



OPEN

# *Dynamozones* are the most obvious sign of the evolution of conformational dynamics in HIV-1 protease

Mohammad Rahimi, Majid Taghdir  & Farzane Abasi Joozdani


Proteins are not static but are flexible molecules that can adopt many different conformations. The HIV-1 protease is an important target for the development of therapies to treat AIDS, due to its critical role in the viral life cycle. We investigated several dynamics studies on the HIV-1 protease families to illustrate the significance of examining the dynamic behaviors and molecular motions for an entire understanding of their dynamics-structure-function relationships. Using computer simulations and principal component analysis approaches, the dynamics data obtained revealed that: (i) The flap regions are the most obvious sign of the evolution of conformational dynamics in HIV-1 protease; (ii) There are dynamic structural regions in some proteins that contribute to the biological function and allostery of proteins via appropriate flexibility. These regions are a clear sign of the evolution of conformational dynamics of proteins, which we call dynamozones. The flap regions are one of the most important dynamozones members that are critical for HIV-1 protease function. Due to the existence of other members of dynamozones in different proteins, we propose to consider dynamozones as a footprint of the evolution of the conformational dynamics of proteins.

Proteins are biomolecules that are regarded as the machinery of life. They are intrinsically dynamic, and their conformational variability is essential to their biological functions<sup>1,2</sup>. The function of a protein besides the structure also relies on its dynamics. Protein flexibility is necessary for biological function, ligand binding, and protein-protein or protein-nucleic acid interactions. A quantitative description of protein dynamics is essential for understanding living systems at a molecular level and probably also for the mechanisms leading to protein malfunction<sup>3,4</sup>.

Protein flexibility refers to the protein structure's internal dynamics, which is beneficial in the structural and functional aspects of proteins<sup>5</sup>. Plasticity and Conformational mobility represent key intrinsic features of proteins through evolution. Internal mobility eases the evolution of proteins to adopt conformational flexibility and thus provides the opportunity to develop novel functions. In addition, conformational flexibility allows proteins to better cope with harmful mutations which can lead to loss of function or altered function that result in disease<sup>5,6</sup>. Despite substantial proof that suggests that protein dynamics are under evolutionary selection, little is known about the molecular basis of the evolution of protein dynamics or how they affect function. An interesting case in the context of the dynamic-function relationship is that structural dynamics play a significant role in protein promiscuity, which almost means the ability of proteins to carry out several more or less related molecular works.

Indeed, every protein has the potential to accept many various conformations in the native state, so, many proteins are capable to perform several functions<sup>7,8</sup>. The structural diversity linked to protein flexibility constitutes a basis of protein evolvability<sup>6</sup>. In some cases, the protein exhibits one well-defined primary function together with several low-level "promiscuous activities." Moonlighting proteins are other cases that can effectively carry out several relevant functions or even different duties linked to various molecular surfaces or active sites. Thus, it appears rational to attend that a particular trait of dynamics in functionally related protein regions (e.g., conformational variety, active-site "flexibility", or "deformability", results in the ability to stabilize different substrates, transition states of leaving groups) may be linked to many cases of protein promiscuity<sup>9,10</sup>.

Protein conformational dynamics play an important role in evolution, normal physiology, and pathophysiology. The evolution of proteins involves mutations that may lead to proteins adopting new functions and, in rare cases, new folds. Indeed, mutations of proteins can alter their conformations, dynamics, and stability, and thereby

Department of Biophysics, Faculty of Biological Science, Tarbiat Modares University, Tehran 14115\_111, Iran.  
email: Taghdir@modares.ac.ir

play critical roles in evolution and diseases. At a molecular level, protein evolution is dominated by neutral or nearly neutral mutations that have little effect on function<sup>11</sup>. Nonetheless, our comprehension of how proteins and species evolve is still elementary. There is a lack of detailed understanding of how proteins have evolved<sup>5,6</sup>.

