

RESEARCH ARTICLE

Mapping allosteric pathway in NIa-Pro using computational approach

Rashmi Panigrahi^{1,*}, Senthilkumar Kailasam^{2,*}

¹ Department of Biochemistry, University of Alberta, Edmonton, T6G 2H7, Canada

² Canadian Centre for Computational Genomics, McGill University, Montréal, H3A 0G1, Canada

* Correspondence: panigrah@ualberta.ca; senthil.duraikannukailasam@mcgill.ca

Received March 7, 2021; Revised January 10, 2022; Accepted February 21, 2022

Background: Computer simulation studies complement *in vitro* experiments and provide avenue to understand allosteric regulation in the absence of other molecular viewing techniques. Molecular dynamics captures internal motion within the protein and enables tracing the communication path between a catalytic site and a distal allosteric site. In this article, we have identified the communication pathway between the viral protein genome linked (VPg) binding region and catalytic active site in nuclear inclusion protein-a protease (NIa-Pro).

Methods: Molecular dynamics followed by *in silico* analyses have been used to map the allosteric pathway.

Results: This study delineates the residue interaction network involved in allosteric regulation of NIa-Pro activity by VPg. Simulation studies indicate that point mutations in the VPg interaction interface of NIa-Pro lead to disruption in these networks and change the orientation of catalytic residues. His142Ala and His167Ala mutations do not show a substantial change in the overall protease structure, but rather in the residue interaction network and catalytic site geometry.

Conclusion: Our mutagenic study delineates the allosteric pathway and facilitates the understanding of the modulation of NIa-Pro activity on a molecular level in the absence of the structure of its complex with the known regulator VPg. Additionally, our *in silico* analysis explains the molecular concepts and highlights the dynamics behind the previously reported wet lab study findings.

Keywords: NIa-Pro; VPg; simulation; residue interaction network; allostery

Author summary: Allosteric control of enzymes is a key regulatory process, ubiquitous in all kingdoms of life. It involves the binding of the specific regulatory molecule (either an activator or an inhibitor) to a location distal from the enzyme's active site. Our research provides a simple *in silico* methodology for gaining insight into the allosteric pathways that regulate specific enzyme activity. In the presence of a 3D structure of the enzyme and the knowledge of specific interaction sites with its regulator, this methodology can be applied to understand enzyme regulation, even in the absence of the structure of the enzyme-regulator complex.

INTRODUCTION

Potyvirus are the largest genus of RNA viruses that infect crop plants, making them concerns to farmers worldwide [1]. The pepper vein banding virus (PVBV) belongs to the potyvirus group and infects Solanaceae plants, such as chilli and capsicum. It also encodes for a large 340–370 kDa polyprotein [2]. The 5' terminus of the genomic RNA is covalently associated via a

phosphodiester bond to the viral protein genome linked (VPg) [3]. VPg acts as a primer for the initiation of positive and negative strand RNA synthesis during viral replication [4]. The polyprotein is subsequently cleaved by three proteinases, namely P1-protease (P1-Pro), Helper component-protease (HC-Pro), and nuclear inclusion protein-a protease (NIa-Pro), all of which are encoded in the virus. This proteolytic cleavage, or polyprotein processing, occurs after translation at

different stages of the viral life cycle [5]. NIa-Pro is a serine-like cysteine proteinase with an active site made up of a catalytic triad consisting of His46, Asp81, and Cys151 [3,6,7]. This proteinase recognizes heptapeptide sequences and cleaves at Gln/Ala, Gln/Ser, or Gln/Thr with amino acids flanking at either side of the cleavage site. Contrary to P1-Pro and HC-Pro that demonstrate *cis* activity only, NIa-Pro has both *cis* and *trans* cleavage activity [8,9]. P1-Pro and HC-Pro cleave at their own C-termini. However, NIa-Pro can cleave at seven different sites along the translated polyprotein [10]. One such *cis* cleavage site exists between VPg and NIa-Pro, which is known to be suboptimal [11] as a glutamate is found in the P1 position of the recognition sequence instead of a glutamine. Any mutation at the cleavage site leads to defective viral replication [11,12]. Therefore NIa-Pro exists as VPg-Pro, with an N-terminal VPg and C-terminal NIa-Pro. VPg is an intrinsically disordered protein and acts as a protein interaction hub [13,14]. It harbors the walker motifs A and B, which is functional only upon interaction with NIa-Pro [14]. Furthermore, the deletion of the N terminal residues of VPg led to a decrease in protease activity as the interaction of VPg with NIa-Pro is impaired [13]. Surface plasmon resonance studies have shown that the affinity of VPg for NIa-Pro is drastically decreased upon deletion of 113 residues from the N-terminus of VPg. Additionally, enzyme kinetic analyses have shown that VPg increases enzymatic turnover when associated with NIa-Pro [15]. The increase in protease activity of NIa during interaction with VPg can be described as an allosteric event.

Allosteric regulation, in which the perturbation at one site of the protein affects the distal catalytic site, is a widely observed phenomenon [16]. This mechanism helps to regulate the activity and dynamic adaptability of proteins while they perform their functions [17]. Protein structures have evolved in the constraint of their designed functions and need to be robust against mutations [18,19]. Hence, establishing good communication paths between distal sites and the catalytic site of the protein bring about the precision in function. Some of these perturbations may not induce visible conformational changes; however, they do affect the catalytic site [20,21].

Allosteric communications are best described by residue-residue interaction networks. Network models that give rise to a high degree of interaction cooperativity based on three-dimensional structure have been successfully used to delineate mode of allosteric communication on a molecular level [22–24]. Elucidating these communication networks are challenging as they involve multiple modal communication pathways between the binding site and distal active site. These

trajectories are not ascertainable by a single crystal structure, which provides a snapshot of the interactions [23]. Hence, theoretical approaches, such as molecular dynamics (MD) simulation studies are precious tools used in obtaining an atomistic insight of the above phenomenon over a time period [25]. The comparative analysis of specific patterns of inter-residue fluctuations can aid in identifying plausible mechanisms of coordinated conformational changes in the protein due to specific mutations or binding of the allosteric partner. The shortest path is identified as the path in which two distal site residues are connected by the minimal number of intermediate residues. These intermediate residues are centrally conserved and are known as conserved interconnectivity determinants (CICD). Mutations in these residues can cause changes in the optimal communication pathway [18], leading to a disruption of protein function. This explains why experimental evidences suggest that a number of single site mutants have little or no impact on the protein function, whereas mutations in key residues perturb the function to a larger extent [18,26].

