

Original Research Article

A comprehensive computational mutation structure-function approach for determining potential drug target sites in poliovirus 2A protease

Amna Younus*, Muhammad Faraz Bhatti

Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Sector H-12, Kashmir Highway, Islamabad 44000, Pakistan

*For correspondence: **Email:** amnayounus82@yahoo.com, **Tel:** 0092-305-5869935; **Fax:** 0092-51-90856102

Sent for review: 5 September 2017

Revised accepted: 16 November 2017

Abstract

Purpose: To investigate a computational approach for analysing the structure-function relationship of poliovirus 2A protease using various bioinformatics tools.

Methods: The three-dimensional structure of 2Apro was modelled and analyzed using the crystal structure of coxsackievirus B4 as a template to understand the function of this protein. Structural validation programs, VADAR and QMEAN, were used to verify the 2Apro model. Analysis of protein stability changes in poliovirus 2A protease-mutated sequences using various servers was also performed. Furthermore, mutation pattern, intrinsic disorder regions (IDRs), hydrophobic regions, drug binding sites (DBS) and subcellular localization were identified.

Results: Hydrophobicity results confirmed the suitability and reliability of 2A protease as a potential drug target. Less IDRs were observed in the protein. Moreover, the results showed the presence of various important drug binding targets among conserved regions of the protease. The predicted drug binding sites indicate their suitability for the inhibition and development of anti-viral drugs against poliovirus 2A protease.

Conclusion: The current study resulted in the detection of important ligand interactions with respect to the binding site of the targeted protein. Thus, these compounds may be potent drug candidates and their potency may be increased against poliovirus 2A protease with relatively simple structural changes.

Keywords: 2A Protease, Computational analysis, Drug binding sites, Intrinsic disorder regions, Hydrophobicity

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Tropical Journal of Pharmaceutical Research is indexed by Science Citation Index (SciSearch), Scopus, International Pharmaceutical Abstract, Chemical Abstracts, Embase, Index Copernicus, EBSCO, African Index Medicus, JournalSeek, Journal Citation Reports/Science Edition, Directory of Open Access Journals (DOAJ), African Journal Online, Bioline International, Open-J-Gate and Pharmacy Abstracts

INTRODUCTION

Viral proteins involved in the cellular functions and manipulation of cellular processes by viruses have gained significant attention from the pharmaceutical and biotechnology industries as

a potential source of drug targets [1]. Poliovirus proteinases are excellent substrates for drug development due to their key roles in the replication cycle.

Drug targets are proteins or protein products

including G-protein coupled receptors, enzymes, ion channels and amino acids. To accomplish desirable therapeutic effects drugs tend to bind to their matching targets [2]. The extreme genetic variability and high mutation rate of RNA viruses such as poliovirus makes the design of antiviral drugs complicated. Polioviruses are therefore termed “quasi-species” [3-5].

Poliovirus (PV) of the *Picornaviridae*, has been categorized into three stable subtypes (1, 2 and 3). The PV genome comprises +ss RNA of approximately 7,440 nucleotides. The genome is packaged in a non-enveloped icosahedral protein capsid [6]. PV replication is carried out through a single open reading frame termed a polyprotein. Two cysteine proteases (2A and 3C) are involved in proteolytic processing of polyprotein, which are crucial for viral life cycle [7]. The poliomyelitis epidemic resulted in the discovery of two important vaccines, OPV and IPV. No drug is available for treating poliomyelitis.

PV 2A^{pro} plays a leading role in autocatalytic cleavage process of poliovirus polyprotein. 2A^{pro} is involved in cleavage of eIF4G (eukaryotic translation initiation factor 4G [8]. Studies have demonstrated the role of poliovirus 2A^{pro} in

activating the apoptotic process in PV-infected cells, and it also exhibits anti-apoptotic activity [9].

The poliovirus 2A^{pro} crystal structure is unknown; however, it is possibly very close to the known structure of HRV2 2A^{pro} [10]. Research on 2A^{pro} inhibitors has been limited so far. However, a number of peptide-based inhibitors have been developed against 2A^{pro}. Anti-2A inhibitors are expected to show the additional advantage of suppressing genetic reversion along with negative effect on capsid formation [11]. Poliovirus proteases display minimal similarity to known mammalian enzymes, making them excellent drug targets [12]. The aim of this study was to explore poliovirus 2A protease as a probable drug target through computational approaches. Sequences in which mutations were already reported were used [13]. Analysis was performed to investigate drug target sites of 2A^{pro}. We also investigated the effect of mutations on protein stability. The degree of conservation was mapped on the 3D structure of the proteins and compared with the novel predicted intrinsic disordered regions (IDRs) of poliovirus 2A protease (Figure 1).

