A bioinformatics protocol for the identification of spatial clusters and the calculation of higher order residue interactions in protein structures

CURRENT STATUS: POSTED

Sowdhamini Ramanathan
National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore, 560065, India

✉ mini@ncbs.res.in*Corresponding Author*

Sundaramurthy Pandurangan
National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore, 560065, India and Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India

Shameer Khader
National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore, 560065, India

Raashi Sreenivasan
National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore, 560065, India Current Address : University of Wisconsin-Madison, Madison, WI 53706-1481, USA

Sunita Gakkhar
Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India

Introduction

Sundaramurthy Pandurangan, Shameer Khader and Raashi Sreenivasan contributed equally to this work.

Folding of a protein from its sequence to its three-dimensional structure is controlled by an ensemble of local and global interactions. Pairwise, triplet and quadruplet residue interactions play a vital role in attaining the stable conformation of the protein structure. As higher order interactions make significant contribution to the potential energy landscape of folded proteins, it is important to identify them to estimate their contributions to the overall stability of a protein structure. In the current era of high-throughput sequencing, due to huge lacunae in the sequence to structure ratio, computational approaches are playing a significant role in understanding the design principles of protein structures. Web servers are currently used as a stable platform for the analysis of protein structures. We have developed a web server "**HORI : Higher Order Residue Interactions**":http://caps.ncbs.res.in/hori in proteins for the calculation of global and local higher order interaction patterns in protein structures. In this article, we report a detailed bioinformatics protocol to access the HORI Server and to employ this server for the identification of spatial clusters and the calculation of higher order residue interactions in protein structures. HORI Server can be accessed from the URL:

"http://caps.ncbs.res.in/hori":http://caps.ncbs.res.in/hori

Reagents

Background

Structural bioinformatics[1] is a mature sub-domain of bioinformatics due to the support of community efforts to deposit and share macromolecule structural data and tools for analysis. As PDB[2, 3] remains to be the cynosure of the structure-related data storage and retrieval, several useful structure analysis servers play major roles in analysing structures from experimental and modelling approaches. As current web server technology is a convenient medium for analysis of individual protein structures, they influence and accelerate structural bioinformatics research to a great extent. A typical web server requires limited computational resource in the client side to perform computationally intensive tasks to perform through the internet. Web servers are relevant in the

analysis of macromolecule structures: for example, in the latest release of Bioinformatics Links Directory[4], 856 resources are available under the category of proteins that explains the importance and availability of web-based tools for protein sequence and structure analysis. In this report, we are introducing a bioinformatics protocol to recognise spatial clusters and higher order residue interaction calculations using a new web server that we developed called HORI (Higher Order Residue Interactions in proteins)[5]. HORI server can be employed for the calculation of higher order residue interactions in proteins using PDB file as input and can be accessed from the URL: "http://caps.ncbs.res.in/hori":http://caps.ncbs.res.in/hori .

Interpretation of the crucial principles behind the three-dimensional structure and folding of a protein from its sequence still remains as the major challenge in structural bioinformatics. Derivation of a three-dimensional structure of a protein from its primary sequence is controlled by a complex and largely unknown set of principles called as folding code. Folding of a protein into its three-dimensional structure is controlled by both local and global interactions within a protein[6-13]. Extracting such interactions from protein structures are of paramount importance due to its potential applications in protein structure prediction, modelling, protein engineering experiments and to correlate structural and functional aspects of structures. HORI server has been designed as an interactive, multiple module server for the computation of all possible higher order residue interactions in protein structures of single chain and higher oligomer structures with two chains. HORI offers enormous possibilities to understand details about higher order residue interaction within protein structures in atomic level. The basic algorithm of HORI utilises the classical concept of four-body nearest-neighbour propensities of amino-acid residues. It has been proved that higher order residue interactions up to the level of quadruple interactions will play a major role in the three dimensional structure of proteins. According to the earlier studies[14], if we approximate each residue as a sphere centred on its location, it is possible for a maximum of four closely packed spheres to make mutual contact, thus giving rise to pairwise, triplet and quadruple interactions. Just as no more than four same-sized spheres can be in mutual contact in 3D space, higher order interaction beyond quadruple interactions

3

are generally not considered. Graphical representation about the approach used for the calculation of pairwise, triplet and quadruple interactions are explained in Figure 1.