The current study was carried out with an aim to delineate the communication pathway in NIa protease between its VPg binding site and the catalytic site. Enzyme kinetics have shown that Ser129, a surface exposed residue on NIa-Pro, is essential for enzyme activity, despite it being located 16.9 Å from the Cys151 C α (catalytic cysteine) [15]. This serine is known to be phosphorylated by a host cell kinase. NIa-Pro Ser129Asp (phospho-mimetic mutant) demonstrate a drastic loss in activity compared to the wild type as observed during enzymatic analysis, suggesting that this phosphorylation could be regulating the protease activity of NIa-Pro [15]. This suggests that changes in the distal Ser129 is being relayed to the active site through a residue-to-residue interaction pathway.

In addition, Ser129 lies in close proximity to Trp143, which lies distal to the Cys151 (14.47 Å apart). It had been shown using fluorescence studies that Trp143 is present on the VPg binding interface of NIa-Pro [15]. Furthermore, the mutation of this tryptophan to alanine did not affect the binding of the substrate, but rather abrogated the cleavage reaction and the subsequent product release. Simulation studies have shown that the Trp143Ala mutation altered the orientation of Cys151 (causing it to move away from the catalytic residues His46 and Asp81) [15]. The WC loop that connects Trp143 with Cys151 was postulated to be crucial for the substrate-product interaction [15,27,28]. It has also been speculated that the binding of VPg on the face of NIa-Pro containing Trp143 might influence the structural integrity of the WC loop, thus aid in the regulation of catalytic activity. Trp143 is connected to Cys151 via

noncovalent interactions mediated by His142 and His167. These residues are conserved across all members of the *Potyvirus* genus [15].

In order to investigate allostery on a molecular level in NIa protease and its mutants, MD simulation studies followed by analysis of residue interaction networks (RIN) have been undertaken. The comparison of allosteric communication pathways between different mutants elucidated in this study highlights the diverse means adopted by the protease mutants to perform their enzymatic function. In the absence of a three-dimensional model of VPg and/or NIa-VPg complex, our findings elucidate the communication pathway and key hub residues between the VPg binding site, as well as the catalytic site and key hub residues of NIa protease, thus mapping the residue-residue interaction network that might be involved in the allosteric regulation of NIa-Pro.

RESULTS

Analysis of trajectory

In order to gain molecular level insight into the residue-residue interaction network of NIa-Pro and its mutants, MD simulation studies have been undertaken. The results provide an in-depth understanding of the

structure and dynamics of these proteins. I-TASSER was used to generate the best model of NIa protease for this study. NIa protease from PVBV shares a sequence identity of 47% and 45% with respect to the TEV protease and TVMV protease, respectively. The RMSD between the model generated using ITASSER and the crystal structure of TEV protease (PDB: 1LVM) was 0.90 Å, while that with the crystal structure of TVMV protease was 1.60 Å. Furthermore, the Ramachandran plot analyses showed that 90% residues were present in the favored regions (Supplementary Fig. S1). The RMSD of the C α atoms for each simulation was calculated using the starting structure as the reference structure to assess the overall stability (Fig. 1A). The RMSD values were found to be 4.76 Å, 3.60 Å, 3.42 Å, 3.18 Å, and 3.15 Å for NIa_wt, NIa_S129, NIa_H142, NIa_W143, and NIa_H167 systems, respectively. The RMSD values of the mutants were lower compared to the wild type, which correlates with the previously reported findings [15]. An analysis of C α atomic fluctuations for each of the systems showed that the N and C termini exhibited higher RMSF compared to other regions. The C-terminal that comprises of 200–242 residues demonstrated extensive fluctuations in NIa_wt and NIa_H167 systems (Fig. 1B). NIa-Pro wild type and H167A mutant have 35 hydrogen bonds in each of their systems. The S129A, W143A, and H142A mutants have

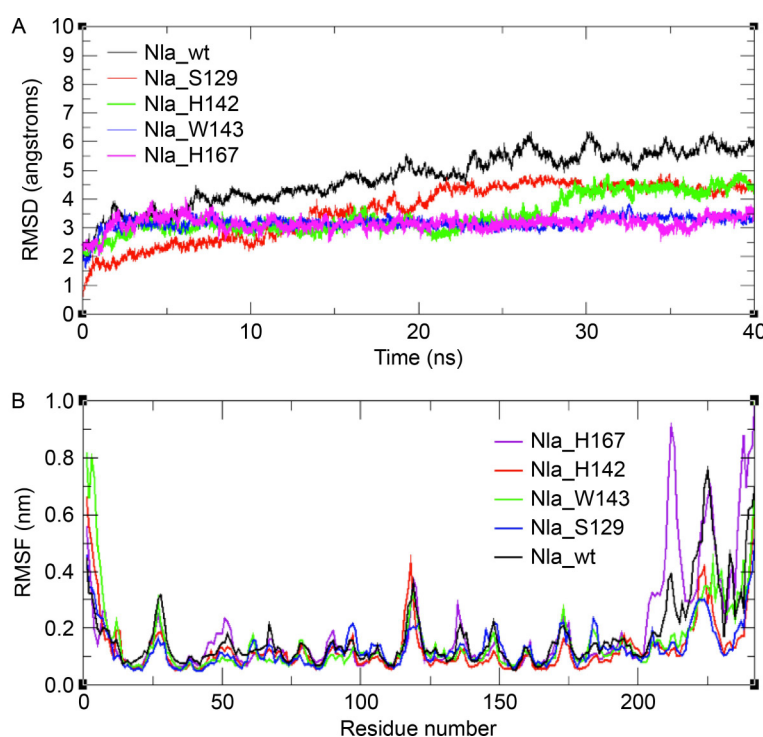


Figure 1. Analysis of MD simulation trajectories for NIa-Pro and the mutants. (A) Root mean square deviation (RMSD) plots for all the systems. (B) The atomic positional fluctuation (RMSF) plots obtained for each of the above system.

similar quantities of hydrogen bonds *i.e.*, 44, 43, and 41, respectively. The total number of salt bridges in Nla-Pro and H142A mutant is 16 and 19, respectively. The S129A, W143A, and H167A mutants have similar quantities of salt bridges *i.e.*, 23, 23, and 22, respectively.

Essential dynamics

The harmonic and large-scale motions in various systems occurring in the essential subspace were analyzed using principal component analysis. The overall motions observed in each system were decomposed into eigenvectors. The first vector accounted for the majority (>60%) of the total movement in all the systems. The conformation sampled in the subspace spanned by the first two eigenvectors was visualized using two-dimensional projection plots designed for all trajectories (Fig. 2). The conformational space sampled was found to be diffused over a larger area in the wild type compared to the mutants. However, multiple local minima were observed in all systems. In the case of the Nla_W143 system, the conformational spread was comparatively restricted and the number of local minima was reduced.