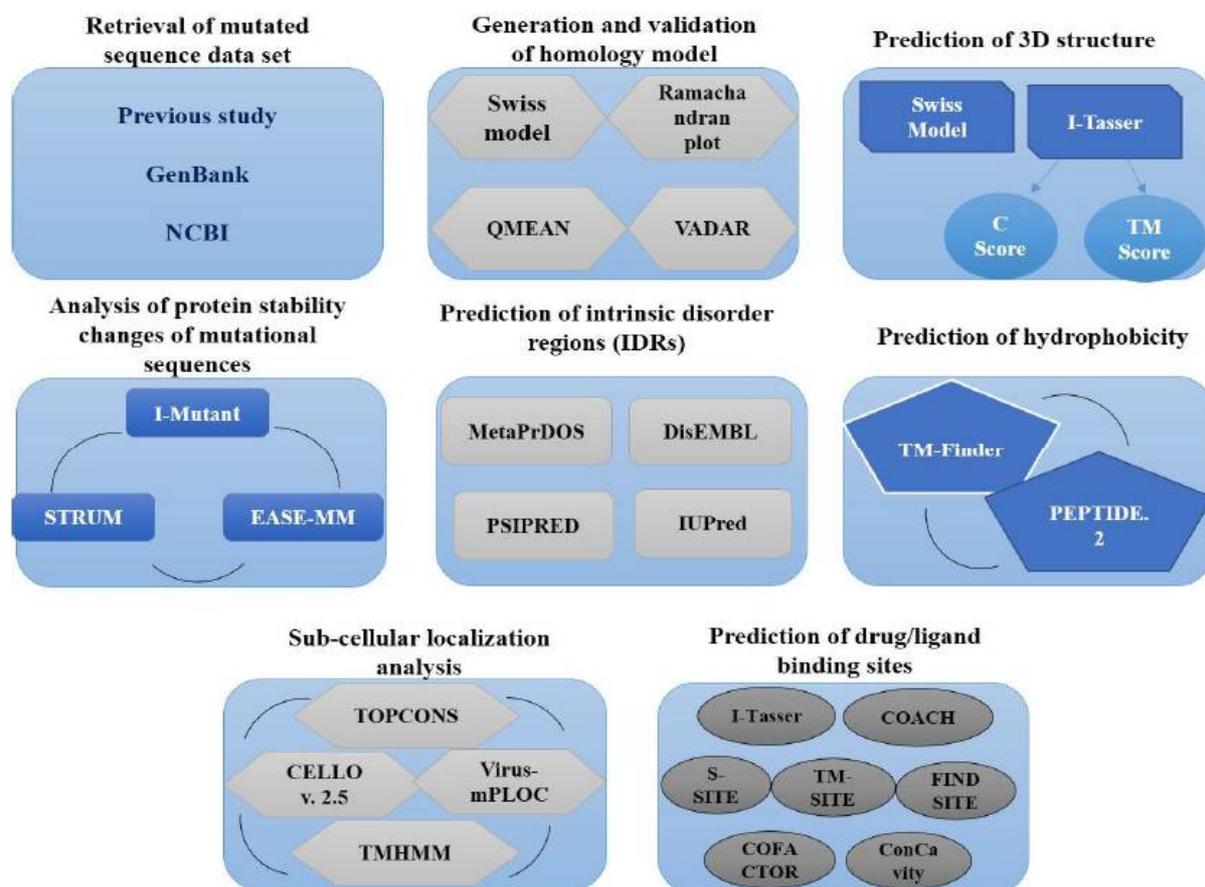


Figure 1: Schematic representation of prioritization pipeline in identifying regions with more DBS and least IDS based on various in-silico tools

EXPERIMENTAL

Sequence retrieval

The 2A protease was amplified from blood samples (001 and 002) (from polio patients in the KP district of Pakistan with the help of Frontier Development Organization) then cloned into a TA cloning vector (PCR 2.1) and sent for sequencing (Eurofin, USA). The sequences were analyzed and submitted to Genbank (KF193869 and KF193870 for 001 and 002 respectively) [13]. Total of four mutations in the poliovirus 2A protease were observed after sequencing. Alanine residue (at position 12) was mutated into glycine, while wild type aspartic acid residue at position 36 was mutated into asparagine and normal serine was mutated to threonine at position 134 and isoleucine mutated to valine at amino acid position 136.

Homology modelling and model reputation

A PDB file of 001 and 002 samples of 2A protease was generated by the Swiss-Model server [14]. The structure of coxsackievirus B4 (1z8r) model was selected as a template to obtain the 2A model. The backbone conformation of the modelled structure of 001 and 002 proteases was calculated by investigating the torsion angles *via* Ramachandran plot version 2.0 [15]. Further analysis of the predicted model was done by QMEAN [16] and ProSA [17]. VADAR was performed to determine the volume area dihedral angle for fractional accessible surface area [18].

Protein structure prediction

I-Tasser [19] server was used to predict protein structure by giving precise structural and functional predictions by means of algorithms [20]. A total of five models were obtained based on the five largest structure clusters. Confidence scores (C-score) were calculated based on the significance of threading template alignments. C-scores lied in the range of [-5, 2], where a higher value indicates a model with a high confidence and vice-versa. A TM-score (template modelling score) of > 0.5 predicts a model of correct topology while a TM-score < 0.17 specifies a random similarity. 2A^{pro} sequences containing mutations were subjected to structure prediction by I-Tasser.

Protein stability prediction

I-Mutant 2.0 [21] network predicts protein stability changes upon single point mutation from protein sequence/structure [22]. Output is obtained in the

form of protein stability change and Gibbs-free energy change (DDG) using Eq 1.

$$\text{DDG} = \text{DG (mutant Protein)} - \text{DG (wild-type, kcal/mol)} \dots (1)$$

Furthermore, STRUM was used to predict fold stability change ($\Delta\Delta G$) of protein molecules upon single-point mutations [23].

EASE-MM sequence-based prediction of mutation-induced stability changes [24], was also used to affirm the stability of the protein upon mutations. Results are displayed in the form of an increase or decrease in stability upon mutation induction.

Protein intrinsic disorder region (IDR) analysis

Variations in IDR structural properties can play a regulatory role in protein activity. The IDRs of the 001 and 002 poliovirus 2A protease, were estimated using four software packages: MetaPrDOS [25], DisEMBL [26], PSIPRED [27] and IUPred [28]. A disorder score (0-1) for each amino acid position was calculated. Position of an amino acid was considered IDR if its disorder score was observed above the cut-off value of 0.5.

Prediction of hydrophobicity

Prediction of overall hydrophobic residues of 001 and 002 poliovirus 2A protease was performed using the TM Finder tool [29]. The TM-Finder predicts trans-membrane proteins. The regions are predicted to be trans-membrane based on the hydrophobicity and helicity of the adjacent amino acid sequences. The threshold value for hydrophobicity is 0.4 and values exceeding the threshold level predict the transmembrane nature of the protein [29]. Hydrophobicity results of the protein were also confirmed using the PEPTIDE.2 server.