Equipment

1. A computer with access to the internet and a web browser. HORI Server can be accessed using different browsers, but we recommend : Mozilla Firefox (Windows/Linux) and Apple Safari (Mac OS)

Equipment setup

Data

HORI server requires protein structure data in a standard PDB file format for the successful computation of higher order interactions. Several parameters are available as optional input for specific sub modules within the HORI server. The input parameters required for a successful HORI run is provided in the Procedure section.

Programs

HORI is a web server to compute higher order residue interactions in protein structures. HORI (Higher Order Residue Interactions in proteins) server offers **21 programs classified into three modules as HORI "Global, HORI" Lite and HORI – Cluster to calculate pairwise, triplet and quadruple interactions** within the protein structures. The three modules are as follows:

1. **HORI-Global :** complete structure level computation of different higher order residue interactions

2. **HORI-Lite :** specific higher order residue interaction computation based on residue numbers in input PDB files

3. **HORI-Cluster :** higher order residue interaction computation based on residue types in input PDB files HORI-Global offers two programs for computationally intensive calculation of Pairwise, Triplet and Quadruplet interactions in the protein structures. HORI-Lite and HORI-Clusters offers a set of 9 programs for comparatively quick calculations of specific/generic interactions within protein structures. HORI-Global,

4

HORI-Lite and HORI-Cluster computations can be performed using C^{alpha} or C^{beta} based distance calculation methods. Easy-to-use web interfaces are available for all programs in HORI server to calculate using single chain and multi-chain proteins (two chains at a time).

In HORI-Global, it is possible to:

Compute Higher Order Residue Interactions for single chain and multi-chain Proteins (Complete structure based calculation - pairwise, triplet and quadruple Interactions)
Compute Higher Order Residue Interactions for Single Chain and multi-chain Proteins (Residue range based - pairwise, triplet and quadruple Interactions)

In HORI-Lite & HORI-Cluster, it is possible to:

Compute Pairwise Distance for any 2 residues in PDB file (pairwise interactions)
Compute Pairwise Distance around any 1 residues in PDB file (pairwise interactions)
Compute Triplet Distances between any 3 residues in a PDB file (triplet interactions)
Compute Triplet Distances around any 2 residues in a PDB file (triplet interactions)
Compute Triplet Distances around 1 residue (triplet interactions)
Compute Quadruple Distances between any 4 residues in a PDB file (quadruple interactions)
Compute Quadruple Distances around any 3 residues in a PDB file (quadruple interactions)
Compute Quadruple Distances around 2 residues in a PDB file (quadruple Interactions)
Compute Quadruple Distances around 1 residue in a PDB file (quadruple Interactions)

NB: HORI-Lite and HORI-Cluster offers similar set of programs for the computations, but the input options are different. Interaction types are available for each program as it is given in brackets.

Procedure

1. If you would like to identify higher order residue interactions in a PDB file (for example PDB ID: 1CRN), go to HORI server page ["URL":http://caps.ncbs.res.in/hori/hori.html]. As 1CRN is a single chain protein, you can select the link HORI - Global for Single Chain Proteins ["URL":http://caps.ncbs.res.in/hori/hori_global_single_chain.html]. If your protein of interest is a multi chain protein, you can opt the HORI - Global for Multi Chain Proteins ["URL":http://caps.ncbs.res.in/hori/hori_global_multi_chain.html]. HORI-Global calculation can also be performed for a range of residues using options available for range-based calculations. HORI server separate web interfaces for programs in three different modules (HORI-Global, HORI-Lite and HORI-Cluster) single chain and multi-chain proteins. Availability of individual programs in different modules and computation of different types of interactions possible with

individual programs are explained in programs section. Screenshots of web-interfaces of HORI-Global, HORI-Lite and HORI-Cluster is provided in Figure 2.

2. **Enter your valid E-mail ID:** HORI-Global is a computationally intensive module that requires considerable amount of time for computation of complete structure level higher order interactions. Hence we designed this particular module as email based server, where user should provide a valid email ID for the submission and process. Please enter your valid email ID (preferably academic email ID) to receive HORI-Global results. Check your email address for any 'typos' before submitting the HORI process to the server.