The direction and magnitude of movement captured by the first eigenvector in all systems were displayed using porcupine plots (Fig. 3). The N and C terminal regions of the proteins in all the systems showed higher motion compared to other regions. In the case of the wild type, there was random movement observed in the region encompassing residues 206–242, as observed in

the rms fluctuations. Similar movement was observed in the Nla_S129 system. In the Nla_H142 system, the motions were relatively reduced, while the C-terminal seemed to move away from the protein core in the Nla_H167 system. Interestingly, the porcupine plot showed relatively restricted movement of the C-terminal region in the Nla_W143 system.

Residue interaction networks

Different interaction networks for communication

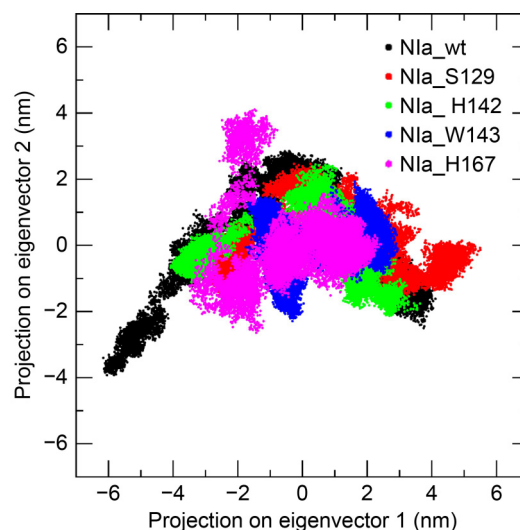


Figure 2. Essential dynamics—2D projection of individual trajectories with their first 2 eigenvectors: vector 1 and vector 2.

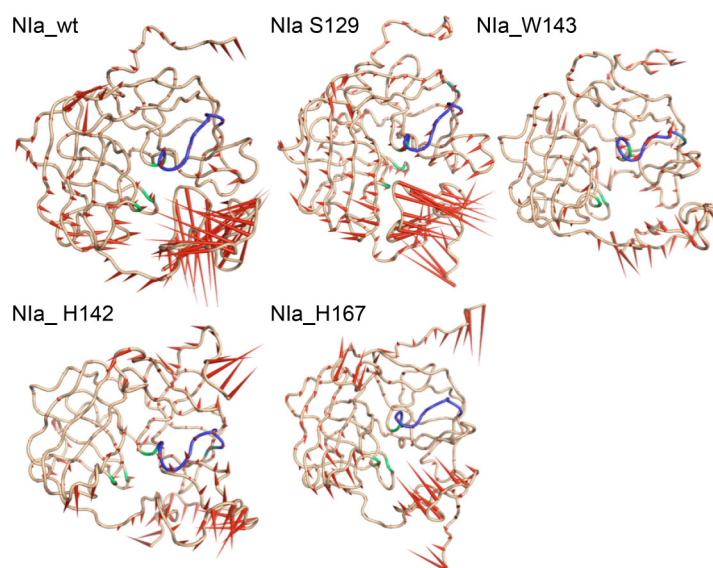


Figure 3. Porcupine plots for the wild type and the mutant systems are presented in a cone model. The length and direction of the cone (red) represents the magnitude and direction of motion.

between the VPg binding site and catalytic site were evaluated using the last 5 ns averaged structures of MD trajectories. A comparative analysis between the RINs from the mutant systems and that of the wild type showed interesting differences. In Nla_{wt}, the WC loop and several other noncovalent interaction networks mediated the communication from Trp143 to Cys151. The shortest noncovalent interaction network was Trp143-His142-His167-Cys151 (Supplementary Fig. S2). This network was distorted and the position of the catalytic site residues and/or their side chains were affected and are discussed below.

Nla_{S129} system

With the mutation of Ser129 to alanine, the side chain interaction with Trp143 is lost. A slight displacement of the side chains of the three active site residues was observed (Fig. 4 and Supplementary Fig. S3). The shortest communication pathway was Trp143-His142-Asn177-Tyr178-His167-Gly152-Cys151.

Nla_{W143} system

When W143 was mutated to alanine, flipping of the side

chain of Cys151 was observed. The imidazole ring of His46 was orientated into the active site pocket (Fig. 5 and Supplementary Fig. S4). Although the side chain of Trp143 was lost with the mutation to alanine, the interaction with Ser129 was maintained. However, a displacement of His142 and a subsequent rearrangement of the interaction with His167 through Asn177 were observed. Thus, the shortest pathway noted was Ala143-His142-Asn177-His167-Cys151. The side chain of Cys151 flipped away from the active site pocket and the imidazole ring of His46 faced into the pocket.

Nla_{H142} system

When His142 was mutated to alanine, a flip in the side chain of Cys151 was observed and the two catalytic residues His46 and Asp81 were displaced (Fig. 6 and Supplementary Fig. S5). Furthermore, the shortest interaction network observed in the wild type was disrupted and additional pathways were established with the involvement of Asn177 and the residues of the WC loop. Additionally, the noncovalent interaction between Trp143 and Ser129 was disrupted because of the change in orientation in the Ser129. Moreover, the new interaction networks in Nla-Pro (H142A) and Nla-Pro

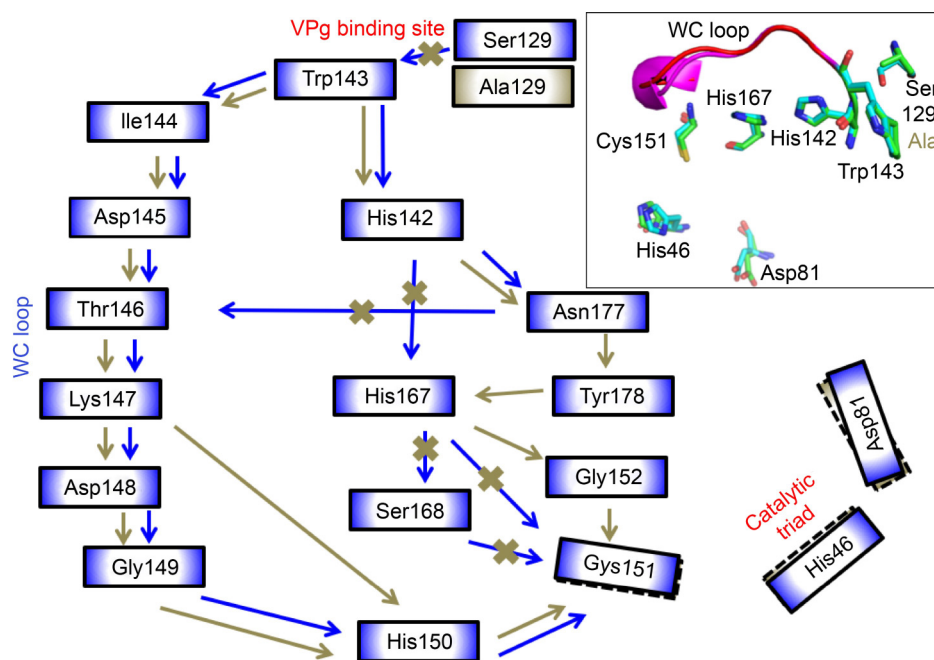


Figure 4. Schematic diagram showing the residue interaction networks in the Nla-Pro wild type and Nla_{S129} system. The inset shows the overlay of WC loop conformations (wild type: magenta, and mutant: red), orientations of the catalytic triad along with the two crucial residues, His142 and His167 (wild type: green sticks, and mutant: cyan). The residue communication pathway from the VPg interaction interface to the catalytic site for the wild type is shown using blue arrows. The modified pathway with the Ser129Ala mutation is shown using golden green arrows. The native communication pathway between the two residues that is no longer effective in the mutant is marked by golden green crosses. The communication pathway is based on hydrogen bonding, hydrophobic, and main-chain interactions, which is also demonstrated in ligplot in the supplemental.