One of the important questions about the evolution of proteins that should be further investigated is: How do conformational dynamics evolve as proteins evolve? The results of some studies show that the dynamics and the evolution of proteins share similarities<sup>12,13</sup>. Tang et al.<sup>14</sup> found that there is a correspondence between the dynamics and evolution of protein structures. Their results show that the evolutionary mechanism of the proteins obtains both dynamical flexibility and evolutionary structural variation. Studying the evolution of protein conformational dynamics would not be possible without the use of computational and biophysical methodologies, that allow structural dynamics to be dissected in different protein variants. Much work has been done on the evolution of protein structure, but the role of protein dynamics in evolution has received attention in recent years.

Molecular dynamics (MD) simulations are a powerful tool to investigate the dynamic behavior of proteins in an aqueous solution and deepen our understanding of the relationship between protein structure and function<sup>3</sup>. The global molecular motions of the proteins can be obtained by applying the combination of MD simulation and essential dynamics (ED) analysis technique<sup>15</sup>.

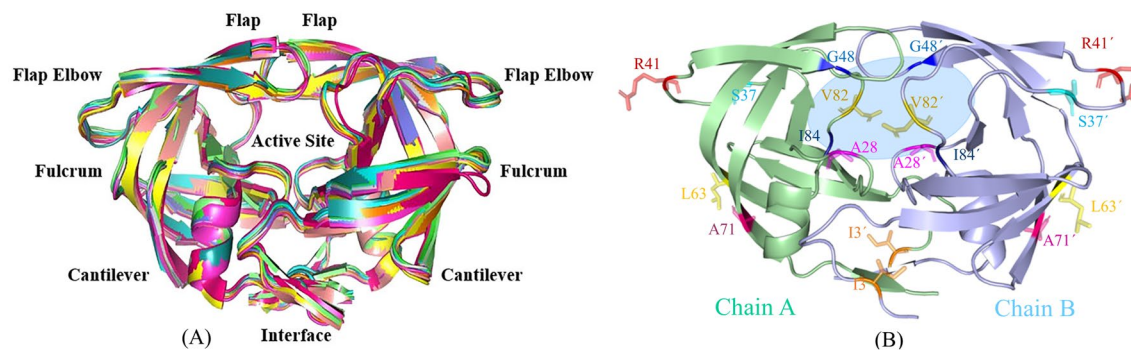
HIV-1 protease (human immunodeficiency virus type 1 protease), an aspartyl protease, is responsible for the generation of structural proteins and viral enzymes critical to HIV viral maturation and infectivity. Thus, HIV-1 protease is a major drug target in the battle against HIV-1 infection, where the inactivation of the HIV-1 protease causes the production of immature, noninfectious viral particles<sup>16,17</sup>. There are about 750 experimentally determined available structures of this enzyme and this wide structural knowledge allows a study of a large number of conformations of protease complexes. The HIV-1 protease is one of the best-characterized cases of protein molecular evolution<sup>18</sup>. The HIV-1 protease is a homodimer with two identical monomers (chain A and chain B) each consisting of 99 amino acids. Each monomer has one  $\alpha$ -helix (usually residues from 86 to 90) and nine  $\beta$ -sheets in the secondary structure. The residues of HIV-1 protease are numbered 1–99 and 100–198 (or 1'–99') for chains A and B, respectively.

HIV-1 protease is a consisting of six structural segments (Fig. 1A): interface (residues 1–5/100–104, 95–99/194–198), fulcrum (residues 11–22/110–121), active site (residues 23–30/122–129), flap (residues 43–58/142–157), flap elbow (residues 35–42/134–141), and cantilever (residues 59–75/158–174).

The active site of the enzyme is formed at the dimer interface with each monomer a conserved catalytic triad (Asp25-Thr26-Gly27). It's gated by two extended  $\beta$ -hairpin loops known as flaps. Two Asp25 residues (one from each monomer or chain) act as the catalytic residues and the conserved active site residues forms a symmetrical and highly hydrogen-bonded arrangement<sup>19,20</sup>.