3. **Upload PDB File:** HORI server requires a PDB file as its input for the computation of higher order residue interactions. HORI Server provides separate interface for single chain and multi chain proteins. If you would like to submit an NMR structure, please submit Chain IDs of two chains at a time in multi-chain based programs. Due to computational intense nature of HORI calculations, currently the server will accept only two chains for calculation. Users may remove HETATM / ANISOU records from the PDB Files. HORI Server can perform computations on experimentally (X-ray, NMR) solved structures as well as modelled structures using software like MODELLER[15], SWISSMODEL[16, 17], SCWRL[18] in standard PDB file format. Submit either single chain from an NMR based PDB file as input in Single-chain based programs or enter the chain IDs of any two chains in multi-chain based programs. Please remove HETATM/ANISOU records in the input file, HORI server will not consider these records during calculations, only records tagged as ATOM will consider for computation. In HORI-Lite, Please specify the residue numbers present in the given input PDB file for various computations. In HORI-Cluster, Please specify the residue types that are present in the given input PDB file for various computations. In this example, input file "1CRN.pdb" is uploaded. 1CRN indicates the PDB code of a hydrophobic plant seed protein, Crambin at atomic resolution[19].

4. **Select Atom type for Distance Calculations (C^{alpha} or C^{beta}):** HORI Server offers wide array of programs for the computation of higher order residue interactions. All computations in

HORI server can be performed using two distance parameters C^{alpha} or C^{beta} atom type. Atom type for distance calculation is available as an option, so that user can decide the atom type for distance calculation. In this example, Atom type for Distance Calculation is selected as C^{beta}.

5. **Select Interaction Type:** HORI-Global offers different options to compute higher order residues interactions in protein structure. Users can select any of the four options provided in the HORI-Global. The Interaction type options are Pairwise, Triplet, Quadruplet and All. Here 'All' indicates computation using all three interaction types available in HORI-Global. In this example, interaction type is selected as "ALL".

6. **Enter Distance Cut-off:** HORI calculations are distance dependent, users can specify the distance cut-off for the calculations. Users can enter any number between 0-14, it is highly recommended to perform HORI computations within a reliable distance of 7 Angstroms. For example Lower Limit: 0 and Upper Limit: 7. Distance cut-off is one of the major parameters in HORI server to identify residues interacting in higher orders like pairwise, triplet and quadruple. In this example, Distance Cut-off is selected in the range of 1Å to 7Å.

7. **Enter Chain ID:** HORI offers all different computations in single chain and multi-chain options. Multi-chain option will be useful for the users interested to perform the HORI computations on higher oligomers. Due to limitation of computation, current version of HORI multi-chain programs performs computation using only two chains at a time. Users uploading multi-chain (PDB files with 2 or more chains) should enter their chain ids for the successful HORI computation.

8. Click on the "HORI" button to submit the input details once you fill up the required information. If you need to modify filled up information you can reset the page by clicking on "Reset" button.

9. Check for any messages. Depending on the success or failure of input data, different warning messages will be displayed. If you have not given required information, an error message will be displayed, you need to recheck and fill the required information for successful submission. If all required input data is submitted to the HORI Server, you can see the message. Your job has

been successfully uploaded. Following, you can also see the parameters which you have

selected for HORI run.

10. Once your submission is successful, the input details will be processed by server and the results

will be sent to the email address which you have provided. Time required for processing of your

query highly depends on the number of residues in the PDB file and the parameters you have

chosen (please see the section on Time Taken to obtain an indication).

11. Once the input structure is processed with the parameters provided by the user, the results are

generated by the server and an automated email with the results URL will be send to the email

provided by the user in the submission form. Subject of messages from HORI server begins with

the following "HORI Server Results for Query: PDB-ID". In the HORI server results email, a link

will be provided that can be used to access the results.

