

R. Nagarajan¹, A. Archana¹, A. Mary Thangakani^{2,3}, S. Jemimah¹, D. Velmurugan^{2,3} and M. Michael Gromiha¹

¹Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India. ²CAS in Crystallography and Biophysics, University of Madras, Chennai, India. ³Bioinformatics Infrastructure Facility, University of Madras, Chennai, India.

ABSTRACT: Understanding the structure–function relationship in proteins is a longstanding goal in molecular and computational biology. The development of structure-based parameters has helped to relate the structure with the function of a protein. Although several structural features have been reported in the literature, no single server can calculate a wide-ranging set of structure-based features from protein three-dimensional structures. In this work, we have developed a web-based tool, PDBparam, for computing more than 50 structure-based features for any given protein structure. These features are classified into four major categories: (i) interresidue interactions, which include short-, medium-, and long-range interactions, contact order, long-range order, total contact distance, contact number, and multiple contact index, (ii) secondary structure propensities such as α -helical propensity, β -sheet propensity, and propensity of amino acids to exist at various positions of α -helix and amino acid compositions in high B -value regions, (iii) physicochemical properties containing ionic interactions, hydrogen bond interactions, hydrophobic interactions, disulfide interactions, aromatic interactions, surrounding hydrophobicity, and buriedness, and (iv) identification of binding site residues in protein–protein, protein–nucleic acid, and protein–ligand complexes. The server can be freely accessed at <http://www.iitm.ac.in/bioinfo/pdbparam/>. We suggest the use of PDBparam as an effective tool for analyzing protein structures.

KEYWORDS: protein three-dimensional structure, physicochemical properties, binding sites, secondary structure propensity

CITATION: Nagarajan et al. PDBparam: Online Resource for Computing Structural Parameters of Proteins. *Bioinformatics and Biology Insights* 2016;10 73–80 doi: 10.4137/BBI.S38423.

TYPE: Software or Database Review

RECEIVED: January 19, 2016. **RESUBMITTED:** April 20, 2016. **ACCEPTED FOR PUBLICATION:** April 24, 2016.

ACADEMIC EDITOR: J. T. Efrid, Associate Editor

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 824 words, excluding any confidential comments to the academic editor.

FUNDING: This study was funded by the Department of Science and Technology, Government of India to MMG (SR/SO/BB-0036/2011). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: gromiha@iitm.ac.in

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE). Provenance: the authors were invited to submit this paper.

Published by Libertas Academica. Learn more about this journal.

Introduction

It is widely accepted that the structure of a protein dictates its function.¹ Most studies of protein structure and function rely on the analysis of the crystal structure of proteins. This is done by calculating various structure-based parameters, which have been developed to describe the folding, stability, and functions of proteins and their complexes, such as the nature of interactions among the amino acid residues and the surrounding solvent molecules, the preferred amino acid residues in the protein environment, the location of residues in the interior/surface of the protein, and the amino acid clusters.²

These parameters focus on specific aspects of the protein structure and are described in the literature. For instance, Lee and Richards³ developed the concept of solvent accessibility of amino acid residues. Chou and Fasman⁴ studied the secondary structures of proteins and deduced the propensity of amino acid residues present in α -helices, β -strands, and turns. Thornton's group developed several algorithms for identifying ion pairs, hydrogen bonds, and catalytic sites in proteins.^{5–7} Manavalan and Ponnuswamy⁸ proposed the concept of

surrounding hydrophobicity to characterize the hydrophobic behavior of amino acid residues in the protein environment. Plaxco et al.⁹ analyzed the contacts between amino acid residues and developed the concept of contact order (CO) to relate the folding rates of two-state proteins. Gromiha and Selvaraj¹⁰ considered contacts that are close in space but far away in the sequence and proposed long-range order (LRO) as a parameter for understanding protein-folding rates. This concept was refined by developing multiple contact index, ie, residues having multiple contacts in two- and three-state proteins.¹¹