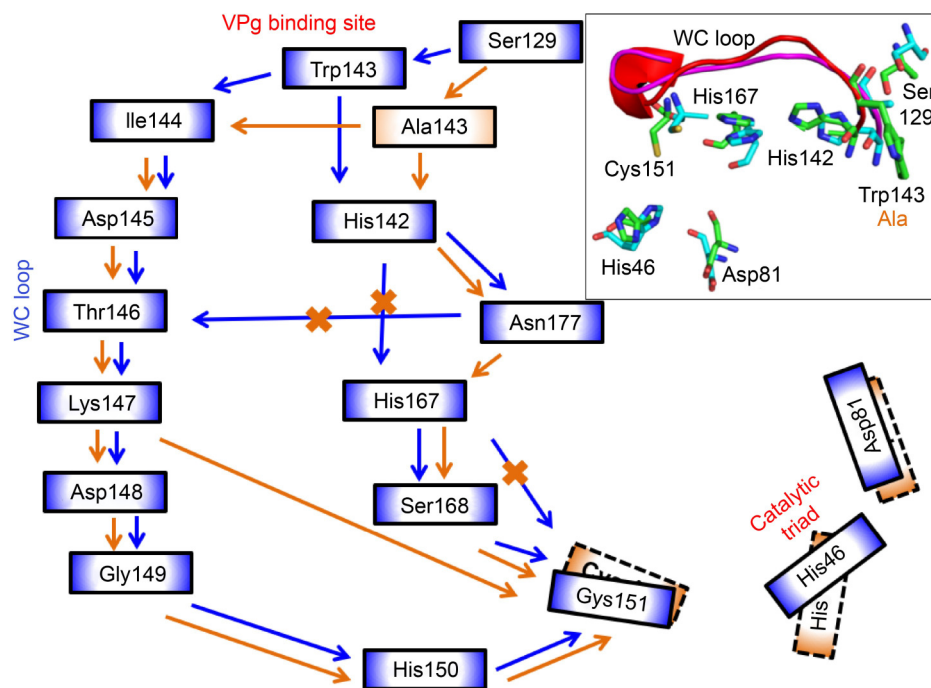


Figure 5. Schematic diagram showing the residue interaction networks in the Nla_W143 system with reference to the wild type. The inset shows the overlay of WC loop conformations (wild type: magenta, and mutant: red), orientations of the catalytic triad (wild type: green sticks, and mutant: cyan), and the two crucial residues, His142 and His167 (wild type: green sticks, and mutant: cyan). The residue communication pathway from the VPg interaction interface to the catalytic site for the wild type is shown using blue arrows. The modified pathway due to the Trp143Ala mutation is shown using orange arrows. The native communication pathway between the two residues that is no longer effective in the mutant is marked by orange crosses. The communication pathway is based on hydrogen bonding, hydrophobic, and main-chain interactions, which is also demonstrated in ligplot in the supplemental.

(H167A) are longer than the non-covalent interaction pathway (Trp143-His142-His167-Cys151) in wild type Nla-Pro. The conformation of the WC loop is different in His142Ala Nla-Pro compared to the wild type (Fig. 6).

Nla_H167 system

For the system where His167 was mutated to alanine, although the interaction between Ser129 and Trp143 was maintained, the interactions involved in the shortest pathway observed in the wild type were disrupted due to absence of side chain interactions of A167 (Fig. 7 and Supplementary Fig. S6). Furthermore, a slight displacement of Asp81 and orientation change of the imidazole ring of His46 was noted. Therefore, another pathway from Trp143 involving WC loop residues exists (Trp143-His142-Asn177-Thr146-Lys147-Asp148-Gly149-His150-Cys151).

DISCUSSION

Understanding allosteric regulation in proteins have

opened doors for the design of allosteric modulators of therapeutically important targets [29,30]. This is due to the improved selectivity offered by allosteric sites in contrast to active sites. Perturbation at these allosteric sites induced by mutation [31], post-translational modifications [32], ligand binding, etc. can induce changes in catalytic activity [33], structure [34], or oligomeric state [35]. The traditional method of mapping allosteric sites involves the creation of mutants or deletion of specific amino acids stretches distal to the catalytic site, followed by subsequent analysis of protein activity. An upregulation or downregulation of activity indicates the involvement of mutated residue(s) in allosteric regulation [36,37]. Directed evolution and high throughput screening methods have also been used [38]. These methods are limited by enzyme signal detection. After the experimental validation of allostery, molecular dynamics (MD) simulations have been performed with either a docked or crystal structure of the complex. The simulation studies capture the binding induced conformational changes and the change in free energy is calculated using MMPB/GBSA algorithms. In the last decade, computational approaches have