Investigating protein flexibility may be important for the study of processes associated with conformational changes and state transitions<sup>21</sup>. Structural and dynamical studies of the HIV-1 protease normally focused on its more flexible region, the flaps, since they control the entrance/stabilization of ligands in the active site<sup>22,23</sup>. There is a large variety in the flap conformations in the unbound state, fluctuating between the closed, semi-open, and wide-open conformations<sup>24,25</sup>. In the closed/semi-open state, the catalytic site is shielded with two flaps and thus limits the entry of most of the ligands. The semi-open conformation is the dominant state in the ligand-free HIV-1 protease<sup>24–26</sup>. The flexibility of the flap is needed to facilitate the substrate access to and product release from the active site of the enzyme by an open and close mechanism<sup>27</sup>. The binding of protease substrate to the active site can be controlled by limiting the movement of the flap, thereby inhibiting HIV-1 protease activity<sup>24,28</sup>.

In order to indicate the relationship between the dynamics of proteins and their structure properties, we examine several dynamics studies. This study includes the investigations of the molecular motions and dynamic behaviors of the HIV-1 protease family in relation to their structure using computer simulation techniques. We adopted here the unbound form of the proteases of HIV-1 to investigate their dynamics-structure-function relationships. We focus mainly on the following facets: (i) dynamic behavior and collective motions of the HIV-1 protease family; (ii) the effect of point mutations on the molecular motions and stability of the HIV-1 protease family; (iii) correlation of some dynamic structural regions with the evolution of conformational dynamics in the HIV-1 proteases family.



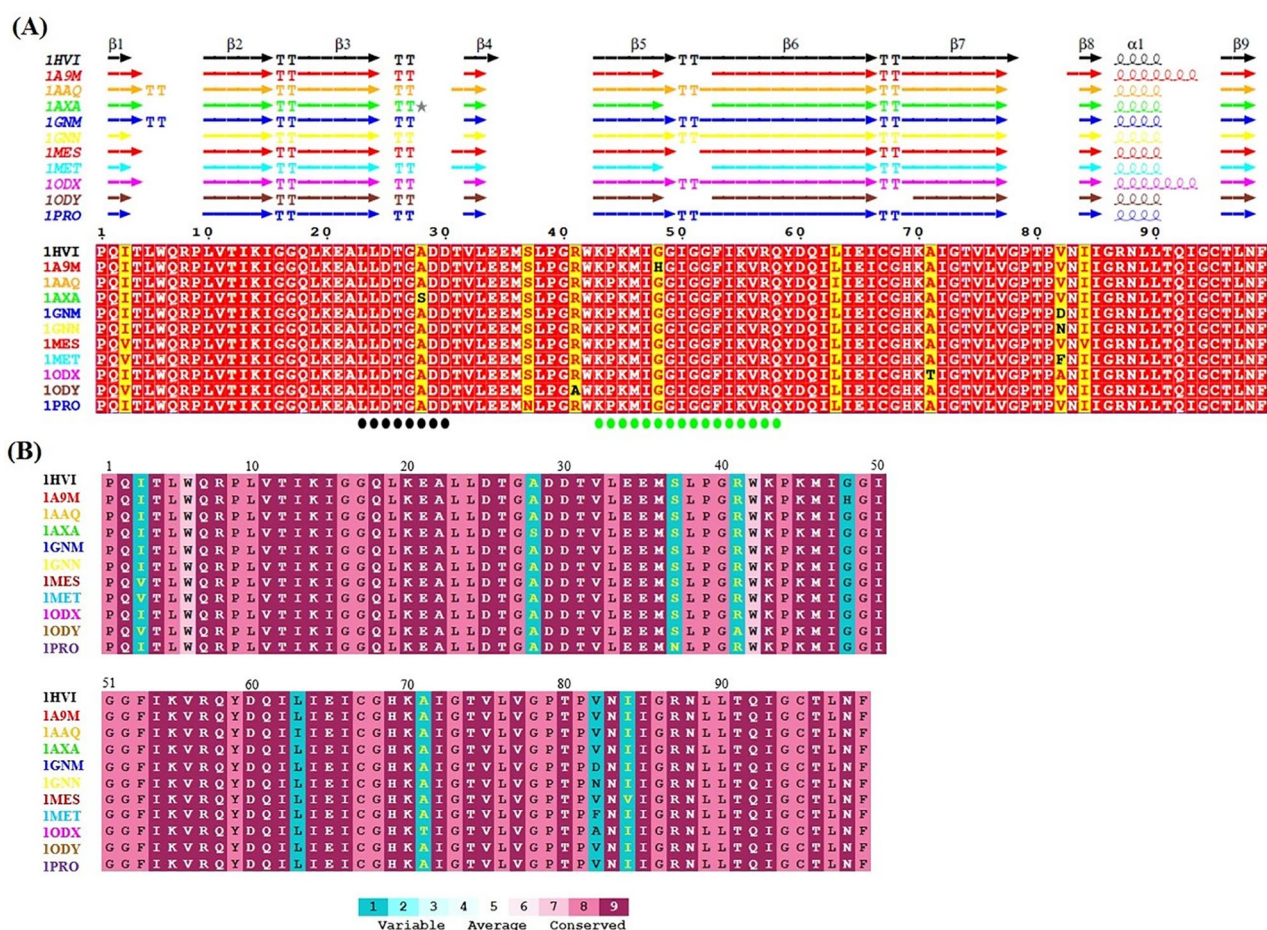
**Figure 1.** (A) Superposition of the crystal structures of the eleven selected HIV-1 proteases. Important regions of the HIV-1 proteases are labeled. (B) 3D structure of HIV-1 Protease (PDB code: 1HVI) showing the positions of mutated residues and also ligand position (light blue ellipse). Rendered using PyMOL.