Methods are also available to identify binding site residues in protein complexes based on distances between atoms, energetic contributions, and changes in accessible surface area upon binding.^{12–14} Many standalone programs and online servers (such as DSSP,¹⁵ NACCESS,¹⁶ HYDROPRO,¹⁷ HYDRONMR,¹⁸ GETAREA,¹⁹ SCide,²⁰ ContPro,²¹ CAPTURE,²² HBPLUS,²³ CALCOM,²⁴ PSAP,²⁵ and SBPS²⁶) are available to calculate various structural parameters. For instance, DSSP¹⁵ provides information on the secondary structure and accessible surface area of



each amino acid residue in a protein. CALCOM is used to locate residues in the interior and surface based on the distance between the residues and the calculated center of mass of the given protein or peptide chain.²⁴ Tina et al.²⁷ developed a server, protein interactions calculator, to calculate the center of mass, hydrogen bond interactions, hydrophobic interactions, aromatic–aromatic interactions, aromatic–sulfur interactions, and cation– π interactions. Kozma et al.²⁸ developed a server to obtain the contact map for any given protein. Magyar et al.²⁹ utilized the concept of surrounding hydrophobicity, LRO, stabilization center, and conservation scores to identify the stabilizing residues in protein structures. ExPASy³⁰ is a collection of tools on various bioinformatic aspects including proteomics, genomics, structural bioinformatics, and systems biology. PDBsum³¹ provides pictorial analyses of several structural features of proteins, DNA, and ligands, as well as the interactions between them.

Although a number of structural parameters have been described in the literature and can be calculated using various servers and standalone programs, no single server exists to calculate a diverse set of parameters and provide the output in a standard format. Hence, we have developed a web server, PDB-param (<http://www.iitm.ac.in/bioinfo/pdbparam/>), to calculate the following four distinct groups of properties: (i) physicochemical properties, (ii) secondary structure propensities, (iii) interresidue interactions, and (iv) identification of binding site residues in protein–DNA/RNA, protein–ligand, and protein–protein complexes. The server and the properties calculated are explained later.

Materials and Methods

A brief description of the properties under the four categories (physicochemical properties, secondary structure propensities, interresidue interactions, and binding site residues in protein complexes) is provided in this section.

Interresidue interactions. For the past three decades, studies on the mechanism of protein folding and stability have focused on interresidue interactions.³² Interactions between amino acid residues of the protein and with the surrounding solvent molecules play an important role in the formation of stable secondary structures and a unique tertiary structure for the protein. These interactions are usually noncovalent and include hydrogen bonds, ion pairs, van der Waals interactions, and hydrophobic interactions. In fact, parameters such as CO and LRO show a very strong correlation with the folding rate of small proteins.^{9,10}

Short-, medium-, and long-range interactions. For a given residue, the surrounding residues within a sphere of 8 Å radius are analyzed in terms of their sequence position. Residues within a distance of two residues from the central residue are considered to contribute to short-range interactions, those within a window between three and four residues to medium-range interactions and those more than four residues apart to long-range interactions.

Number of contacts (8/14 Å, C_α/C_β atoms). The contacts between amino acid residues in the crystal structure are computed with cutoffs of 8 and 14 Å using C_α or C_β atoms, as reported widely in literature.³²

Contact order. This parameter reflects the relative importance of local and nonlocal contacts to the native structure of a protein.⁹ It is defined as

$$CO = \frac{\sum \Delta S_{ij}}{L \times N}$$

where N is the total number of contacts, ΔS_{ij} is the sequence separation between two contacting residues i and j , and L is the total number of residues in the protein.

Long-range order. LRO is derived from long-range contacts (contacts between two residues that are close in space and far in the sequence) in the protein structure.¹⁰ It is defined as

$$LRO = \frac{\sum n_{ij}}{N}$$

$$n = 1 \text{ if } |i - j| > 12; 0 \text{ otherwise}$$

where i and j are the two contacting residues within a distance of 8 Å, and N represents the total number of residues in the protein.

Total contact distance. A new parameter total contact distance was developed by taking the product of CO and LRO. This parameter shows good correlation with the folding rates of proteins.³³

Multiple contact index. It considers the distance between amino acid residues in protein structure, residue separation at the sequence level, and the number of residues that have multiple contacts.¹¹ Multiple contact index has been derived separately for two- and three-state proteins.

Two-state proteins:

$$MCI = \frac{\sum n_{mi}}{N}$$

$$n_{mi} = 1 \text{ if } n_{ci} \geq 4; 0 \text{ otherwise}$$

$$n_{ci} = \sum n_{ij}; n_{ij} = 1 \text{ if } r_{ij} < 7.5 \text{ \AA} \text{ and } |i - j| > 12 \text{ residues};$$

$$0 \text{ otherwise}$$

Three-state proteins:

$$MCI = \frac{\sum n_{mi}}{N}$$

$$n_{mi} = 1 \text{ if } n_{ci} \geq 5; 0 \text{ otherwise}$$

$$n_{ci} = \sum n_{ij}; n_{ij} = 1 \text{ if } r_{ij} < 6.5 \text{ \AA} \text{ and } |i - j| > 3 \text{ residues};$$

$$0 \text{ otherwise}$$

where n_{ci} is the number of contacts for each residue, and r_{ij} is the distance between the residues i and j .



