XML and TXT files: (a) Screenshot of interaction result page showing three sections; 3D interface visualization with JSmol, interaction information in a table, and highlighted interacting residues on sequence; (b) Screenshot of the matrix for all interactions, excluding non-bonded contacts with colored compartments for each type of interactions.

**Table 1:** PIIA default parameters

Interactions	Distance cut off in Å	References
<b>H-Bond</b>		
DH...A (D=N, O, C; A=N,O)	$d_{D-A} \leq 3.5$	[4]
SH...A (A=O, N)	$d_{D-A} \leq 4.3$	[13]
DH...S (D=O, N, C)	$d_{D-A} \leq 4.1$	[13]
SH...S	$d_{D-A} \leq 4.5$	[13]
DH... $\pi$ (D = O, N)	$d_{D-A} \leq 4.3$	[12]
CH... $\pi$	$d_{D-A} \leq 4.5$	[11]
SH... $\pi$	$d_{D-A} \leq 4.5$	[13]
<b>Others</b>		
Disulfide Bridge	$d_{S-S} \leq 2.4 \text{ \AA}$	[2]
Salt Bridge	$d_{(+)-(-)} \leq 4.0 \text{ \AA}$	[5]
Hydrophobic interaction	$d_{HPHOB-HPHOB} \leq 5.0 \text{ \AA}$	[3]
Non-bonded contacts	$d_{ATOM-ATOM} \leq 3.9 \text{ \AA}$	[17]
$\pi$ - $\pi$	$d_{\pi C-\pi C} \leq 7.0 \text{ \AA}$	[6]
Cation- $\pi$	$d_{(+)-\pi C} \leq 6.0 \text{ \AA}$	[7]
Anion- $\pi$	$d_{(-)-\pi C} \leq 6.0 \text{ \AA}$	[8]
Lone pair - $\pi$		
	N- $\pi$ $d_{N-\pi C} \leq 4.0 \text{ \AA}$	[9]
	O- $\pi$ $d_{O-\pi C} \leq 4.0 \text{ \AA}$	[9]
	S- $\pi$ $d_{S-\pi C} \leq 6.0 \text{ \AA}$	[10]

Note: As per IUPAC recommendations,  $\angle_{D-H-A} \geq 110^\circ$  is set for H-bonds; D = Donor; A = Acceptor;  $\pi C$  = centroid of  $\pi$ -ring.

### PIIA (Protein Interaction Identification Algorithm)

PIIA calculates nine different types of interactions between neighboring chains based on (i) the interface area, (ii) the atomic coordinate distance, (iii) the linearity of bonds, (iv) the orientation of aromatic side chains, and (v) the physicochemical properties of the residues. The integration of these methods enables PICI to offer efficient identification of location and types of interactions with greater reliability of results. Default cutoff distance and angle for calculations are shown in **Table 1**. Users can also define their preferred parameters for each type of interactions.

### PIIA considers the following criteria for calculation

(a) For NMR structures, only the first model is considered; (b) Cysteine gamma sulfur atoms which participate in disulfide bridges are not considered as proton donors while calculating hydrogen bonds; (c) Backbone nitrogen atoms (except terminal  $NH_2$ ) are excluded from proton acceptors due to delocalization of lone pair electrons with the neighboring carbonyl group; (d) Hydrophobicities of residues are based on solvent transfer free energies from octanol to water [3]. (e) Both 6 and 5 member rings of tryptophan side-chain are considered as separate  $\pi$ -rings for aromatic interactions; (f) Histidine is considered to be in neutral form and side chain imidazole as an aromatic motif; (g) The ammonium nitrogen (NZ) in Lysine and guanidinium carbon (CZ) in Arginine are considered as cationic charge centers for cation- $\pi$  interaction [7].

### PICI Web Interface

PICI web interface is developed with a set of PHP scripts and JavaScripts to display the generated data interactively by responsive HTML web pages with CSS, which work on all major web browsers. PICI takes a protein structure file in PDB format as input. After pre-processing and processing the input data, it displays the output in an interactive result webpage. On the result viewer page, each button on the top panel corresponds to specific interactions and shares a common web page composed of three panels (structure viewer, sequence viewer, and simple text tables) to facilitate the analysis of the information (**Figure 2a**). PICI also generates two output formats: TXT (plain text) and XML (Extensible Markup Language) files to be downloaded by users.

### 3D interaction viewer panel

The structure visualization panel dynamically displays interacting residues in the JSmol structure viewer which is compatible with all browser types. It allows users to visualize atomic details of all interacting residues at the interface.

### Sequence viewer panel

Amino acid sequences of the multi-chain protein are shown and residues are highlighted for each interaction. The single letter code residues are colored by their physicochemical properties such as green for hydrophobic, yellow for neutral, red for negatively charged, and blue for positively charged residues. The interacting residues in the sequence are dynamically highlighted for each type of interaction. BLAST links are generated for each sequence to perform sequence similarity search on PDB or UniProt database.