Prediction protein–ligand–drug binding site

Tertiary structures of 001 and 002 poliovirus 2A protease were analyzed to predict the number of drug binding sites. The online servers COACH [19, 30] I-Tasser were used to determine the number of binding sites and their positions. COACH is a meta-server approach for the prediction for ligand binding targets through two comparative methods TM-SITE [31] and S-SITE. The BioLiP protein function database is used in these methods to recognize ligand binding sites. These predictions were later combined with the results obtained from other methods including ConCavity [32], FINDSITE [33] and COFACTOR

[20]. To predict the final ligand binding site of the protein, PDB structures of 001 and 002 poliovirus 2A protease were given to the server for analysis. Complete results were given in the form of the position of the potential drug binding site.

Subcellular localization analysis

To understand a protein's function and its suitability as a vaccine/target, subcellular localization analysis is performed. Cytoplasmic proteins tend to have more potential as drug targets compared to surface membrane proteins [34]. The CELLO v.2.5 [35] server was employed to predict subcellular localization of poliovirus 2A protease. The results were further validated with predictions obtained from Virus-mPLoc [36], TOPCONS [37] and TMHMM [38].

RESULTS

Structures of 001 and 002 PV2A protease models

The Swiss-Model predicts the 3D structure of both the models on the basis of known crystal structure of the protein (Figure 2). The maximum scoring and validated model for PV2A protease exhibiting the highest sequence identity with the crystal structure was coxsackievirus B4 (PDB ID: 1Z8R.A).

The stereo-chemical attributes of the predicted protein models were investigated through QMEAN (Table 1) and evaluated by Ramachandran plots (Figure 3).

The QMEAN4 and QMEAN6 score [39] of the model was close to the value of 0, showing the reliability of the model (model reliability in between 0 and 1; Table 1). Predicted structure analysis delivered solid evidence of the good quality of predicted 3D structure of poliovirus 2A protease.

Protein model quality and accessible surface area (ASA) was predicted through VADAR. Accessible surface area to water molecules was measured as fractional ASA range (0 to 1). Results of ASA showed that the major portion of the protein is hydrophobic, having ASA scores less than 0.8 signifying tight folding of the protein (Figure 4).

Predicted drug binding sites (DBS)

DBS analysis predicted functional ligand binding sites. All of these binding site residues of both models were also confirmed when predicted

through different methods including COFACTOR, FINDSITE, TM-SITE and ConCavity. In the case of COFACTOR His 4, Gln 5, Asn 6, Lys 7, Ala 8, His 20, Leu 21, Asp 38, Cys 55, Cys 57, Gln 81, Tyr 82, Met 83, Glu 84, Asn 86, Tyr 88, Tyr 89, Arg 92, Gln 94, Asp 108, Cys 115, His 117, Ile 123, Thr 124, Ala 125, Gly 126, Gly 127, Leu 130 and Ala 132 were predicted as potential binding sites for 001 and 002 (Figure 5A and 5B). In contrast, FINDSITE predicted Phe 2, Gly 3, His 4, Gln 5, Asn 6, Lys 7, His 20, Leu 21, Ala 22, Thr 24, Glu 25, Asp 26, Leu 27, Ser 29, Val 30, Asn 31, Val 33, Asp 38, Val 41, Thr 42, Arg 45, Gln 47, Ser 51, Lys 69, Tyr 83, Asn 86, Tyr 89, Ala 91, Arg 92, Gly 107, Asp 108, Cys 109, Ile 123, Thr 124, Ala 125, Gly 126, Gly 127, Gly 129, Val 131 and Phe 133 as potential binding sites (Figure 5C and 5D).

His 20, Leu 21, Asp 38, Cys 55, Cys 57, Met 83, Glu 84, Arg 92, Gln 94, Asp 108, Cys 115, His 117, Ile 123, Thr 124, Ala 125 and Gly 126 were predicted as potential drug binding sites of 001 and 002 according to TM-SITE results (Figure 6A and 6B), while ConCavity results predicted Gln 5, Asn 6, Lys 7, Ala 8, Asn 18, Tyr 19, His 20, Leu 21, Asp 38, Asp 50, Tyr 62, Tyr 71, Tyr 83, Asn 87, Tyr 88, Tyr 89, Gln 94, Gly 100, Gly 102, Phe 103, Ala 104, Pro 106, Gly 107, Cys 109, Gly 110, Ile 112, Arg 114, Ile 123, Thr 124, Gly 126, Gly 129 and Val 131 as potential drug binding residues (Figure 6C and 6D).

Intrinsic disorder region prediction

To understand protein folding and function, precise prediction of a protein's tendency to have intrinsic disordered regions is an essential step. The current investigation used four software packages to identify disordered positions based on the majority voting by these software tools. In this prediction server, a residue is considered as disordered if its score is above 0.5. Based on the real valued disorder propensities generated by MetaPrDOS [25], DisEMBL [26], PSIPRED [27] and IUPred [28] analysis, disordered regions are defined as regions with scores above the threshold line (0.5); however, scores below this line specify a degree of flexibility. In the case of 001 and 002, regions GFGH (1-4) and EAMEQ (145-149) were found in ID regions.

Subcellular localization

Subcellular localization prediction of PV2A protease specifies its localization in the cytoplasmic membrane with no signal peptide and transmembrane helix.