Propensities. Propensities indicate the preference of amino acid residues for different secondary structures. The propensities listed in PDBparam are given below.

α -Helical, β -strand, and coil tendencies. The α -helical propensities can be computed by taking into account the frequency of amino acids in these regions.

$$\begin{aligned} & \% \text{ of residue } i \text{ in } \alpha\text{-helix} \\ &= \frac{\text{count of residue } i \text{ in the } \alpha\text{-helix}}{\text{total count of residue } i \text{ in the whole protein}} \end{aligned}$$

i varies from 1 to 20, number of amino acid residues. Similar equations have been used to compute strand and coil propensities.

Frequency of occurrence in β -bends. Certain segments in the polypeptide chain help in bringing the distant residues into close proximity during the folding process. For example, β -bends³⁴ allow hydrogen bonds to form between the C = O group of residue i and NH group of residue $(i + 3)$.

Criteria to occur in β -bends:

- Distance between $C_{\alpha}(i)$ to $C_{\alpha}(i + 3)$ carbon atoms should be less than 7 Å.
- The $(i + 1)$ th or $(i + 2)$ th residue is not in an α -helix.

Amino acid compositions in turns. An open turn exists in a protein if the distance between C_{α}^1 to C_{α}^4 carbon atoms is < 5.7 Å.³⁵ Turns are usually present where a strand of β -sheet reverses itself to form the next antiparallel strand or keep the helices, β -sheets, and random coils in a compact globular form and are thus used to predict protein structure.

Normalized frequency of helix. Helical regions are divided into three zones³⁵: the first three residues represent the N-helix, the last three represent the C-helix, and the residues in the middle represent the M-helix. The amino acid frequency in each helical zone divided by the total frequency (in the entire protein) constitutes normalized frequency.

Propensity to form multiple contact index. The frequency of occurrence of amino acid residues that form multiple contacts (f_{mc}) and in the protein as a whole (f_t) is computed.¹¹ The propensity, P_{mc} can be calculated as follows:

$$P_{mc}(i) = \frac{f_{mc}(i)}{f_t(i)}$$

where i represents each of the 20 amino acid residues.

Amino acid composition in high B -value regions. Temperature factors (ie, B -values) provide a measure of the degree of uncertainty in the position of an atom due to thermal motion and/or positional disorder. Analyzing B -values provides insights into protein flexibility and protein dynamics. The B -values at C_{α} atoms are normalized and residues with B -values greater than $B_{mean} + 0.5 \times B_{\sigma}$ are labeled as high B -value residues.³⁶

Physicochemical properties of proteins. *Center of mass.* The center of mass can be used to define constraints in

predicting protein tertiary structures to assess the global shape of the protein partners in protein–protein complexes and to measure their distance.²⁴ It is given by

$$x_{COM} = \frac{\sum_{i=1}^N m_i x_i}{\sum_{i=1}^N m_i}$$

where x_i is the X coordinate of the atom i and m_i is the atomic mass. The Y and Z coordinates of the center of mass can be calculated using a corresponding formula.

Radius of gyration. The radius of gyration describes the compactness of the protein. It is calculated as follows:

$$ROG = \sqrt{\frac{\sum_i m_i |x_i - COM|^2}{\sum m_i}}$$

where m_i denotes the mass of each atom, COM denotes the center of mass of protein, and x_i represents the atomic coordinate.

Surrounding hydrophobicity. The sum of hydrophobic indices assigned to the residues that appear within a distance of 8 Å from the central residue⁸ can be used to characterize the hydrophobic behavior of each amino acid residue in the protein environment. It is defined as

$$H_p(i) = \sum_{j=1}^{20} n_{ij} \times h_j;$$

where n_{ij} is the total number of surrounding residues of type j around the i th residue of the protein, and h_j is the hydrophobicity index (kcal/mol) obtained from thermodynamic transfer experiments.^{37,38}

Gain in surrounding hydrophobicity of a residue. For a given amino acid, the increase in surrounding hydrophobicity as the protein transitions from its unfolded state to its native (ie, folded) state represents the enrichment in the hydrophobic property of that residue. To compute the gain in surrounding hydrophobicity³⁹ for each residue in the protein molecule, it is assumed that the fully extended chain conformation is the unfolded reference state.