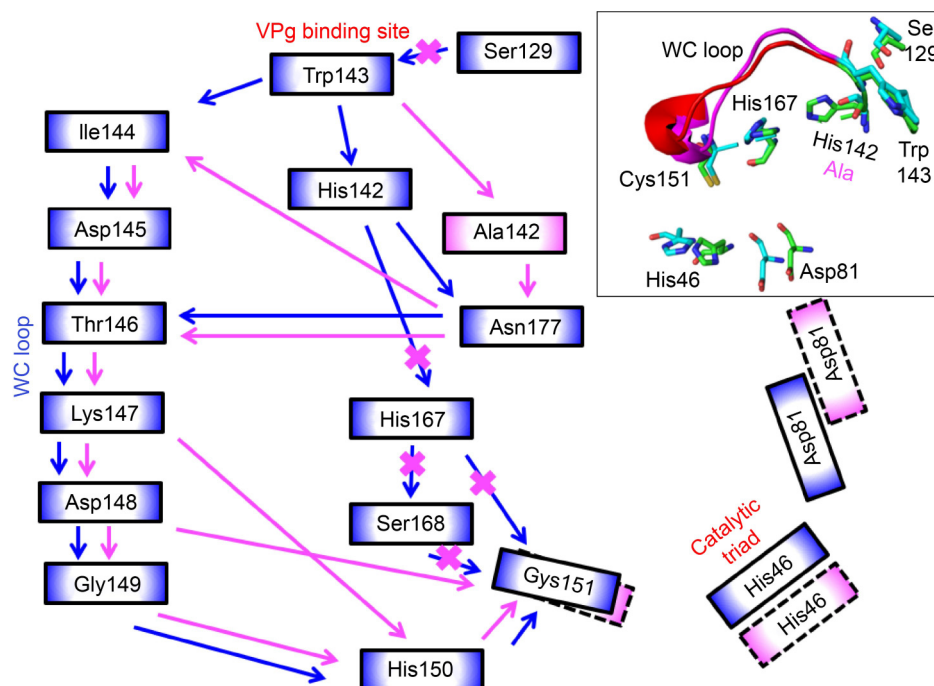


Figure 6. Schematic diagram showing the residue interaction networks in the Nla_H142 system with reference to the wild type. The inset shows the overlay of WC loop conformations (wild type: magenta, and mutant: red), orientations of the catalytic triad, and the two crucial residues, His142 and His167 (wild type: green sticks, and mutant: cyan). The residue communication pathway from the VPg interaction interface to the catalytic site for the wild type is shown using blue arrows. The modified pathway due to the His142Ala mutation is shown using pink arrows. The native communication pathway between the two residues that is no longer effective in the mutant is marked by orange crosses. The communication pathway is based on hydrogen bonding, hydrophobic, and main-chain interactions, which is also demonstrated in ligplot in the supplemental.

provided the inexpensive first step to predict such allosteric sites, trace crucial residues in the allosteric pathway, validate experimental findings, and design small molecules as drug candidates. For instance, tools like AlloPred [39] and PARS [40] use normal mode analysis to predict allosteric pockets on proteins. The dynamics upon the binding of modulators are modelled alongside support vector machine algorithms (in AlloPred) and structural conservation scores (in PARS). As the allosteric type and pathways are controlled by forces imposed upon ligand binding to the target protein, a prediction method named AlloType [41] has been recently developed. In this method, coarse grained anisotropic network model is used to define the response of the protein structures. Structural changes due to ligand binding are described using linear response theory. The free energy changes along with conformational changes determine the allosteric coupling strength. Several comprehensive allosteric data collections have been curated, including Allosteric Databases (containing protein structures and modulators) and AlloMAPS (containing allosteric mutants) [42]. These databases have aided the development of allosteric drug design platforms, including AlloSite [43,44], CavityPlus

[45], PARS, AlloSigMA for allosteric residue analysis [46,47], and AlloFinder for identification of modulators [48]. Allosteric coupling networks can be deduced from pairwise interactions at residue level, revealing densely connected elements. Molecular dynamics has been commonly used to trace such allosteric networks [49,50]. Simulation studies of large proteins can be computationally expensive. In such cases, fluctuation dynamics study using elastic network model (ENM) with normal mode analysis using single parameter harmonic potential has provided success [51–53]. The residues were considered as the nodes and linkers of the pathway and are the inter-residue potentials stabilizing the folded state. A key limitation in ENM is the presence of cutoff distance. Ohm is a platform for mapping allosteric communication and is based on the perturbation propagation algorithm. It predicts allosteric sites, inter-residue correlations, and communication pathways [54]. A combination of graph theory-based methods like bond-bond propensity analysis and Markov transient analysis have been successfully used to map the allosteric communication pathway in proteins, including the main protease of severe respiratory syndrome coronavirus 2 [55,56]. Recently,

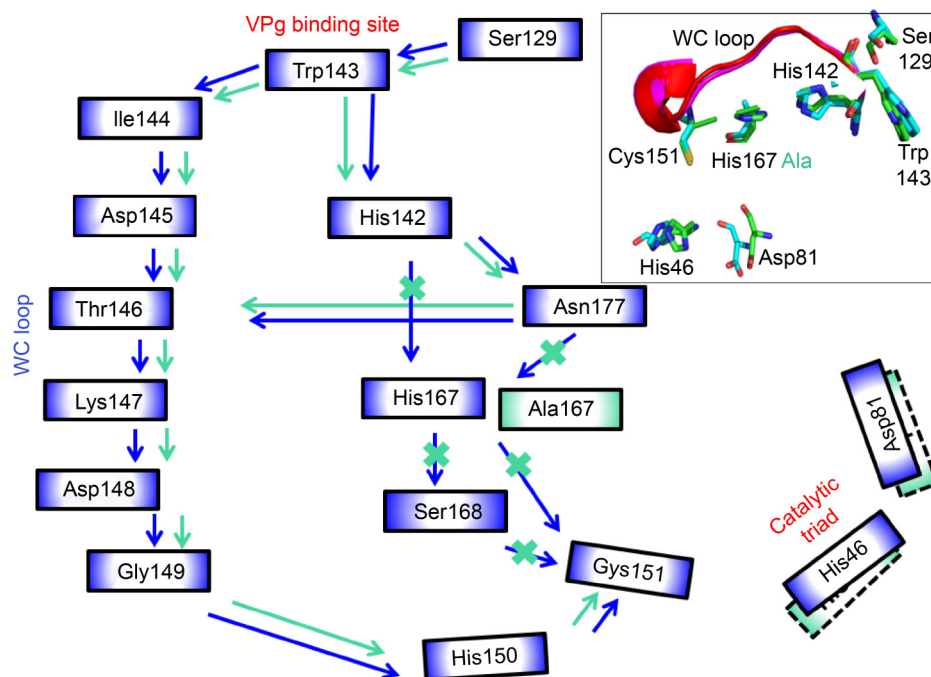


Figure 7. Schematic diagram showing the residue interaction networks in the Nla_H167 system with reference to the wild type. The inset shows the overlay of WC loop conformations (wild type: magenta, and mutant: red), orientations of the catalytic triad, and the two crucial residues, His142 and His167 (wild type: green sticks, and mutant: cyan). The residue communication pathway from the VPg interaction interface to the catalytic site for the wild type is shown using blue arrows. The modified pathway due to the His167Ala mutation is shown using green arrows. The native communication pathway between the two residues that is no longer effective in the mutant is marked by orange crosses. The communication pathway is based on hydrogen bonding, hydrophobic, and main chain interactions, which is also demonstrated in ligplot in the supplemental.

Duan *et al.* wrote a comprehensive review, which is an excellent read on the advancement of computational methods applied to understand allostery [42]. Thus, various computational strategies have been promising in our journey to understanding allostery in various proteins, including apoptosis related protein, Caspase1 [54], the regulator protein CheY [54], ClpY ATPase [57], Hsp90 [58], monoacylglycerol lipase [19], G protein coupled receptors [59], etc.