12. Click on the URL in the email for to access results from HORI server. For the sample file

1CRN.pdb with the parameters explained above a typical HORI output consists of 3 links for

browsing and 3 links for file download. An example output file is available at this URL with the

parameters explained above from this

"URL":http://caps.ncbs.res.in/hori/hori_results/er_hori_SatDec13185917IST2008.html. Details of

the input file and parameters used for the HORI submission are presented in the beginning of

the result page. A typical HORI-Global output page follows that is divided in to three sections as

HORI - Pairwise Interactions, HORI - Triplet Interactions and HORI - Quadruple Interactions. Each

of these sections provides links for browsing the results and visualizes the individual higher

order interactions using RasMol or Jmol. The second link can be used to download higher order

interaction data in a text format for parsing or further distance calculations and analysis. In the

given example link, if user click on the "Show Pairwise - Interactions" link, the section will

expand and results will be displayed in a tabular format. At the same time "Show Pairwise -

Interactions" link will replaced with the "Hide Pairwise - Interactions" link. This link can be used

to hide this list. If you click on the "Show Pairwise - Interactions" it shows a detailed table with

No., Residue No., Interacting Residues, Distance, and links to download a copy of RasMol script file for visualization of individual interactions in client machine and visualization in client browser using Jmol. Here the first "No" column indicates the serial number of interaction. This can be used as a parameter when user wanted to compare total number of different interactions across different members from same protein family. This is followed by "Residue No" column which provides number of residues involved in a particular interaction. Third column is giving details about the "Interacting Residues" in terms of three-letter amino acid code. Fourth column provides "Distance" in terms of Angstroms. Fifth column provides an image of RasMol logo that provides link to download the RasMol script file. For visualization using RasMol[20, 21], user should install RasMol. A RasMol script file can be visualized using command line interface of RasMol. The coordinates of query protein highlighted by HORI can be downloaded and conveniently displayed with standard macromolecule structural visualization software such as RasMol. In order to view a HORI RasMol script file, Please follow the steps:

a. Click the link "Rasmol view"

b. Save the text file. In this example, the file name will be "hori_rasmol_doublet.cgi"

c. Use "rasmol -script hori_rasmol_doublet.cgi" command

Sixth column consists of an image of Jmol logo provides a link to access the dynamic Jmol interface to visualize the interactions in the browser. If user would like to visualize a particular interaction within the browser, based on Java compatibility of the browser, user can visualize a particular interaction within the browser using Jmol[22]. The number of the three columns "Residue No.", "Interacting Residues" and "Distance" in the result page varies between different higher order interactions. For example in pairwise interactions, there is one set of these 3 columns, but in case of triplet interactions these columns are repeated three times to describe a pairwise interaction and for quadruple interaction this columns are repeated 6 times to describe the details of residues involved in a particular quadruple interactions. A screen shot collage with details about different sections of output is provided in Figure 3.

Timing

Time taken varies between different programs in HORI modules and the parameters used to perform the run. Time varies also depend up on the size of the protein structures and load on the server.

Critical Steps

Please refer to Procedure and Troubleshooting sections. Various steps explained in the Procedure & Troubleshooting sections are critical for the successful analysis of protein structures using HORI server.

Troubleshooting

1. For a successful HORI submission, user must submit all the input details for specific operation.

2. If users are interested to understand role of specific residues in active site or other sites of interest, users can perform calculations using programs in HORI-Lite module.

3. If users are interested in performing full-structure level calculations, users can perform such calculations using programs in HORI-Global modules. Depending on the size of protein and the interaction type, the content of the output from HORI server varies and this may cause loading problem in some browsers. In such situations, user can either perform a HORI-Global calculation using "Residue range based programs" or they can download the interactions file or RasMol file and visualize then on local machines.

Anticipated Results

A detailed explanation of results obtained using "1CRN.pdb":http://www.pdb.org/pdb/explore.do?structureId=1CRN as input file in HORI server is explained in Section - 12 of PROCEDURE. HORI server can be utilized to understand interesting higher-order interactions from protein structures. Detailed analysis of different higher order interactions and residues that mediate such interactions using set of protein structures of interest will be providing better insights to understand structural features.