Surrounding hydrophobicity in the unfolded

$$\text{state of the } j\text{th residue} = \sum_{k=j-2; k \neq j}^{k=j+2} h_k;$$

The average gain ratio in surrounding hydrophobicity is given by

$$G_j = \frac{H_j^f}{H_j^u}$$

where H^f and H^u denote the hydrophobic index of the j th residue in the folded state and unfolded state of the protein, respectively.



Surface hydrophobicity. This is computed from the protein crystal structure by considering the hydrophobic contribution of exposed amino acid residues. Surface hydrophobicity³⁸ is given by

$$\Phi_{\text{surface}} = \frac{\sum s_i \times \Psi_i}{s_p}$$

where s_i is the solvent accessible surface area occupied by the i th residue, Ψ_i is the hydrophobicity value assigned to the residue, and s_p is the solvent accessible surface area of protein.

Hydrophobic accessible area. It is calculated as the solvent accessible surface area of the hydrophobic residues on the protein surface.⁴⁰ We considered Ala, Val, Leu, Ile, Met, Phe, and Pro as the hydrophobic residues to calculate the hydrophobic accessible area.

Accessible surface area for the native protein. The accessible surface area (ASA) for the native protein is calculated as the sum of the accessible surface area of each residue present in the protein, which is obtained from DSSP.¹⁵

Buriedness. The buriedness² of each residue is calculated as the ratio of number of residues in the interior of the protein and the total number of residues in the protein.

Mean area buried on transfer. The mean area buried on transfer⁴¹ is given by difference in the accessible area in the unfolded and folded states of the protein.

$$\text{Mean area buried on transfer} = A^0 - \langle A \rangle$$

where A^0 and $\langle A \rangle$ represent the accessible areas in unfolded and folded states of protein, respectively.

Mean fractional area loss. During the process of folding, the nonpolar residues avoid contact with solvent molecules and are buried inside the protein. The area lost when a residue is buried is proportional to its hydrophobic contribution. This is termed as solvent accessible reduction ratio⁴¹ or mean fractional area loss, denoted as $\langle R_A \rangle$:

$$\langle R_A \rangle = \frac{A^0 - \langle A \rangle}{A^0}$$

where A^0 and $\langle A \rangle$ represent the accessible areas in unfolded and folded states of protein, respectively.

Normalized flexibility parameters (B-values). This parameter can be computed from the temperature factors extracted from the PDB for the N, C α , C, and O atoms. Based on the deviation of B-value from the mean, each residue was classified as flexible or rigid. The normalized B-values⁴² were determined for each residue type, ie, when surrounded by none, one, or two rigid neighbors.

Noncovalent interactions. Several interactions (hydrophobic, hydrogen bond, ionic, aromatic, cation- π , and disulfide bonds) have been described in terms of the amino acid residues involved and the distance between two specific amino

acid residues. The details of the amino acid residues in each interaction along with the distances²⁷ are given in Table 1.

Hydrophobic-free energy. The hydrophobic-free energy⁴³ is expressed as

$$G_{\text{hy}} = \sum_i \Delta \sigma_i \left[A_i(\text{folded}) - A_i(\text{unfolded}) \right]$$

where $A_i(\text{folded})$ and $A_i(\text{unfolded})$ represent the accessible surface areas of each atom in the folded and unfolded (extended) states of the protein, respectively.

The solvent accessible surface areas of all the atoms in the folded state were computed using the program NACCESS.¹⁶ The extended state ASA of the atom was obtained from literature. They are in the form of a Gly-X-Gly (where X is the amino acid) sequence in a typical extended conformation. σ_i (atomic solvation parameters) for the five classes of atoms (namely, carbon, neutral nitrogen and oxygen, charged nitrogen, charged oxygen, and sulfur) are determined by a least-squares fit of above equation. The σ_i values are C: 12.02, N/O: -5.86, N⁺: -19.46, O⁻: -34.98, and S: 35.51 (in units of cal/mol Å²).⁴³

Free energy due to disulfide interactions. The free energy due to disulfide interactions is calculated using the formula:

$$G_{\text{SS}} = 2.3 N_{\text{SS}}$$

where N_{SS} is the number of disulfide bonds in the protein.

Hydrogen bond interactions. It is classified into the following three main categories: main chain-main chain, main chain-side chain, and side chain-side chain interactions. These interactions are calculated using HBPLUS,²³ a hydrogen bond calculation program.