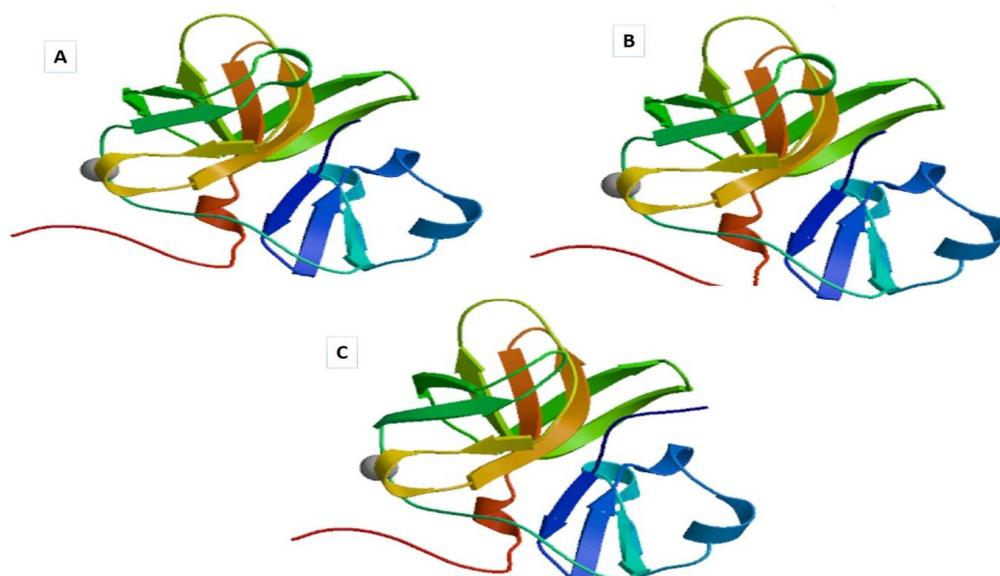


Figure 2: 3D model generation through Swiss model software. A and B indicate the model produced for poliovirus 2A protease for sample 001 and 002 respectively. Whereas, C is the template model (1z8r) used for analysis

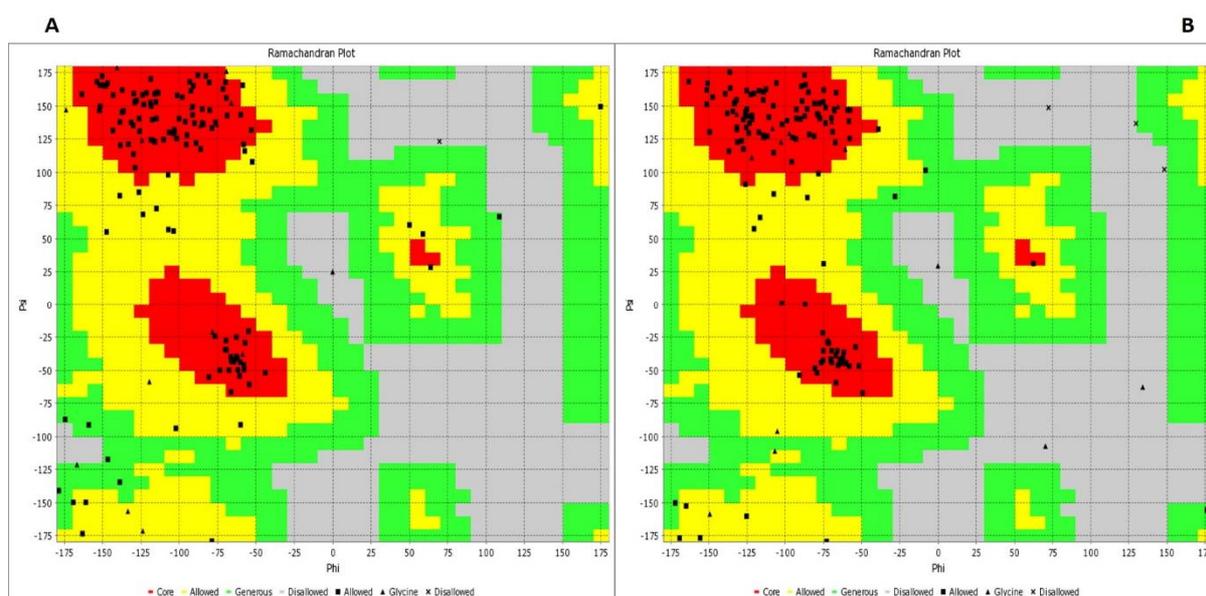


Figure 3: Ramachandran plot analysis of 001 (A) and 002 (B) protein. Red colour indicates core region, yellow indicates allowed region, green is allowed region, grey is disallowed region

Table 1: QMEAN global scores for 001 002, and template models obtained from Swiss modelling

Scoring function term	Z-score (Reliability range 0-1)		
	001	002	Template 1Z8R
QMEAN6 score	-4.16	-3.55	-4.18
QMEAN4 score	-4.12	-3.41	-3.92
All-atom pairwise energy	-2.23	-2.28	-2.48
C_beta interaction energy	-0.71	-0.84	-1.43
Solvation energy	-1.51	-1.56	-1.91
Torsion angle energy	-3.57	-2.82	-3.17
SS Agreement	-1.38	-1.04	-1.34
ACC Agreement	-0.91	-1.24	-1.31