In PVBV, the binding of VPg regulates the catalytic function of Nla-Pro as evidenced by enzyme kinetics and binding studies [13]. The molecular details underlying this communication pathway was poorly understood. VPg is unstructured and its residues involved in its binding to Nla-Pro, along with the conformational changes that might occur due to this interaction, are yet to be studied. In the absence of the complex structure, we use traditional mutation studies to shed light on the communication pathway. In this study, we provide molecular insights into the allosteric regulation of Nla-Pro and the residue interaction network between the VPg interacting interface and catalytic site of Nla-Pro. Our analysis from MD studies shows significant correlation with previously documented biochemical

studies [13]. The biochemical experiments suggest that the protease activity of the Nla-Pro mutants, Ser129Ala, Trp143Ala, His142Ala, and His167Ala are impaired compared to the wild type [60]. Our simulation studies indicate that the decrease in catalytic activity of the mutants could possibly be associated with a conformational change in the catalytic triad. The point mutations Ser129Ala and Trp143Ala are located distal from the catalytic triad. However, these mutations affect the enzymatic activity as the optimal residue interaction pathway is disrupted. In the case of the Nla_S129 system, the short interaction network involves His142 and His167, along with other residues. This finding is similar to that of the Nla_W143 system; however, distortion in the side chains of Cys151 and His46 was observed. Interestingly, the residue interaction networks of the Nla_H167 and Nla_H142 systems are longer and involve the residues of the WC loop. Furthermore, the orientations of catalytic site residues are different from that in the wild type. The changes in the orientation of the residues of the catalytic triad are associated with the activity of Nla-Pro. Our findings delineate the residue level communication pathway between Trp143 present on the VPg interaction interface and the catalytic site.

The residues operate in a concerted manner and aid the allosteric regulation the NIa-Pro. In case a VPg-NIa-Pro complex can be obtained in the future using any experimental technique, simulation studies on the complex would reveal more details and add to the current study.

MATERIALS AND METHODS

Starting structures

The sequences of NIa-Pro (accession number NP_188424, At3g17970) was used for the study. In order to obtain a reliable model for *in silico* studies, the protein sequence was submitted for automated protein structure modelling using the I-TASSER pipeline. This server builds 3D models based on multiple threading alignments and iterative structure assembly simulations [61–64]. Since the above method used multiple threading alignments, there is minimal bias towards a particular structural model as in the case of homology modelling [65,66]. The model systems for the mutants were built using the coordinates of the wild type obtained from the I-TASSER server as templates. The wild type and mutant starting structures were further validated using Ramachandran plots [67].

Six systems were prepared: NIa wild type (NIa_wt), NIa-Pro Ser129Ala mutant (NIa_S129), NIa-Pro His142Ala mutant (NIa_H142), NIa-Pro Trp143Ala mutant (NIa_W143), and NIa-Pro His167Ala mutant (NIa_H167).

Preparation

The simulation studies were performed using assisted model building with energy refinement (AMBER) suite version 12 [68] associated with the all-atom ff12SB force field [69,70]. The starting structures for the simulations were neutralized using Na⁺ and Cl[−] ions. The hydrogen atoms were positioned using the *tleap* module from AMBERTOOLS12. The protein was centered in a solvent truncated octahedron box, which was comprised of TIP3P (3-point charged) triangulated water molecules with a 12 Å cut off in all directions [71]. The total number of atoms including water molecules was approximately 20,000 across each of the various systems. The systems were minimized using a two-phase energy minimization procedure, which included 2,500 cycles of steepest descent and 2,500 cycles of conjugate gradient with solute atoms restrained by a harmonic potential with a constant force of 50 kcal mol^{−1} Å². This was followed by 5,000 steps of unrestrained whole system minimization. 50 ps of density equilibration with weak harmonic restraints

(2 kcal mol^{−1} Å²) on the solute molecule was performed followed by unrestrained equilibration for 500 ps using under constant pressure and temperature conditions. All simulations ran with constraints using the SHAKE algorithm [72] on hydrogen-linked bonds with a tolerance of 0.0001. To evaluate long-range electrostatic interactions, the Particle Mesh Ewald (PME) method [73] was used with a 9 Å cut-off. A 2 fs integration time step was used to numerically solve Newton's equations of motion. Langevin dynamics was used to maintain a constant temperature of 300 K throughout the simulations. All the simulations were carried out using the PMEMD module in AMBER12. Production runs were performed for 50 ns in an explicit solvent environment and isothermal-isobaric (NPT) ensemble.

Analysis

After 40 ns of production run, each of the trajectories was analyzed based on the variation in kinetic and potential energies under the NPT ensemble using the *ptraj* program. The root-mean-square deviation (RMSD) and atomic positional fluctuation per residue (RMSF) were analyzed to understand the overall conformational change throughout the trajectory. Hydrogen bonding, hydrophobic interactions, and their occupancies were calculated using HBPLUS using the criteria described by McDonald and Thornton for the definition of a hydrogen bond [74]. Only the hydrogen bonds with occupancies above 50% were considered for analysis. LIGPLOT was used to map the hydrogen and hydrophobic bonding patterns between the WC loop residues, mutated residues, and other residues of the NIa protease in various systems [75]. The last 5 ns averaged structures from each simulation were used for further analysis. Visual molecular dynamics (VMD) was used for visualizing the trajectories of the simulations of the six systems [76]. Furthermore, performing principal component analysis (PCA) substantiated the above results [77]. Collective coordinates for protein motions were extracted using Covariance matrices. These matrices were constructed using the backbone atoms of the protein (N, Ca, C) and were used to calculate the eigenvalues of maximum magnitude using GROMACS [78]. PyMOL was used to generate all the graphical representations [79].

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.15302/J-QB-022-0296>.

ACKNOWLEDGMENTS

We thank Prof. H. S. Savithri from Indian Institute of Sciences for

providing intellectual inputs on the project and funding Dr. Panigrahi for two months while she designed and performed this computational project using her personal resources.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Rashmi Panigrahi and Senthilkumar Kailasam declare that they have no conflict of interests.