Table 1. Distance criteria for noncovalent interactions and disulfide bonds.

NAME OF THE INTERACTIONS	INTERACTING RESIDUES	DISTANCE CRITERIA
Disulfide	Pair of cysteines	2.2Å
Ionic Interactions	(R,K) with (D,E,H)	6.0 Å
Hydrophobic interactions	A,V,L,I,M,F,W,P,Y	5.0 Å
Hydrogen bond interactions	Donor-acceptor distance cut-off (O and N)	3.5Å
	Donor-acceptor distance cut-off (sulfur)	4.0Å
Aromatic-Aromatic interactions	Pairs of phenyl ring	4.5 to 7.0 Å
Aromatic-sulfur interactions	Sulfur atoms of C, M and thearomatic rings of F,Y,W	5.3 Å
Cation- π interactions	Cationic side chain (Lys or Arg) is near an aromatic side chain (Phe, Tyr, or Trp)	6.0 Å



Identification of binding sites in protein–DNA/RNA and protein–protein complexes. Protein–DNA interactions play a key role in many vital processes, including regulation of gene expression, DNA replication and repair, and packaging. The binding sites for a protein–DNA/RNA complex can be identified using the following distance criteria¹²: an amino acid residue within a protein is designated as a binding site residue if its side chain or backbone atoms are within a cut-off distance (eg, 3.5 Å) from any atom in DNA/RNA.^{44–46} The binding sites for protein–protein complexes were also computed using the distance criteria between different chains present in the protein.

Server Description and Implementation

The PDBparam server can calculate more than 50 parameters from the three-dimensional structure of a protein. Each parameter has been treated as a separate module, and the script has been written using perl. The perl-CGI scripts are used to render the HTML web pages. The PDBparam server works with the PDB file as input and provides the computed results in a single output page. The output can be downloaded as a PDF file. The results for all the parameters were cross-checked manually with several structures of proteins and their complexes. Furthermore, the documentation has been provided for all the parameters listed in PDBparam on the website. It is

Please enter the following

Enter a PDB-id Chain name

Or

Upload a PDB File No file selected. Chain name

Select one of the following

Identification of binding site

Inter-residue interactions

Secondary structure propensities

Physicochemical properties

Identification of binding site

Protein-Protein Protein-Ligand Protein-DNA/RNA

Chain Name Distance threshold

Identification of binding site - Protein-DNA/RNA

PDB ID : 6CRO.pdb

Residue name	Atom name	chain name	Residue name	Atom name	chain name	Distance		
PHE	14	O	A	DT	1	O2	U	3.140
GLN	16	N	A	DT	3	OP2	R	3.000
GLN	16	N	A	DT	3	OP2	U	3.000
THR	17	OG1	A	DT	3	C7	R	3.436
THR	17	OG1	A	DT	3	C7	U	3.436
VAL	25	CG2	A	DT	15	OP2	R	2.791
VAL	25	CG2	A	DT	15	OP2	U	2.792
TYR	26	N	A	DT	15	OP2	R	2.843
TYR	26	N	A	DT	15	OP2	U	2.844
TYR	26	CB	A	DT	15	OP2	R	3.390
TYR	26	CB	A	DT	15	OP2	U	3.390

Figure 1. Steps to identify the binding sites in a protein–DNA complex.



linked with other online tools available in the literature. The utility of the server is described with a few examples.

Example 1: Identify the binding site residues in a protein–DNA complex (PDB code: 6CRO) using the distance cutoff of 3.5 Å.

Steps:

1. Enter the PDB code and chain (optional; case sensitive); eg, PDB code: 6CRO.

2. Check “identification of binding site” and submit.
3. In the new page, check protein–DNA/RNA.
4. Give the distance (default cutoff is 3.5 Å).
5. Click on submit.