Application of HORI server in protein structure analysis:

Users can utilize the different programs for calculating higher order residue interactions in protein structures in different contexts. For example the higher order interactions can be utilized in the identification of alternate binding sites based on higher order residue interactions, interaction

patterns for mutational studies and protein engineering, analysis of higher order residue interactions of active site residues, to distinguish protein folds through map of higher order interactions. As a general protein structure analysis web server that can perform computation of higher order residue interactions from a single residue level to whole structure. In order to illustrate the usefulness of higher order interactions and to demonstrate where triplet and quadruplet interactions might be crucial, we take the example of four different proteins belong to different superfamily from SCOP database [23, 24].

Example 1: PDB ID: 1NNQ

3D Domain swapping is a structural phenomenon observed in proteins with diverse sequence, structure and function[25-27]. Ruberythrin from *Pyrococcus furiosus* is reported to be involved in 3D domain swapping[28]. We have used the structure (PDB ID: "1NNQ":http://www.pdb.org/pdb/explore.do?structureId=1NNQ) and performed a HORI-Global based computation and retrieved all pair wise interactions. HORI server reported a total of 580 pair wise interactions with in the cut-off of 1Å to 7Å. Out of these pairwise interactions, 13% of interactions are mediated by the residues between the two chains. Such detailed analysis on proteins with peculiar structural features will give a better understanding of the role higher order interactions in orchestrating the structural integrity.

Example 2: PDB ID: 1BF2

Isoamylase from *Pseudomonas amyloderamosa* (PDB ID: "1BF2":http://www.pdb.org/pdb/explore.do?structureId=1BF2) [29] is a hydrolase reported with active site residues ASP, GLU, ASP. We have used these active site residues reported in Catalytic Site Atlas[30] database and calculated the triplet interactions mediated by this active site residues. The link to access the results is provided in Table 1.

Example 3: PDB ID: 1QJ4

Hydroxynitrile lyase is a lyase from *Hevea brasiliensis* (PDB ID: "1QJ4":http://www.pdb.org/pdb/explore.do?structureId=1QJ4) with the classical Ser, His, Asp triad. The active sites are reported in Catalytic Site Atlas based on PSI-BLAST alignment to the PDB ID: 1C4X. We performed the option to compute quadruple distances around any 3 residues. List of

11

quadruple interactions mediated by active site is obtained. This kind of analysis will be useful to identify potential alternate active site and residues suitable for mutational experiments. Results can be accessed from the link provided in Table 1.

Example 4: PDB ID: 1AQ0

The functionally important residues, including at the active site, may not be strictly conserved for protein domains at the level of superfamily[31]. HORI server can be used to search for alternate active site residues based on specific higher order residue interactions. Barley 1,3-1,4-beta-glucanase[32] from Hordeum vulgare belong to the TIM barrel glycosyl hydrolases superfamily. This PDB entry (PDB ID: "1AQ0":http://www.pdb.org/pdb/explore.do?structureId=1AQ0) belongs to the Glycosyl hydrolases family 17 (Pfam ID: "PF00332":http://pfam.sanger.ac.uk/family/PF00332) in Pfam database[33, 34]. We have used three of the four active site residues reported in Catalytic Site Atlas[30] in the option to compute quadruple distances around any three residues in a PDB file and successfully identified the fourth active site residues in a quadruple interaction flexibly.

Links to access HORI-Global results or automated example pages to run the computation of higher order residues interactions mentioned in this protocol is provided in Table 1.

Future work

Currently the higher order residue interaction calculation is only possible through the web interface available from the URL: "http://caps.ncbs.res.in/hori/hori.html":http://caps.ncbs.res.in/hori/hori.html. We will be extending this platform for the programmatic access through a web service implementation of the programs using "REST":http://en.wikipedia.org/wiki/Representational_State_Transfer and "SOAP":http://en.wikipedia.org/wiki/SOAP based technologies.

Scope of protocol based on HORI Server:

Higher order interactions calculated using set of computationally intensive algorithms available in HORI server will be useful in protein structure prediction, protein modelling, protein-protein interaction, active site identification and to understand higher order interaction characteristics of

active site residues within specified distance shells. Knowledge about the higher order interaction will be of great importance in the structural biology domain due to its wide range of applications in fold recognition, structural analysis, protein engineering and to understand mechanism of action of enzymes.

References

1. Altman RB, Dugan JM: Defining bioinformatics and structural bioinformatics. Methods Biochem Anal 2003, 44:3-14.

2. Berman HM: The Protein Data Bank: a historical perspective. Acta Crystallogr A 2008, 64(Pt 1):88-95.

3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: The Protein Data Bank. Nucleic Acids Res 2000, 28(1):235-242.