### Interaction matrix

PICI also generates a dynamic matrix of interacting residues in the selected chain to other chains with colored compartments (**Figure 2b**). Each color represents an interaction type, such as yellow for hydrogen bond, purple for disulfide bridge, red for the salt bridge, green for hydrophobic, blue for  $\pi$ - $\pi$ , violet for cation- $\pi$ , pink for anion- $\pi$ , orange for lone pair- $\pi$  interactions, and gray for non-bonded contacts. The compartments, which show multiple interactions, are represented by multiple colors. The matrix representation of interaction provides an overall view of the protein-protein interface graphically.

### Discussion:

PICI web-server will serve as a useful resource for the structural bioinformatics and the proteomics community to perform analysis of protein structures based inter-chain interactions. Although there are some existing web servers doing similar jobs such as PIC [18], PDBsum (Protein-protein interface) [19] and PISA (Interface) [20] for protein interactions, they target some specific research area and have their unique data representations. PICI server is unique in calculating non-classical hydrogen bonds, anion- $\pi$ , and  $L_p$ - $\pi$  interactions, which are being ignored by other existing protein interaction algorithms. PDBsum and PISA show only donor-acceptor distances for hydrogen bonds without showing their bond angles. Although PIC calculates the hydrogen bond angle, it does not consider bond linearity as a criterion for a hydrogen bond. Also, PIC server does not calculate the interface area and

non-bonded contacts between residues in close proximity. The uniqueness of PDBsum's protein-protein interface program lies in its schematic colored diagram of inter-chain interactions. PISA interface is unique in its ability of deep analyses of PPI interface and informative colored representation of the interface data table. A newly developed PIMA server [21] is unique in data presentation with a dot plot for interaction energies, but it is also limited to few interaction types. PICI's PIIA program is unique in analyzing protein-protein interactions based on multiple structural and physicochemical parameters. Data representation in PICI is unique in presenting the interaction data in (i) highlighted interacting residues in a colored sequence viewer, (ii) color coded interaction matrices, (iii) Java independent structure viewer, and (iv) a textual data table. (A comparative analysis of different web servers along with example are given in supplementary documents)

### Acknowledgement:

AAD was supported by Pondicherry University fellowship for Ph.D. scholars. The authors thank to M. Kannan for fruitful discussions on proteins interactions and S. Manikandan for technical help on server implementation.

### References:

- [1] Jones S & Thornton JM, *Proc Natl Acad Sci U S A*. 1996 **93**: 13 [PMID: 8552589]
- [2] Sowdhamini R *et al. Protein Eng.* 1989 **3**: 95 [PMID: 2594728]
- [3] Matthews BW *Encyclopedia of Life Sciences* 2001 (DOI: 10.1038/npg.els.0002975)
- [4] Weiss MS *et al. Trends Biochem Sci.* 2001 **26**: 521 [PMID: 11551776]
- [5] Barlow DJ & Thornton JM, *J Mol Biol.* 1983 **168**: 867 [PMID: 6887253]
- [6] Burley SK & Petsko GA, *Science* 1985 **229**: 23 [PMID: 3892686]
- [7] Gallivan JP & Dougherty DA, *Proc Natl Acad Sci U S A*. 1999 **96**: 9459 [PMID: 10449714]
- [8] Schwans JP *et al. Proc Natl Acad Sci U S A*. 2013 **110**: 11308 [PMID: 23798413]
- [9] Mooibroek TJ *et al. CrystEngComm* 2008 **10**: 1501
- [10] Reid KSC *et al. FEBS Letters* 1985 **190**: 209
- [11] Brandl M *et al. J Mol Biol.* 2001 **307**: 357 [PMID: 11243825]
- [12] Steiner T & Koellner G, *J Mol Biol.* 2001 **305**: 535 [PMID: 11152611]
- [13] Zhou P *et al. Proteins.* 2009 **76**: 151 [PMID: 19089987]
- [14] Das AA *et al. Genomics Proteomics Bioinformatics.* 2013 **11**: 241 [PMID: 23896518]
- [15] <http://www.bioinf.manchester.ac.uk/naccess/> (Comment: Solvent accessible area calculations)
- [16] Word JM *et al. J Mol Biol.* 1999 **285**: 1735 [PMID: 9917408]
- [17] Wallace AC *et al. Protein Eng.* 1995 **8**: 127 [PMID: 7630882]
- [18] Tina KG *et al. Nucleic Acids Res.* 2007 **35**: W373 [PMID: 17584791]
- [19] Laskowski RA, *Nucleic Acids Res.* 2009 **37**: D355 [PMID: 18996896]
- [20] Krissinel E & Henrick K, *J Mol Biol.* 2007 **372**: 774 [PMID: 17681537]
- [21] Mathew OK & Sowdhamini R, *Bioinformatics* 2016 **12**: 9

Edited by P. Kanguane

Citation: Das & Krishna, *Bioinformatics* 12(2): 78-81 (2016)

**License statement:** This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