Figure 1 shows the relevant items to be checked, the required information, and the output. The output contains information on the residue name, residue number, atom name, and chain name of both protein and DNA and the distance

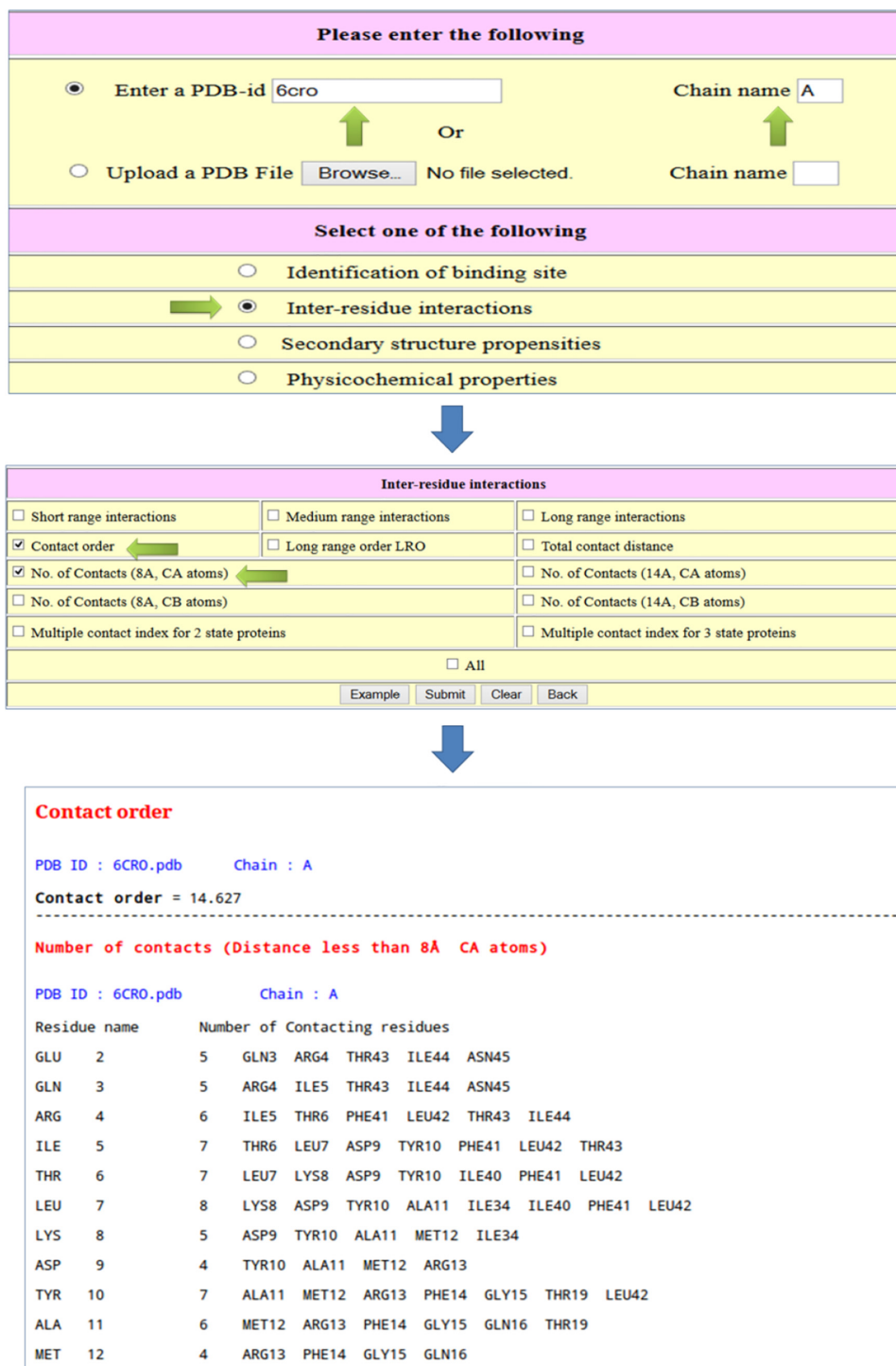


Figure 2. Example to compute the contact order of a protein and the number of contacts for all the amino acid residues in a protein.



between the atoms. These residues are identified as binding sites. We have also provided options to display the structure of the complex, highlighting the binding site residues.

Example 2: Calculate the CO of the protein, 6CRO (A chain), and the number of contacts for all the residues using C_{α} atoms within the limit of 8 Å.

Steps:

1. Enter the PDB code and chain (optional; case sensitive).
2. Check “interresidue interactions” and submit.
3. In the new page, check “contact order and number of contacts (8 Å, CA atoms)”.
4. Click on submit.

Figure 2 shows the relevant items for computing the CO and number of contacts and the output. The output displays the CO for the protein and the number of contacts for all the residues with residue name and number. The contacting residues are also shown in the output.

Availability of PDBparam. PDBparam is freely available at <http://www.iitm.ac.in/bioinfo/pdbparam>.