The cross-correlation analysis and principal component analysis (PCA) were also performed to probe the difference in internal dynamics and conformational changes of the selected proteins induced by mutation. It is evident from the correlation map that almost in all proteases the flaps and flap elbow motions are highly correlated.

## Results and discussion

**Sequence analysis.** Proteins with sequence identity > 30% typically belong to the same family and have similar conformation and function. Such clear homologues are probably to have separated from a common ancestor and their sequences may show conserved differences between species of organisms. The simultaneous comparison of sequence and structure information is of significance to detect biological specificities in a group of proteins<sup>29,30</sup>. Multiple sequence alignment was performed using T-COFFEE and rendered by ESPrpt 3.0 using default parameters for residue similarity calculations, where boxed residues represent identical (red box, white character) and similar (Yellow box, red character) amino acid conservation (Fig. 2A). Residues boxed in red indicate strict conservation, while residues boxed in yellow indicate greater than or equal to 80% identity across the 11 homologs. Interestingly, the  $\alpha$ -helix structure in proteases of the 1A9M and 1ODX is somewhat longer.

The structural and functional significance of a residue in protein structure is substantial for its evolutionary conservation. The importance of a given residue in conserving the structure and function of a protein can be inferred from the degree of conservation of the residue in a multiple sequence alignment of the protein and its homologues. The ConSurf<sup>31</sup> is a bioinformatics tool for calculating the conservation pattern of a protein, which quantifies the degree of conservation at each aligned position. This program first identifies conserved positions using multiple sequence alignment, then calculates the evolutionary conservation rate using the empirical Bayesian method and provides the evolutionary conservation profiles of the structure or the sequence of the protein.