4. Brazas MD, Fox JA, Brown T, McMillan S, Ouellette BF: Keeping pace with the data: 2008 update on the Bioinformatics Links Directory. Nucleic Acids Res 2008, 36(Web Server issue):W2-4.

5. Sundaramurthy P, Shameer K, Raashi S, Sunita G, and Sowdhamini R: HORI : a web server to compute higher order residue interactions in protein structures. BMC Bioinformatics 2010, 11(Suppl 1):2.

6. Levinthal C: Are there pathways for protein folding? J Chim Phys 1968, 65:44.

7. Anfinsen CB: The formation and stabilization of protein structure. Biochem J 1972, 128(4):737-749.

8. Betancourt MR, Thirumalai D: Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci 1999, 8(2):361-369.

9. Munson PJ, Singh RK: Statistical significance of hierarchical multi-body potentials based on Delaunay tessellation and their application in sequence-structure

alignment. Protein Sci 1997, 6(7):1467-1481.

10. Pace CN, Shirley BA, McNutt M, Gajiwala K: Forces contributing to the conformational stability of proteins. FASEB J 1996, 10(1):75-83.

11. Sippl MJ: Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 1990, 213(4):859-883.

12. Sippl MJ: Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995, 5(2):229-235.

13. Rojnuckarin A, Subramaniam S: Knowledge-based interaction potentials for proteins. Proteins 1999, 36(1):54-67.

14. Singh RK, Tropsha A, Vaisman, II: Delaunay tessellation of proteins: four body nearest-neighbor propensities of amino acid residues. J Comput Biol 1996, 3(2):213-221.

15. Eswar N, Eramian D, Webb B, Shen MY, Sali A: Protein structure modeling with MODELLER. Methods Mol Biol 2008, 426:145-159.

16. Kopp J, Schwede T: The SWISS-MODEL Repository: new features and functionalities. Nucleic Acids Res 2006, 34(Database issue):D315-318.

17. Arnold K, Bordoli L, Kopp J, Schwede T: The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006, 22(2):195-201.

18. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr.: A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 2003, 12(9):2001-2014.

19. Teeter MM: Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. Proc Natl Acad Sci U S A 1984, 81(19):6014-6018.

20.   Bernstein HJ: Recent changes to RasMol, recombining the variants. Trends Biochem Sci 2000, 25(9):453-455.

21.   Sayle RA, Milner-White EJ: RASMOL: biomolecular graphics for all. Trends Biochem Sci 1995, 20(9):374.

22.   Herráez A: Biomolecules in the Computer: Jmol to the rescue. Biochem Educ, 2006, 34:7.

23.   Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 2008, 36(Database issue):D419-425.

24.   Murzin AG, Brenner SE, Hubbard T, Chothia C: SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995, 247(4):536-540.

25.   Rousseau F, Schymkowitz JW, Itzhaki LS: The unfolding story of three-dimensional domain swapping. Structure 2003, 11(3):243-251.

26.   Bennett MJ, Schlunegger MP, Eisenberg D: 3D domain swapping: a mechanism for oligomer assembly. Protein Sci 1995, 4(12):2455-2468.

27.   Bennett MJ, Choe S, Eisenberg D: Domain swapping: entangling alliances between proteins. Proc Natl Acad Sci U S A 1994, 91(8):3127-3131.

28.   deMare F, Kurtz DM, Jr., Nordlund P: The structure of Desulfovibrio vulgaris rubrerythrin reveals a unique combination of rubredoxin-like FeS4 and ferritin-like diiron domains. Nat Struct Biol 1996, 3(6):539-546.

29.   Katsuya Y, Mezaki Y, Kubota M, Matsuura Y: Three-dimensional structure of Pseudomonas isoamylase at 2.2 A resolution. J Mol Biol 1998, 281(5):885-897.

30.   Porter CT, Bartlett GJ, Thornton JM: The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res

2004, 32(Database issue):D129-133.

31.   Todd AE, Orengo CA, Thornton JM: Evolution of function in protein superfamilies, from a structural perspective. J Mol Biol 2001, 307(4):1113-1143.