Applications. PDBparam computes various structure-based parameters on interresidue interactions, amino acid propensities, physicochemical properties, and binding sites. This information can be used to understand the structure and functions of proteins and their complexes. The contacts between amino acid residues in protein structures provide data on the location of amino acid residues and preferred contacts in the protein environment, which can be used to comprehend protein folding and predict protein structures.³² The topological parameters, such as CO, LRO, total contact distance, and multiple contact distance, are helpful in understanding protein-folding rates and folding kinetics.^{9–11} Specific physicochemical interactions between amino acid residues in protein structures, such as cation- π , aromatic clusters, and hydrogen bonds, reveal the importance of these interactions in protein stability.²⁷ The combination of secondary structure and solvent accessibility is useful in identifying functionally important residues in proteins.^{15,16} Furthermore, the identification of binding sites in protein-protein, protein-nucleic acid, and protein-ligand complexes can be effectively used to compute the binding propensity and affinity and understand the recognition mechanism of protein complexes.^{46–51}

PDBparam can be used to compute important parameters for any specific protein, providing deep insights into its structure-function relationship. It can also be used for large-scale analysis of different types of proteins to explore potential interactions and contacts, which will provide insights on the similarities and differences crucial to understanding the function.

Conclusion

The PDBparam server can calculate more than 50 parameters from the three-dimensional structure of a protein, classified into the following four categories: physicochemical properties,

interresidue interactions, secondary structure propensities, and identification of binding sites in protein-DNA/RNA and protein-protein complexes. All the parameters have been coded using perl. Furthermore, perl-CGI scripts are used to render the HTML web pages. Detailed documentation for the protein properties and links of other available web servers related to such properties are provided, in order to enhance the user's ease of access.

Acknowledgment

We thank the Bioinformatics Facility, Department of Biotechnology, and IIT Madras for computational facilities.

Author Contributions

Conceived and designed the study: MMG, DV. Web server development: AA, AMT, RN. Discussions: AA, AMT, RN, SJ, DV, MMG. Wrote the first draft of the article: AA, MMG. Contributed to the writing of the article: AMT, RN, SJ, DV. All the authors reviewed and approved the final article.

REFERENCES

1. Branden C, Tooze J. *Introduction to Protein Structure*. New York: Garland Science; 1999:13–34.
2. Gromiha MM. *Protein Bioinformatics: From Sequence to Function*. Cambridge, MA: Academic Press; 2010.
3. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*. 1971;55:379–400.
4. Chou PY, Fasman GD. Prediction of protein conformation. *Biochemistry*. 1974;13:222–245.
5. Barlow DJ, Thornton JM. Ion-pairs in proteins. *J Mol Biol*. 1983;168:867–85.
6. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. *J Mol Biol*. 1984;238:777–93.
7. Furnham N, Holliday GL, de Beer TA, Jacobsen JO, Pearson WR, Thornton JM. The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res*. 2014;42:D485–9.
8. Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. *Nature*. 1978;275:673–4.
9. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*. 1998;277:985–94.
10. Gromiha MM, Selvaraj S. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol*. 2001;310:27–32.
11. Gromiha MM. Multiple contact network is a key determinant to protein folding rates. *J Chem Inf Model*. 2009;49:1130–5.
12. Ahmad S, Gromiha MM, Sarai A. Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information. *Bioinformatics*. 2004;20:477–86.
13. Tjong H, Zhou HX. DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces. *Nucleic Acids Res*. 2007;35:1465–77.
14. Gromiha MM, Fukui K. Scoring function based approach for locating binding sites and understanding the recognition mechanism of protein-DNA complexes. *J Chem Inf Model*. 2011;51:721–9.
15. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22:2577–637.
16. Hubbard SJ, Thornton JM. Naccess. [Computer Program]. *Biochemistry and Molecular Biology, University College London*; 1993. Available at: <http://www.bioinf.manchester.ac.uk/naccess/>.
17. Ortega A, Amorós D, García de la Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J*. 2011;101:892–8.
18. García de la Torre J, Huertas ML, Carrasco B. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*. 2000;147:138–46.
19. Fraczekiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem*. 1998;19:319–33.