**Figure 2.** (A) Sequence alignment produced by T-COFFEE and the figure was prepared with ESPrpt 3.0. Identical residues are written in white bold characters and boxed in red whereas similar residues are in red bold characters and boxed in yellow. Secondary structure elements related to each protein are indicated at the top of sequence blocks ( $\alpha$ ,  $\alpha$ -helix;  $\beta$ ,  $\beta$ -strand; TT, turn). Residues forming the active site (23–30) and the flap region (43–58) are labelled with a black circle and a green circle under the sequence alignment, respectively. (B) The Evolutionary Conservation prediction analysis of amino acid residues of all the selected proteins by ConSurf server. The amino acids are colored according to their conservation grades and conservation levels. A grade of 1 indicates rapidly evolving sites, which are color-coded in turquoise; 5 shows sites that are evolving at an average rate, which are colored white; and 9 shows evolutionarily conserved sites, which are color-coded in maroon.

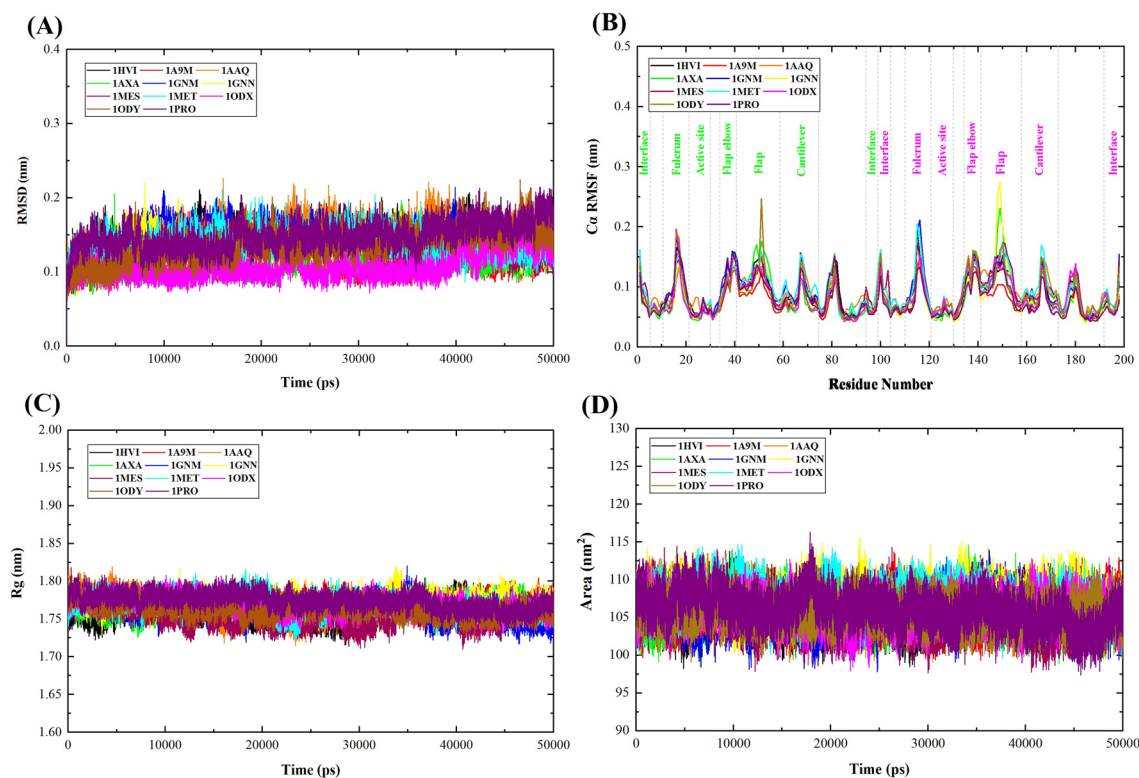
ConSurf identifies functional regions in proteins, taking into account by considering the evolutionary relationships among their sequence homologs. ConSurf score ranges from 1 to 9, with 1 denoting rapidly evolving (variable) sites, 5 depicting sites that are evolving at an average rate, and 9 representing slowly evolving (evolutionary conserved) sites. The degree of conservation of the amino acid sites among the eleven homologues with similar sequences was estimated (Fig. 2B). Importantly, the information from the sequence logo of the proteins indicates that sequences are highly conserved in different proteins. As expected, the ConSurf analysis has revealed, that most of the amino acids in all the selected proteins are highly conserved. Similarly, this analysis indicated that the functional regions of all proteins are highly conserved.