All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Martínez, F., Rodrigo, G., Aragonés, V., Ruiz, M., Lodewijk, I., Fernández, U., Elena, S. F. and Daròs, J. A. (2016) Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. *BMC Genomics*, 17, 87
- Anindya, R., Joseph, J., Gowri, T. D. and Savithri, H. S. (2004) Complete genomic sequence of pepper vein banding virus (PVBV): a distinct member of the genus *Potyvirus*. *Arch. Virol.*, 149, 625–632
- Joseph, J. and Savithri, H. S. (2000) Mutational analysis of the NIa protease from pepper vein banding potyvirus. *Arch. Virol.*, 145, 2493–2502
- Anindya, R., Chittori, S. and Savithri, H. S. (2005) Tyrosine 66 of pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. *Virology*, 336, 154–162
- Urcuqui-Inchima, S., Haenni, A. L. and Bernardi, F. (2001) Potyvirus proteins: a wealth of functions. *Virus Res.*, 74, 157–175
- Ghabrial, S. A., Smith, H. A., Parks, T. D. and Dougherty, W. G. (1990) Molecular genetic analyses of the soybean mosaic virus NIa proteinase. *J. Gen. Virol.*, 71, 1921–1927
- Dougherty, W. G. and Parks, T. D. (1989) Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites in cell-free assays. *Virology*, 172, 145–155
- Merits, A., Rajamäki, M. L., Lindholm, P., Runeberg-Roos, P., Kekarainen, T., Puustinen, P., Mäkeläinen, K., Valkonen, J. P. T. and Saarma, M. (2002) Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *J. Gen. Virol.*, 83, 1211–1221
- Pasin, F., Simón-Mateo, C. and García, J. A. (2014) The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses. *PLoS Pathog.*, 10, e1003985
- Riechmann, J. L., Laín, S. and García, J. A. (1992) Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol.*, 73, 1–16
- Schaad, M. C., Haldeman-Cahill, R., Cronin, S. and Carrington, J. C. (1996) Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J. Virol.*, 70, 7039–7048
- Carrington, J. C., Haldeman, R., Dolja, V. V. and Restrepo-Hartwig, M. A. (1993) Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus *in vivo*. *J. Virol.*, 67, 6995–7000
- Sabharwal, P., Srinivas, S. and Savithri, H. S. (2018) Mapping the domain of interaction of PVBV VPg with NIa-Pro: Role of N-terminal disordered region of VPg in the modulation of structure and function. *Virology*, 524, 18–31
- Mathur, C. and Savithri, H. S. (2012) Novel ATPase activity of the polyprotein intermediate, viral protein genome-linked-nuclear inclusion-a protease, of pepper vein banding potyvirus. *Biochem. Biophys. Res. Commun.*, 427, 113–118
- Mathur, C., Jimsheena, V. K., Banerjee, S., Makinen, K., Gowda, L. R. and Savithri, H. S. (2012) Functional regulation of PVBV nuclear inclusion protein—a protease activity upon interaction with viral protein genome-linked and phosphorylation. *Virology*, 422, 254–264
- Boerner, J. E., Lyle, J. M., Daijogo, S., Semler, B. L., Schultz, S. C., Kirkegaard, K. and Richards, O. C. (2005) Allosteric effects of ligands and mutations on poliovirus RNA-dependent RNA polymerase. *J. Virol.*, 79, 7803–7811
- Changeux, J. P. (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu. Rev. Biophys.*, 41, 103–133
- del Sol, A., Fujihashi, H., Amoros, D. and Nussinov, R. (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.*, 2, 0019
- Tyukhtenko, S., Ma, X., Rajarshi, G., Karageorgos, I., Anderson, K. W., Hudgens, J. W., Guo, J. J., Nasr, M. L., Zvonok, N., Vemuri, K., *et al.* (2020) Conformational gating, dynamics and allostery in human monoacylglycerol lipase. *Sci. Rep.*, 10, 18531
- Cooper, A. and Dryden, D. T. (1984) Allostery without conformational change. A plausible model. *Eur. Biophys. J.*, 11, 103–109
- Popovych, N., Sun, S., Ebright, R. H. and Kalodimos, C. G. (2006) Dynamically driven protein allostery. *Nat. Struct. Mol.*