20. Dosztanyi Z, Magyar C, Tusnady G, Simon I. SCide: identification of stabilization centers in proteins. *Bioinformatics*. 2003;19:899–900.
21. Firoz A, Malik A, Afzal O, Jha V. ContPro: a web tool for calculating amino acid contact distances in protein from 3D-structures at different distance threshold. *Bioinformatics*. 2010;5:55–7.
22. Gallivan JP, Dougherty DA. Cation- π interactions in structural biology. *Proc Natl Acad Sci*. 1999;96:9459–64.
23. McDonald IK, Naylor D, Jones D, Thornton JM. HBPLUS. [Computer Program]. Department of Biochemistry and Molecular Biology, University College, London; 1993. Available at: <http://www.ebi.ac.uk/thornton-srv/software/HBPLUS/>.
24. Costantini S, Paladino A, Facchiano AM. CALCOM: software for calculating the center of mass of proteins. *Bioinformatics*. 2008;2:271–2.
25. Balamurugan B, Md Roshan MNA, Hameed BS, et al. PSAP: protein structure analysis package. *J Appl Crystallogr*. 2007;40:773–7.
26. Gurusaran M, Shankar M, Nagarajan R, Helliwell JR, Sekar K. Do we see what we should see? Describing non-covalent interactions in protein structures including precision. *IUCr*. 2015;1:74–81.
27. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. *Nucleic Acids Res*. 2007;35:473–6.
28. Kozma D, Simon I, Tusnady GE. CMWeb: an interactive on-line tool for analysing residue-residue contacts and contact prediction methods. *Nucleic Acids Res*. 2012;40:W329–33.
29. Magyar C, Gromiha MM, Pujadas G, Tusnady GE, Simon I. SRide: a server for identifying stabilizing residues in proteins. *Nucleic Acids Res*. 2005;33:W303–5.
30. Artimo P, Jonnalagedda M, Arnold K, et al. ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012;40:W597–603.
31. de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. *Nucleic Acids Res*. 2014;42:D292–6.
32. Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. *Prog Biophys Mol Biol*. 2004;86:235–77.
33. Zhou H, Zhou Y. Folding rate prediction using total contact distance. *Biophys J*. 2002;82:458–63.
34. Lewis PN, Momany FA, Scheraga HA. Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc Natl Acad Sci*. 1971;68:2293–7.
35. Crawford JL, Lipscomb WN, Schellman CG. The reverse turn as a polypeptide conformation in globular proteins. *Proc Natl Acad Sci*. 1973;70:538–42.
36. Parthasarathy S, Murthy MRN. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng*. 2000;13:9–13.
37. Nozaki Y, Tanford C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale. *J Biol Chem*. 1971;246:2211–7.
38. Tanford C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc*. 1962;84:4240–7.
39. Ponnuswamy PK, Prabhakaran M, Manavalan P. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochim Biophys Acta*. 1980;623:301–16.
40. Mahn A, Lienqueo ME, Asenjo JA. Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *J Chromatogr A*. 2004;1043:47–55.
41. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. *Science*. 1985;229:834–8.
42. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins*. 2004;19:141–9.
43. Ponnuswamy PK, Gromiha MM. On the conformational stability of folded proteins. *J Theor Biol*. 1994;166:63–74.
44. Nagarajan R, Ahmad S, Gromiha MM. Novel approach for selecting the best predictor for identifying the binding sites in DNA binding proteins. *Nucleic Acids Res*. 2013;41:7606–14.
45. Nagarajan R, Gromiha MM. Prediction of RNA binding residues: an extensive analysis based on structure and function to select the best predictor. *PLoS One*. 2014;9:e91140.
46. Gromiha MM, Nagarajan R. Computational approaches for predicting the binding sites and understanding the recognition mechanism of protein-DNA complexes. *Adv Protein Chem Struct Biol*. 2013;91:65–99.
47. Gromiha MM, Siebers JG, Selvaraj S, Kono H, Sarai A. Role of inter and intramolecular interactions in protein-DNA recognition. *Gene*. 2005;364:108–13.
48. Nagarajan R, Chothani SP, Ramakrishnan C, Sekijima M, Gromiha MM. Structure based approach for understanding organism specific recognition of protein-RNA complexes. *Biol Direct*. 2015;10:8.
49. Yugandhar K, Gromiha MM. Feature selection and classification of protein-protein complexes based on their binding affinities using machine learning approaches. *Proteins*. 2014;82:2088–96.
50. Yugandhar K, Gromiha MM. Protein-protein binding affinity prediction from amino acid sequence. *Bioinformatics*. 2014;30:3583–9.
51. Yugandhar K, Gromiha MM. Computational approaches for predicting binding partners, interface residues and binding affinity of protein-protein complexes. *Methods Mol Biol*. 2016 (in press).