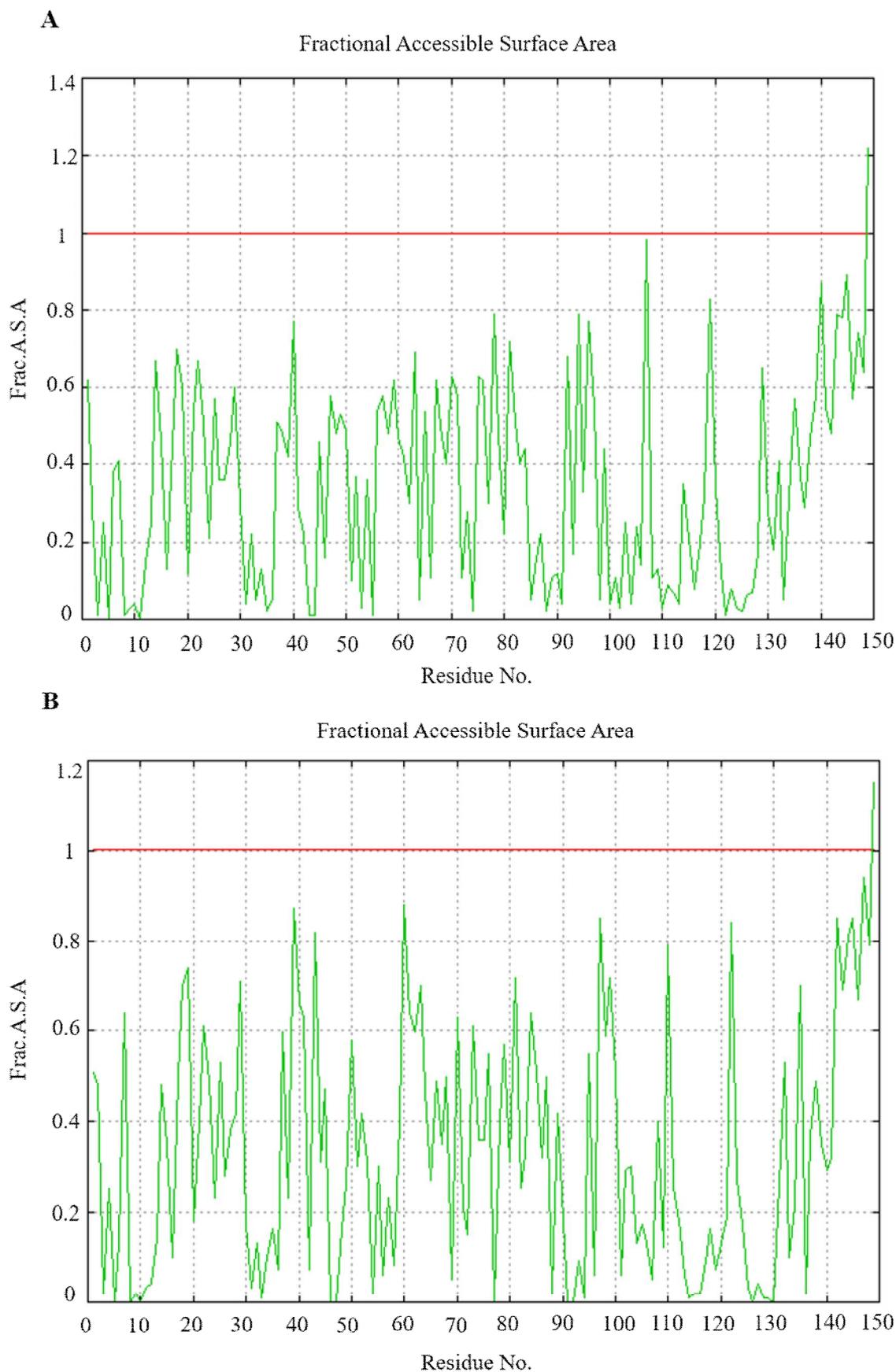


Figure 4: Accessible surface area (ASA) prediction of 001 (A) and 002 (B) protease models using VADAR server

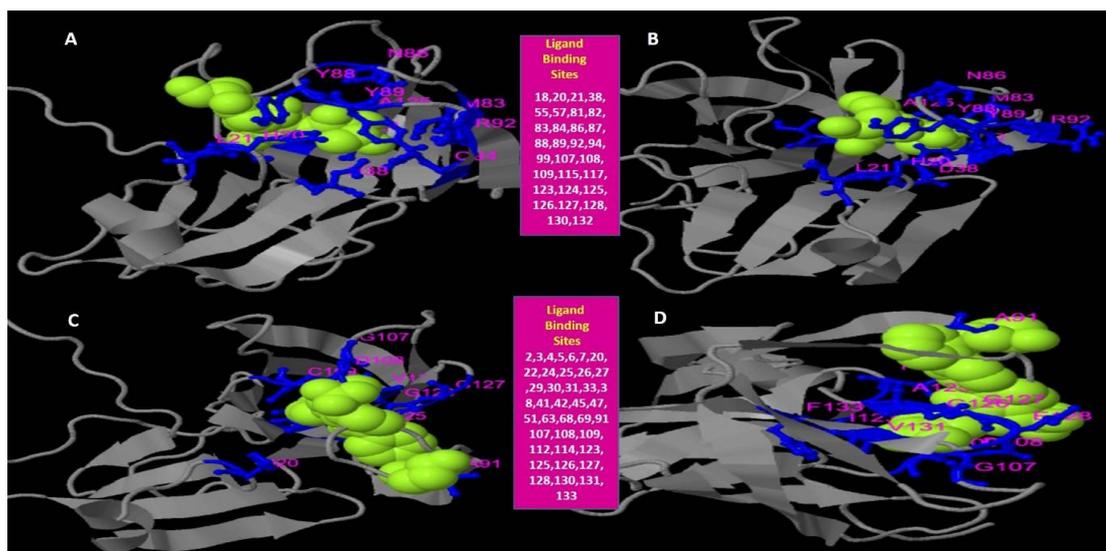


Figure 5: Ligand binding sites prediction through COACH server. A and B are the predictions from COFACTOR of 001 and 002 respectively, while C and D are the results of FINDSITE methods. Ligand binding sites are shown in purple colour