**Root mean square deviation (RMSD).** To obtain information about the conformational stability and assess the reliability of MD simulation, the RMSD of the backbone atoms of all the selected proteins was calculated. The value of RMSD has a negative correlation with the stability of the backbone atoms. The larger the value of RMSD, the more unstable the backbone atoms are<sup>32</sup>. Figure 3A shows the plot of RMSD for 1HVI (native protein) and ten proteins mutated. Initially, in the first 5000 ps, the RMSD was raised due to the "relaxation" of the proteins in the water environment, which is commonly observed in all MD simulation types. It is observed that all the proteins reach equilibrium after ~5 ns and present a steady behavior throughout the triplicates and thus suitable for exploring the dynamics of selected proteins.

The RMSD average values for 1HVI, 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY and 1PRO were found to be 0.146, 0.121, 0.146, 0.115, 0.144, 0.133, 0.133, 0.138, 0.107, 0.135 and 0.148 nm, respectively. 1PRO showed a higher RMSD value as compared to other proteins, whereas 1ODX showed the least value, which confirmed its greater stability than other proteins. The RMSD of wild and mutated proteins did not fluctuate convincingly, and all RMSD average values were less than 0.2 nm, thus, the equilibrium of all MD simulations is reliable. RMSD plots of these proteins showed that they displayed relatively similar stability. The replicates of each protein had slight variations in the RMSD values. Minor deviations in RMSD suggest the stable conformation of all proteins during this time period.

The plateau of RMSD values, observed at all simulations, is approximately similar between the wild-type and mutant proteins, indicating that all structures fluctuate around a stable average conformation. So, it is reasonable to evaluate its local fluctuations<sup>33</sup>.

**Root mean square fluctuation (RMSF).** RMSF was calculated to study better the effect of amino acid mutations on the conformational flexibility of WT and mutant HIV-1 protease variants. The high value of RMSF



**Figure 3.** Plot illustrating (A) RMSD, (B) RMSF, (C) Rg, and (D) SASA for WT, and mutant proteins. The symbol coding scheme is as follows: wild-type protein (black color), mutant proteins 1A9M (red color), 1AAQ (orange color), 1AXA (green color), 1GNM (blue color), 1GNN (yellow color), 1MES (maroon color), 1MET (cyan color), 1ODX (magenta color), 1ODY (brown color), and 1PRO (violet color).

shows the flexible region, while the low value of RMSF denotes limited movements during MD simulation. A fluctuation value of less than 2 Å is acceptable for a small protein<sup>34</sup>.

The comparison of the fluctuations between WT and mutant structures evidenced that the presence of the mutation resulted in no significant local flexibility alterations (Fig. 3B). The variation in the RMSF values suggested that the fluctuating behaviors were almost similar in the wild-type and mutant proteins except at the flap and fulcrum regions. The RMSF average values for mutant proteins were 0.0762, 0.0904, 0.0817, 0.0887, 0.0804, 0.0886, 0.0919, 0.0816, 0.0848, 0.0797 nm, for 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY, and 1PRO, respectively, while the RMSF value for WT is 0.0864 nm (Table 1). According to the fluctuation score, we ranked the collected values as follows: 1MET > 1AAQ > 1GNM > 1MES > 1HVI (WT) > 1ODY > 1AXA > 1ODX > 1GNN > 1PRO > 1A9M. Therefore, 1MET showed larger fluctuations as compared to other proteins, whereas 1A9M showed the least fluctuations. In proteins 1A9M and 1ODX, the number of helix pitches in  $\alpha$ -helix has increased, as a result, these two proteins have become more stable and show fewer fluctuations.