- Biol., 13, 831–838
22. Blacklock, K. and Verkhivker, G. M. (2014) Computational modeling of allosteric regulation in the hsp90 chaperones: a statistical ensemble analysis of protein structure networks and allosteric communications. *PLOS Comput. Biol.*, 10, e1003679
 23. Ghosh, A. and Vishveshwara, S. (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA.*, 104, 15711–15716
 24. Atilgan, A. R., Turgut, D. and Atilgan, C. (2007) Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. *Biophys. J.*, 92, 3052–3062
 25. Ghosh, A., Sakaguchi, R., Liu, C., Vishveshwara, S. and Hou, Y. M. (2011) Allosteric communication in cysteinyl tRNA synthetase: a network of direct and indirect readout. *J. Biol. Chem.*, 286, 37721–37731
 26. Taverna, D. M. and Goldstein, R. A. (2002) Why are proteins so robust to site mutations? *J. Mol. Biol.*, 315, 479–484
 27. Sun, P., Austin, B. P., Tözsér, J., and Waugh, D. S. (2010) Structural determinants of tobacco vein mottling virus protease substrate specificity. *Protein Sci.*, 19, 2240–2251
 28. Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. K. 3rd, Kapust, R. B., Li, M., Wlodawer, A. and Waugh, D. S. (2002) Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.*, 277, 50564–50572
 29. Wenthur, C. J., Gentry, P. R., Mathews, T. P. and Lindsley, C. W. (2014) Drugs for allosteric sites on receptors. *Annu. Rev. Pharmacol. Toxicol.*, 54, 165–184
 30. Sheik Amamuddy, O., Veldman, W., Manyumwa, C., Khairallah, A., Agajanian, S., Oluyemi, O., Verkhivker, G. and Tastan Bishop, O. (2020) Integrated computational approaches and tools for allosteric drug discovery. *Int. J. Mol. Sci.*, 21, 847
 31. Sinha, N. and Nussinov, R. (2001) Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc. Natl. Acad. Sci. USA.*, 98, 3139–3144
 32. Vaish, N. K., Dong, F., Andrews, L., Schweppe, R. E., Ahn, N. G., Blatt, L. and Seiwert, S. D. (2002) Monitoring post-translational modification of proteins with allosteric ribozymes. *Nat. Biotechnol.*, 20, 810–815
 33. Brzović, P. S., Ngo, K. and Dunn, M. F. (1992) Allosteric interactions coordinate catalytic activity between successive metabolic enzymes in the tryptophan synthase bienzyme complex. *Biochemistry*, 31, 3831–3839
 34. Hilser, V. J. and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA*, 104, 8311–8315
 35. Schlessinger, J. (1988) Signal transduction by allosteric receptor oligomerization. *Trends Biochem. Sci.*, 13, 443–447
 36. Caldo, K. M. P., Acedo, J. Z., Panigrahi, R., Vederas, J. C., Weselake, R. J., and Lemieux, M. J. (2017) Diacylglycerol acyltransferase 1 is regulated by its N-terminal domain in response to allosteric effectors. *Plant Physiol.*, 175, 667–680
 37. Panigrahi, R., Matsui, T., Song, A. H., Caldo, K. M. P., Young, H. S., Weselake, R. J., and Lemieux, M. J. (2018) Intrinsic disorder in the regulatory N-terminal domain of diacylglycerol acyltransferase 1 from *Brassica napus*. *Sci. Rep.*, 8, 1–13
 38. Buller, A. R., Van Roye, P., Cahn, J. K., Scheele, R. A., Herger, M., and Arnold, F. H. (2018) Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.*, 140, 7256–7266
 39. Greener, J. G., and Sternberg, M. J. (2015) AlloPred: Prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics*, 16, 335
 40. Panjkovich, A., and Daura, X. J. B. (2014) PARS: A web server for the prediction of protein allosteric and regulatory sites. *Bioinformatics*, 30, 1314–1315
 41. Huang, Q., Song, P., Chen, Y., Liu, Z., and Lai, L. (2021) Allosteric type and pathways are governed by the forces of protein-ligand binding. *J. Phys. Chem. Lett.*, 12, 5404–5412
 42. Ni, D., Chai, Z., Wang, Y., Li, M., Yu, Z., Liu, Y., Lu, S., and Zhang, J. (2021) Along the allostery stream: Recent advances in computational methods for allosteric drug discovery. *Wires Comput. Mol. Sci.*, e1585
 43. Huang, W., Lu, S., Huang, Z., Liu, X., Mou, L., Luo, Y., Zhao, Y., Liu, Y., Chen, Z., Hou, T., *et al.* (2013) Allosite: a method for predicting allosteric sites. *Bioinformatics*, 29, 2357–2359
 44. Song, K., Liu, X., Huang, W., Lu, S., Shen, Q., Zhang, L. and Zhang, J. (2017) Improved method for the identification and validation of allosteric sites. *J. Chem. Inf. Model.*, 57, 2358–2363
 45. Xu, Y., Wang, S., Hu, Q., Gao, S., Ma, X., Zhang, W., Shen, Y., Chen, F., Lai, L. and Pei, J. (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.*, 46, W374–W379
 46. Tan, Z. W., Guarnera, E., Tee, W.-V., and Berezovsky, I. N. (2020) Allosigma 2: Paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. *Nucleic Acids Res.*, 48, W116–W124
 47. Guarnera, E., Tan, Z. W., Zheng, Z., and Berezovsky, I. N. (2017) AlloSigMA: Allosteric signaling and mutation analysis server. *Bioinformatics*, 33, 3996–3998
 48. Huang, M., Song, K., Liu, X., Lu, S., Shen, Q., Wang, R., Gao, J., Hong, Y., Li, Q., Ni, D., *et al.* (2018) Allofinder: A strategy for allosteric modulator discovery and allosterome analyses. *Nucleic Acids Res.*, 46, W451–W458
 49. Shukla, D., Meng, Y., Roux, B. and Pande, V. S. (2014) Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.*, 5, 3397
 50. Ghosh, A. and Vishveshwara, S. (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA.*, 104, 15711–15716
 51. Yang, L., Song, G., and Jernigan, R. L. (2009) Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA*, 106, 12347–12352
 52. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. and Bahar, I. (2001) Anisotropy of fluctuation

- dynamics of proteins with an elastic network model. *Biophys. J.*, 80, 505–515
53. Stolzenberg, S., Michino, M., LeVine, M. V., Weinstein, H. and Shi, L. (2016) Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochim. Biophys. Acta*, 1858, 1652–1662
 54. Wang, J., Jain, A., McDonald, L. R., Gambogi, C., Lee, A. L. and Dokholyan, N. V. (2020) Mapping allosteric communications within individual proteins. *Nat. Commun.*, 11, 3862
 55. Strömich, L., Wu, N., Barahona, M., and Yaliraki, S. N. (2022) Allosteric hotspots in the main protease of SARS-CoV-2. *J. Mol. Biol.*, 434, 167748
 56. Wu, N., Strömich, L. and Yaliraki, S. N. (2022) Prediction of allosteric sites and signaling: Insights from benchmarking datasets. *Patterns*, 3, 100408
 57. Kravats, A. N., Tondast-Navaei, S. and Stan, G. (2016) Coarse-grained simulations of topology-dependent mechanisms of protein unfolding and translocation mediated by ClpY ATPase nanomachines. *PLOS Comput. Biol.*, 12, e1004675
 58. Chandramohan, A., Krishnamurthy, S., Larsson, A., Nordlund, P., Jansson, A. and Anand, G. S. (2016) Predicting allosteric effects from orthosteric binding in hsp90-ligand interactions: Implications for fragment-based drug design. *PLOS Comput. Biol.*, 12, e1004840
 59. Wooten, D., Reynolds, C. A., Smith, K. J., Mobarec, J. C., Furness, S. G., Miller, L. J., Christopoulos, A. and Sexton, P. M. (2016) Key interactions by conserved polar amino acids located at the transmembrane helical boundaries in Class B GPCRs modulate activation, effector specificity and biased signalling in the glucagon-like peptide-1 receptor. *Biochem. Pharmacol.*, 118, 68–87
 60. Sabharwal, P. (2017) Molecular insights into the structure and function of pepper vein banding virus encoded proteins and endocytic uptake pathway of virus-like particles into mammalian cells. Doctorate of Philosophy, Bangalore: Indian Institute of Science
 61. Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2011) A protocol for computer-based protein structure and function prediction. *J. Vis. Exp.*, 57, e3259
 62. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40
 63. Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5, 725–738
 64. Roy, A., Yang, J. and Zhang, Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.*, 40, W471–477
 65. Chung, S. Y., and Subbiah, S. (1996) A structural explanation for the twilight zone of protein sequence homology. *Structure*, 4, 1123–1127
 66. Venclovas, C. and Margelevicius, M. (2005) Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. *Proteins*, 61, 99–105
 67. Lovell, S. C., Davis, I. W., Arendall, W. B. 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*, 50, 437–450
 68. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S. and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with amber on GPUS. 1. Generalized born. *J. Chem. Theory Comput.*, 8, 1542–1555
 69. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., *et al.* (2012) Amber 12. San Francisco: University of California
 70. Case, D. A., Cheatham, T. E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, 26, 1668–1688
 71. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79, 926–935
 72. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, 23, 327–341
 73. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with amber on GPUS. 2. Explicit solvent particle Mesh Ewald. *J. Chem. Theory Comput.*, 9, 3878–3888
 74. McDonald, I. K. and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, 238, 777–793
 75. Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8, 127–134
 76. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, 14, 33–38., 27–28
 77. Amadei, A., Linssen, A. B. and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins*, 17, 412–425
 78. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845–854
 79. DeLano, W. L. (2002) The pymol molecular graphics system. DeLano Scientific, San Carlos, CA, USA