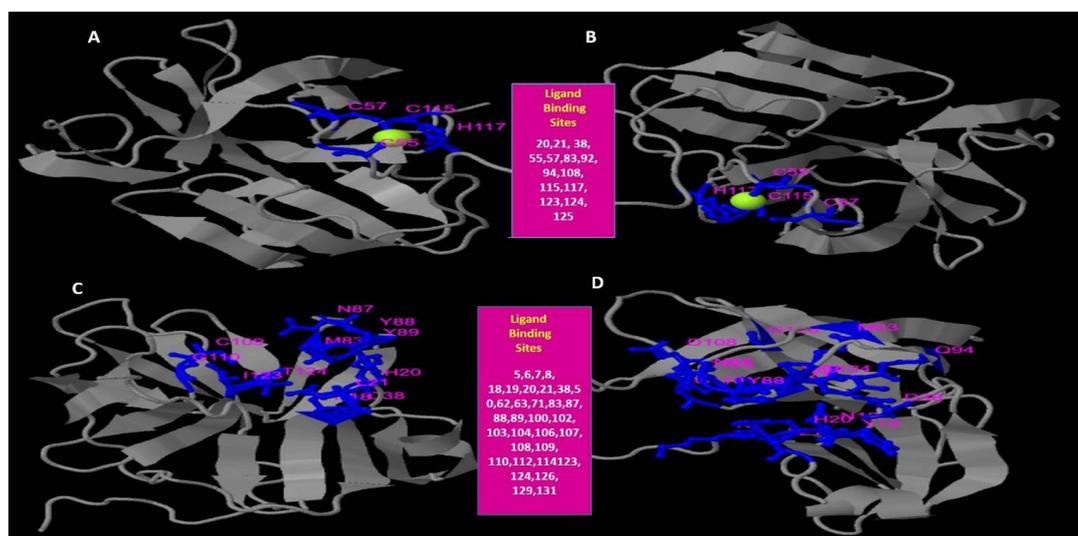


Figure 6: Ligand binding sites prediction through COACH server. A and B are the predictions from TM-SITE of 001 and 002 respectively while C and D are the results of ConCavity. Ligand binding sites are shown in purple colour

Protein stability with mutations

The protein stability change upon mutation was assessed by using I-Mutant 2.0, STRUM and EASE-MM web-servers. The results showed the effect of mutants on stability of protein with reliability index at pH 7.0 and temperature 25 °C. Mutations (A12G, D36N, S134T, I136V) used in this study indicated a decrease in stability of the protein.

Hydrophobicity

Hydrophobic interactions play an important role in drug design involving biophysical events for instance protein-ligand or protein-protein binding.

Hydrophobic regions of the selected 001 and 002 proteins were determined with PEPTIDE 2 and TM-Finder. These software predicted the calculated hydrophobicity and helicity values, along with the predicted TM-segments of the proteins (Figure 8). In case of both proteins most residues were found to be hydrophobic (34.9 %) as compared to hydrophilic residues (11.4 %).

DISCUSSION

Poliovirus 2A protease is involved in many physiological processes, making it an important target in the cysteine protease superfamily of proteins. However, the absence of a high resolution crystal structure of poliovirus 2A

protease is a limitation in understanding atomic details [10]. In this study, we took advantage of the PV2A^{pro} primary structural similarity to other viral proteases, specifically that from Coxsackievirus B4, and performed an *in silico* analysis of its structure. Poliovirus 2A protease sequences (001 and 002) were taken from our previous work where we have reported the mutations in this protein.

No previous studies have reported mutations detected in the poliovirus 2A protease. In this study, we performed a computational analysis on these mutations to determine their impact on protease structure, function and stability. Intrinsic disordered region prediction results proved it to be suitable for drug binding sites, with a low number of residues lying in the disordered regions. Moreover, hydrophobicity results predicted that these proteases have more hydrophobic residues compared to hydrophilic residues. Important drug binding sites were predicted using the COACH server for these mutation containing proteases. These important residues include His20, Leu21, Asp38, Cys55, Cys57, Met83, Tyr88, Tyr89, Asp108, Cys109, Cys115, His117, Ile123, Thr124, Ala125 and Gly126.

Previous studies have reported that His20, Cys109, Cys55, Cys57, Cys115 and His117 are involved in structural maintenance and catalytic activity of 2A^{pro} [40]. Proteinase 2A of human rhinovirus 2 (HRV2 2A) Cys residues and His114 are important residues involved in coordinating Zn²⁺. Moreover, Gly123, Gly124, Thr121 and Cys101 are suggested to form part of the construction of the substrate binding pocket and to offer a suitable setting for the active site of His18, Asp35 and Cys106 [41].

Previously reported mutations (A12G, D36N, S134T, I136V) in poliovirus 2A protease were subjected to predict single nucleotide polymorphisms along with their impact on protease stability. All mutations result in a decrease of protein stability, thus destabilizing the protein structure. Defects in *cis* cleavage with the minute plaque phenotype have been observed in poliovirus point mutants (G60R, S66F, L39F, N18K, D135N, V119M, A22T, C55Y, and S134L) when the poliovirus 2A gene was expressed in yeast [42]. Previously it was reported that single amino acid mutations at the His20 and Cys109 residues led to complete loss of the proteolytic activity of 2A^{pro} [43].