The RMSF plot showed that residual fluctuations are present in all proteins in several regions of the structure of the proteins. In addition to the N- and C-terminal residues, the regions around 17(116), 41(140), 52 (151), 67(166), and 81(180) show the biggest dynamic fluctuations. Residues 1–37 and 59–99 in each monomer are defined as the core region, while residues 43–58 constitute the flap region. It is worth noting that for all the proteins, changes observed in one monomer are almost always present also in the other. In the WT and mutant HIV-1 protease variants, there are two very stable regions in both monomers, one in the active site (residues 23–30/122–129) and the other in the  $\alpha$ -helix formed by residues 86 to 90. As mentioned, the residues near the catalytic D25/D124 present a high degree of rigidity in all proteins, which is expected, as the catalytic function of these residues probably needs a well-defined stable three-dimensional structure.

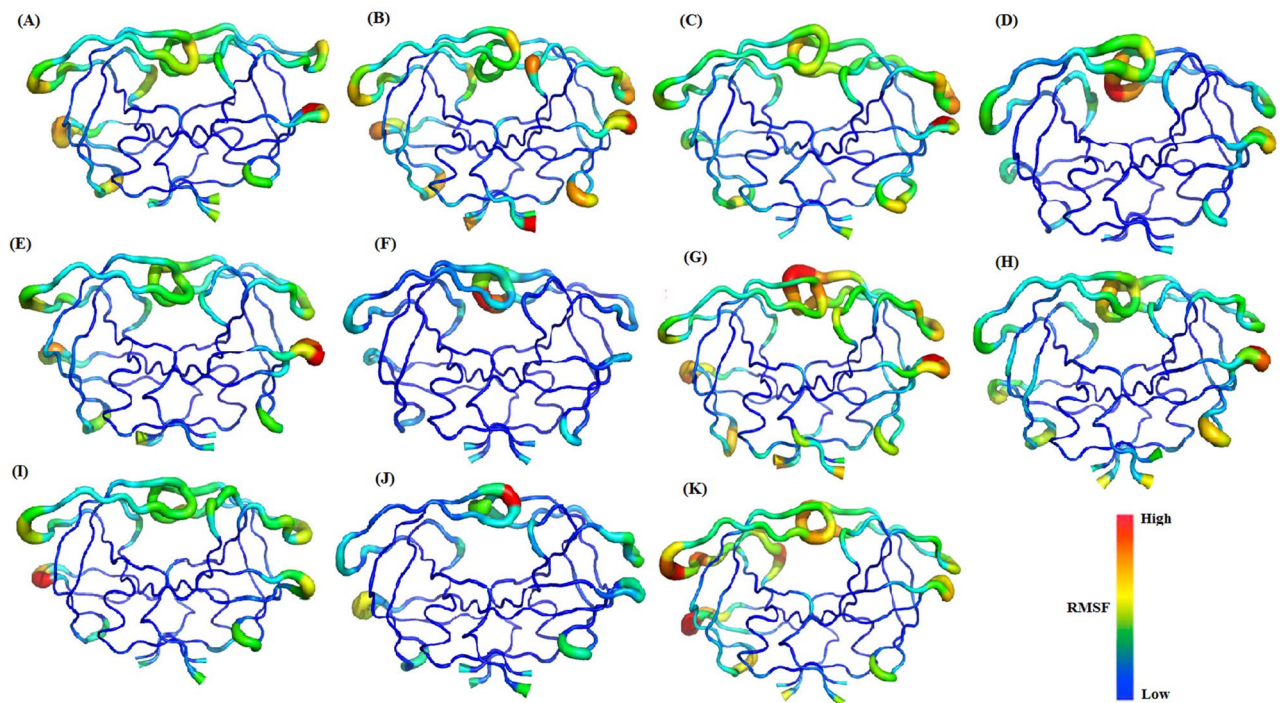
Due to the fact that the conformational dynamics of the flap region of the protease are vital for the catalytic activity, our analysis for fluctuation focused on the flap region. Interestingly, after the check of the flap region, which includes residues 43–58, we noticed higher fluctuations in the 1ODY, 1GNN, and 1AXA proteins compared to the WT; while other areas exhibited similar behavior. The handedness feature of the flaps was also visible in the RMSF values in which one of the flaps has more fluctuation than the other<sup>35</sup>. The RMSF results are consistent with that of the RMSD.