32.   Muller JJ, Thomsen KK, Heinemann U: Crystal structure of barley 1,3-1,4-beta-glucanase at 2.0-A resolution and comparison with Bacillus 1,3-1,4-beta-glucanase. J Biol Chem 1998, 273(6):3438-3446.

33.   Coggill P, Finn RD, Bateman A: Identifying protein domains with the Pfam database. Curr Protoc Bioinformatics 2008, Chapter 2:Unit 2 5.

34.   Sammut SJ, Finn RD, Bateman A: Pfam 10 years on: 10,000 families and still growing. Brief Bioinform 2008, 9(3):210-219.

Acknowledgements

Figures

Quadruple Interactions:

$i = (i + 4)$

Quadruple interaction mapped on to sequence of PDB ID 1CRN: (T-P) (P-A) (A-F) (F-T) (T-A) (P-F)

TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN

Triplet Interactions:

$i = (i + 4)$

Triplet interaction mapped on to sequence of PDB ID 1CRN: (T-P) (P-A)

(A-T)

TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN
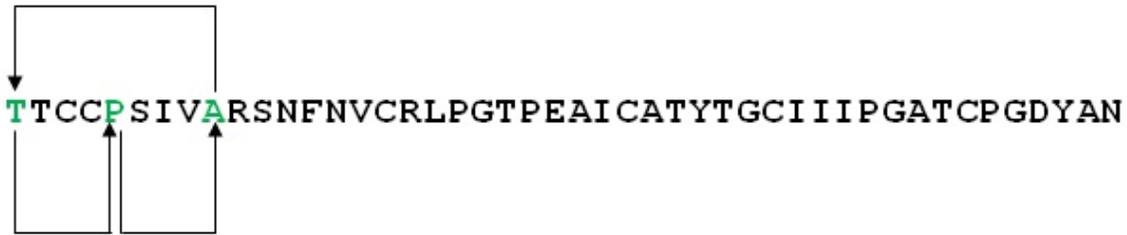
Pairwise Interactions:

i = (i + 4)

Pair wise interaction mapped on to sequence of PDB ID 1CRN: (T-P)

TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN

Figure 1

Graphical representation about the approach used for the calculation of pairwise, triplet and

quadruple interactions

Figure 2

Screenshots of web-interfaces of HORI-Global, HORI-Lite and HORI-Cluster

No.	Residue No.	Interacting Residues	Distance	Rasmol	Jmol
6	1-38	THR-ALA	5.738		
7	2-32	THR-CYS	6.742		
8	2-33	THR-ILE	5.587		
9	2-34	THR-ILE	4.597		
10	2-35	THR-ILE	6.394		
11	3-32	CYS-CYS	5.610		
12	3-33	CYS-ILE	5.365		
13	3-34	CYS-ILE	6.483		
14	3-40	CYS-CYS	5.687		
15	3-44	CYS-TYR	5.309		
16	3-45	CYS-ALA	6.498		
17	3-46	CYS-ASN	6.148		
18	4-9	CYS-ALA	5.999		
19	4-10	CYS-ARG	5.517		
20	4-31	CYS-GLY	6.762		
21	4-32	CYS-CYS	4.090		
22	4-33	CYS-ILE	6.332		
23	4-44	CYS-TYR	4.836		
24	4-45	CYS-ALA	6.286		
25	4-46	CYS-ASN	4.828		
26	5-9	PRO-ALA	6.398		
60	25-29	ILE-TYR	6.140		
61	26-30	CYS-THR	6.087		
62	27-31	ALA-GLY	5.572		
63	27-32	ALA-CYS	4.627		
64	27-33	ALA-ILE	5.364		
65	28-32	THR-CYS	6.935		
66	40-44	CYS-TYR	6.143		
67	40-45	CYS-ALA	6.295		
68	41-45	PRO-ALA	6.297		
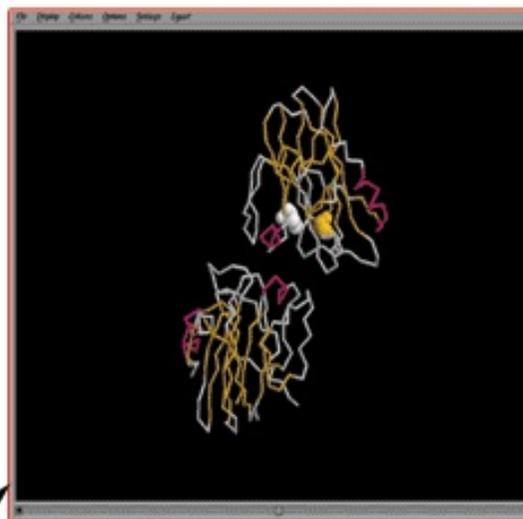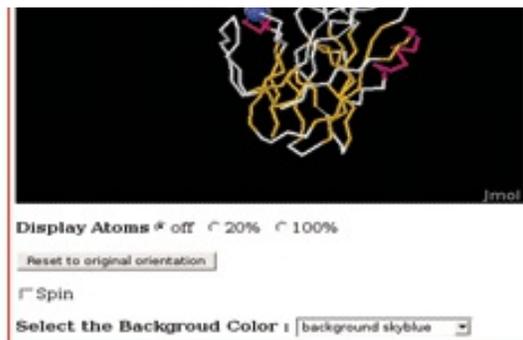
Click Here to download tab-delimited text file of HORI - Pairwise Interactions