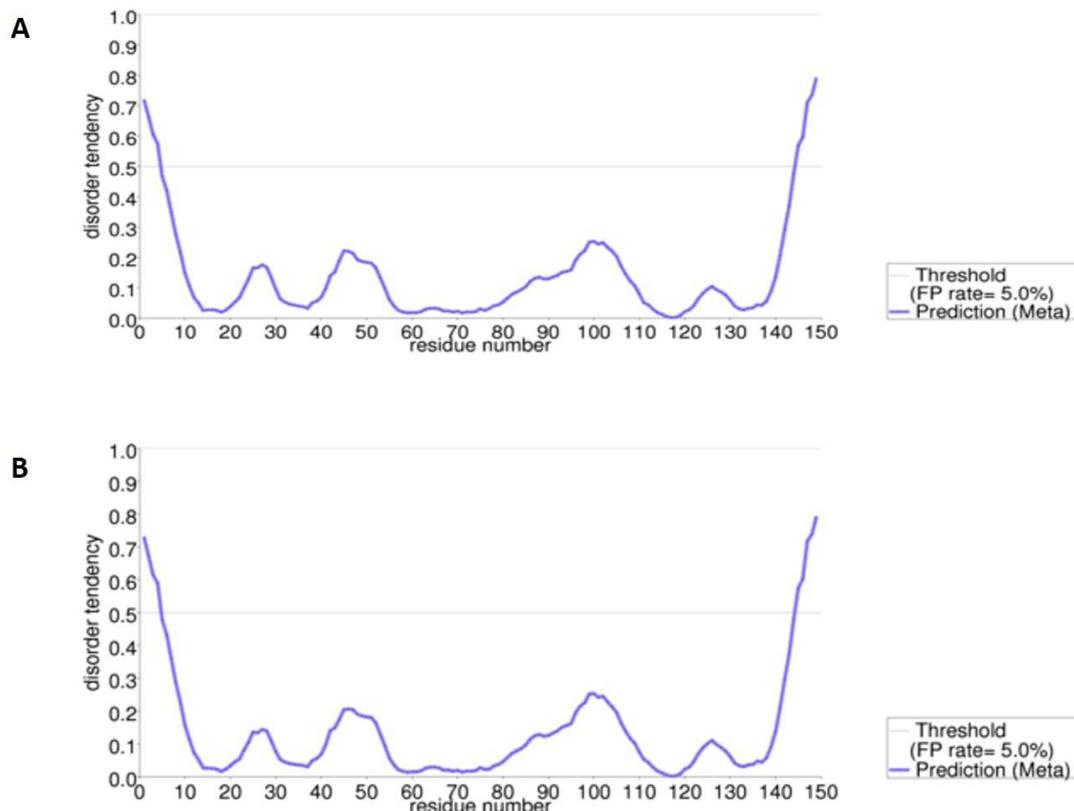


Figure 7: Intrinsic disorder regions of 001 (A) and 002 (B). Scores above the grey line are the disorder region, whereas scores below the grey line show degree of flexibility

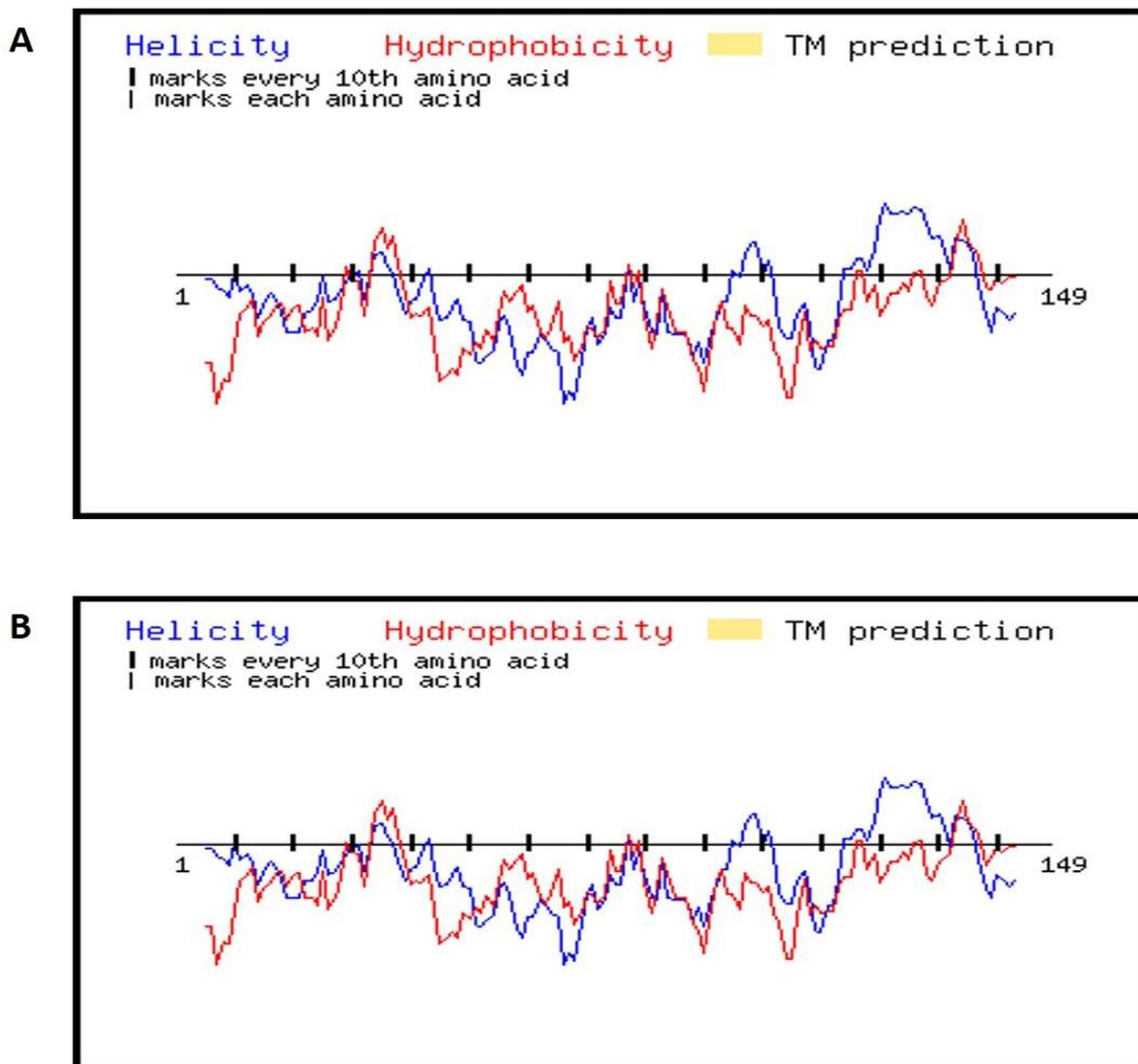


Figure 8: Hydrophobicity results of 001 (A) and 002 (B) model. Red amino acids are amino acids in a TM segment and Blue amino acids are charged amino acids in the TM segment

Important ligand-protein interactions have been identified following docking investigations in this study, providing potential candidates for drug targeting, the potency of which may be increased following relatively simple structural alterations.

CONCLUSION

A structure-function investigation of the poliovirus 2A protease is presented in the study. By using the crystal structure of coxsackievirus B4 as a template, the three-dimensional structure of the poliovirus 2A protease was predicted. I-TASSER and COACH results identified potential binding sites in the protein structure. Use of computational approaches to identify more potential drug binding sites in the conserved regions will lead to better drug design methods in the years ahead. This computational approach may be combined with experimental methods,

thus contributing to functional interpretation of protein structures.

DECLARATIONS

Acknowledgement

The authors would like to acknowledge the guidance and examination committee members, including Nasar Virk, Hussnain Ahmed Janjua and Muhammad Arshad Malik, for their feedback on the work; Faryal Mehwish Awan and Aqsa Ikram for providing valuable suggestions; and National University of Sciences & Technology (NUST) Islamabad for providing financial support.

Conflict of interest

The authors declare that no conflict of interest is

associated with this study.

Contribution of authors

We declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by the authors.

REFERENCES

- De Chassey B, Meyniel-Schicklin L, Aublin-Gex A, André P, Lotteau V: New horizons for antiviral drug discovery from virus-host protein interaction networks. *Curr Opin Virol* 2012; 2: 606-613.
- Landry Y, Gies JP. Drugs and their molecular targets: an updated overview. *Fundam Clin Pharmacol* 2008; 22: 1-8.
- Eigen M. Viral quasispecies. *Scientific American-American edition*. 1993; 269:42-49.
- Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. Rapid evolution of RNA genomes. *Science* 1982; 215: 1577-1585.
- Domingo E, Menendez-Arias L, Holland JJ. RNA virus fitness. *Rev Med Virol* 1997; 7: 87-96.
- Hogle J, Chow M, Filman D. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 1985; 229: 1358-1365.
- Kitamura N, Semler BL, Rothberg PG, Larsen GR, Adler CJ, Dorner AJ, Emini EA, Hanecak R, Lee JJ, van der Werf S, et al. Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* 1981; 291: 547-553.
- Kuyumcu-Martinez NM, Joachims M, Lloyd RE. Efficient cleavage of ribosome-associated poly (A)-binding protein by enterovirus 3C protease. *J Virol* 2002; 76: 2062-2204.
- Mueller S, Wimmer E, Cello J. Poliovirus and poliomyelitis: A tale of guts, brains, and an accidental event. *Virus Res* 2005; 111: 175-193.
- Petersen JF, Cherney MM, Liebig HD, Skern T, Kuechler E, James MN. The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J* 1999; 18: 5463-5475.
- Crowder S, Kirkegaard K. Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nat Genet* 2005; 37: 665-666.
- National Research Council. *Exploring the role of antiviral drugs in the eradication of polio: Workshop report*. National Academies Press, 2006.
- Younus A, Bhatti MF, Ahmad T, Virk N, Arshad M, Janjua HA. Protein structure modelling, ligand docking and active site analysis of mutated poliovirus 2A protease gene isolated from the blood of Pakistani polio infected patients. *Int J Adv Sci Engine Technol* 2016; 4.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014; 42: W252-W258.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 1963; 7: 95-99.
- Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res* 2009; 37: W510-W514.
- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007; 35: W407-W410.
- Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* 2003; 31: 3316-3319.
- Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008; 9: 40.
- Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 2012; 40: W471-W477.
- Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res* 2004; 32: D120-D121.
- Capriotti E, Fariselli P, Casadio R. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics* 2004; 20: 163-168.
- Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 2016; 32: 2936-2946.
- Folkman L, Stantic B, Sattar A, Zhou Y. EASE-MM: Sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *J Mol Biol* 2016; 428: 1394-1405.
- Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 2008; 24: 1344-1348.
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure* 2003; 11: 1453-1459.
- Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 2013; 41: W349-W357.
- Deng X, Eickholt J, Cheng J. PreDisorder: ab initio sequence-based prediction of protein disordered regions. *BMC Bioinformatics* 2009; 10: 1.
- Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, Muskat BL, Cuticchia AJ. TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci* 2001; 10: 212-219.
- Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific

- substructure comparison and sequence profile alignment. *Bioinformatics* 2013; 29: 2588-2595.
31. Wang W, Malcolm BA. Two-stage polymerase chain reaction protocol allowing introduction of multiple mutations, deletions, and insertions, using QuikChange TM site-directed mutagenesis. *In Vitro Mutagenesis Protocols* 2002:37-43.
 32. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comp Biol* 2009; 5: e1000585.
 33. Skolnick J, Brylinski M. FINDSITE: a combined evolution/structure-based approach to protein function prediction. *Brief Bioinform* 2009; 10: 378-391.
 34. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res* 2011; 72: 162-177.
 35. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics* 2006; 64: 643-651.
 36. Shen HB, Chou KC. Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. *J Biomol Struct Dyn* 2010; 28: 175-186.
 37. Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* 2009; 37: W465-W468.
 38. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; 305: 567-580.
 39. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2010; 27: 343-350.
 40. Yu SF, Lloyd RE. Characterization of the roles of conserved cysteine and histidine residues in poliovirus 2A protease. *Virology* 1992; 186: 725-735.
 41. Sommergruber W, Seipelt J, Fessl F, Skern T, Liebig HD, Casari G. Mutational Analyses Support a Model for the HRV2 2A Proteinase. *Virology* 1997; 234: 203-214.
 42. Barco A, Ventoso I, Carrasco L. 1997. The yeast *Saccharomyces cerevisiae* as a genetic system for obtaining variants of poliovirus protease 2A. *J Biol Chem* 1997; 272:12683-12691.
 43. Shuyuan FY, Lloyd RE. Identification of essential amino acid residues in the functional activity of poliovirus 2A protease. *Virology* 1991; 182: 615-625.