Display Atoms ⦿ off ○ 20% ○ 100%

Reset to original orientation

☐ Spin

Select the Backgroud Color : background skyblue ▾

HORI - Triplet Interactions:

Hide Triplet - Interactions

No.	Residue No.	Interacting Residues	Distance	Residue No.	Interacting Residues	Distance	Residue No.	Interacting Residues	Distance	Rasmol	Jmol
1	3-40	CYS-CYS	5.687	40-44	CYS-TYR	6.143	44-3	TYR-CYS	5.309		
1	3-40	CYS-CYS	5.687	40-45	CYS-ALA	6.295	45-3	ALA-CYS	6.498		
1	4-9	CYS-ALA	5.999	9-31	ALA-GLY	6					
1	4-9	CYS-ALA	5.999	9-32	ALA-CYS	6					
1	4-10	CYS-ARG	5.517	10-46	ARG-ASN	6					
1	5-9	PRO-ALA	6.398	9-31	ALA-GLY	6					
1	5-9	PRO-ALA	6.398	9-32	ALA-CYS	6					
1	6-10	SER-ARG	6.483	10-46	ARG-ASN	6					
1	9-13	ALA-PHE	6.065	13-30	PHE-THR	6					
1	13-17	PHE-ARG	6.269	17-26	ARG-CYS	6					
1	13-26	PHE-CYS	5.787	26-30	CYS-THR	6					
1	16-21	CYS-THR	6.402	21-26	THR-CYS	6					
1	17-21	ARG-THR	5.176	21-26	THR-CYS	6					

Click Here to download tab-delimited text file of HORI - Triplet Interac

HORI - Quadruple Interactions:

Show Quadruple - Interactions
Click Here to download tab-delimited text file of HORI - Quadruple Int

Figure 3

Screenshot of details about different sections of a successful HORI-Global output

PDB ID	Molecule	URL to access HORI server results and Example runs using PDB structures mentioned in this protocol
1CRN	Crambin	http://caps.ncbs.res.in/hori/hori_results/er_hori_SatDec13185917IST2008.html http://caps.ncbs.res.in/hori/hori_cluster_quadruple_sc_4_er.html http://caps.ncbs.res.in/hori/hori_results/er_hori_SatDec13185917IST2008.html
1NNQ	Ruberythrin	http://caps.ncbs.res.in/hori/HORI_1NNQ.html
1BF2	Isoamylase	http://caps.ncbs.res.in/hori/hori_lite_ca1_er.html
1QJ4	Hydroxynitrile lyase	http://caps.ncbs.res.in/hori/hori_cluster_quadruple_sc_2_er.html
1AQ0	Barley 1,3-1,4-beta-glucanase	http://caps.ncbs.res.in/hori/hori_lite_quadruple_sc_2_er.html

Figure 4

Table 1 Details about PDB structures mentioned in this protocol with URL to access their

HORI results