To visualize conformational fluctuations in selected proteins, we used a "sausage" plot to show the range of observed motions during simulation trajectories (Fig. 4). The thickness of the sausage plot is proportional to B-factor values and shows the extent of protein chain motion. The thinner segments denote the most stable regions of the protein, while thicker segments represent the more mobile regions. In most of the selected proteases, the highest mobility is found in the flaps and flap elbow regions, which should have significant functional implications because these regions are near the active site.

**Radius of gyration (Rg).** The Rg calculated from the MD trajectory indicates the compactness or rigidity of a protein system during the simulation. Higher Rg values indicate less compactness of protein structure, while lower Rg values indicate more stability and compactness<sup>36</sup>. We performed Rg analysis to observe the conformational alterations and dynamic stability of the wild-type and mutant structures. To understand the changes in Rg with time, a plot was constructed (Fig. 3C). The average values of Rg calculated for each protein can be found in Table 1. The Rg plots of all the protein systems show fluctuations ranging lesser than 2 Å, which shows that the protein systems are stable. The average Rg values for 1HVI, 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY, and 1PRO were found to be 1.759, 1.771, 1.766, 1.767, 1.761, 1.772, 1.754, 1.768, 1.769, 1.762 and 1.773 nm, respectively (Table 1). According to the fluctuation score, we ranked the collected values as follows: 1PRO > 1GNN > 1A9M > 1ODX > 1MET > 1AXA > 1AAQ > 1ODY > 1GNM > 1HVI (WT) > 1MES. The Rg plot suggested that the 1MES has tight packing than other proteins. Protein 1PRO showed a larger radius of gyration than other proteins, indicating that 1PRO, is less tightly packed. During the simulation, the WT and mutant proteins showed almost a similar pattern in terms of Rg values, indicating there were no important changes in

Proteins	Average backbone RMSD (nm)	Average Ca-RMSF (nm)	Average Rg-protein (nm)	Average SASA (nm <sup>2</sup> )
1HVI (WT)	0.146	0.0864	1.759	105.677
1A9M	0.121	0.0762	1.771	106.222
1AAQ	0.146	0.0904	1.766	106.254
1AXA	0.115	0.0817	1.767	105.925
1GNM	0.144	0.0887	1.761	105.518
1GNN	0.133	0.0804	1.772	107.363
1MES	0.133	0.0886	1.754	105.314
1MET	0.138	0.0919	1.768	107.140
1ODX	0.107	0.0816	1.769	105.898
1ODY	0.135	0.0848	1.762	105.717
1PRO	0.148	0.0797	1.773	106.057

**Table 1.** The calculated parameters for all the protein obtained after 50 ns MD simulations.



**Figure 4.** Sausage representation for the selected proteases [(A) 1HVI; (B) 1A9M; (C) 1AAQ; (D) 1AXA; (E) 1GNM; (F) 1GNN; (G) 1MET; (H) 1MES; (I) 1ODX; (J) 1ODY; (K) 1PRO]. The thickness of the sausage indicates the mobility of the region. The color scale goes from blue (low RMSF values—poorly flexible regions) to red (high RMSF values—very flexible regions). The figures were prepared using the PyMOL software.

the overall structure and folding of the protein after the mutation. In all of the proteins, the Rg results are in good agreement with that of RMSD and RMSF.

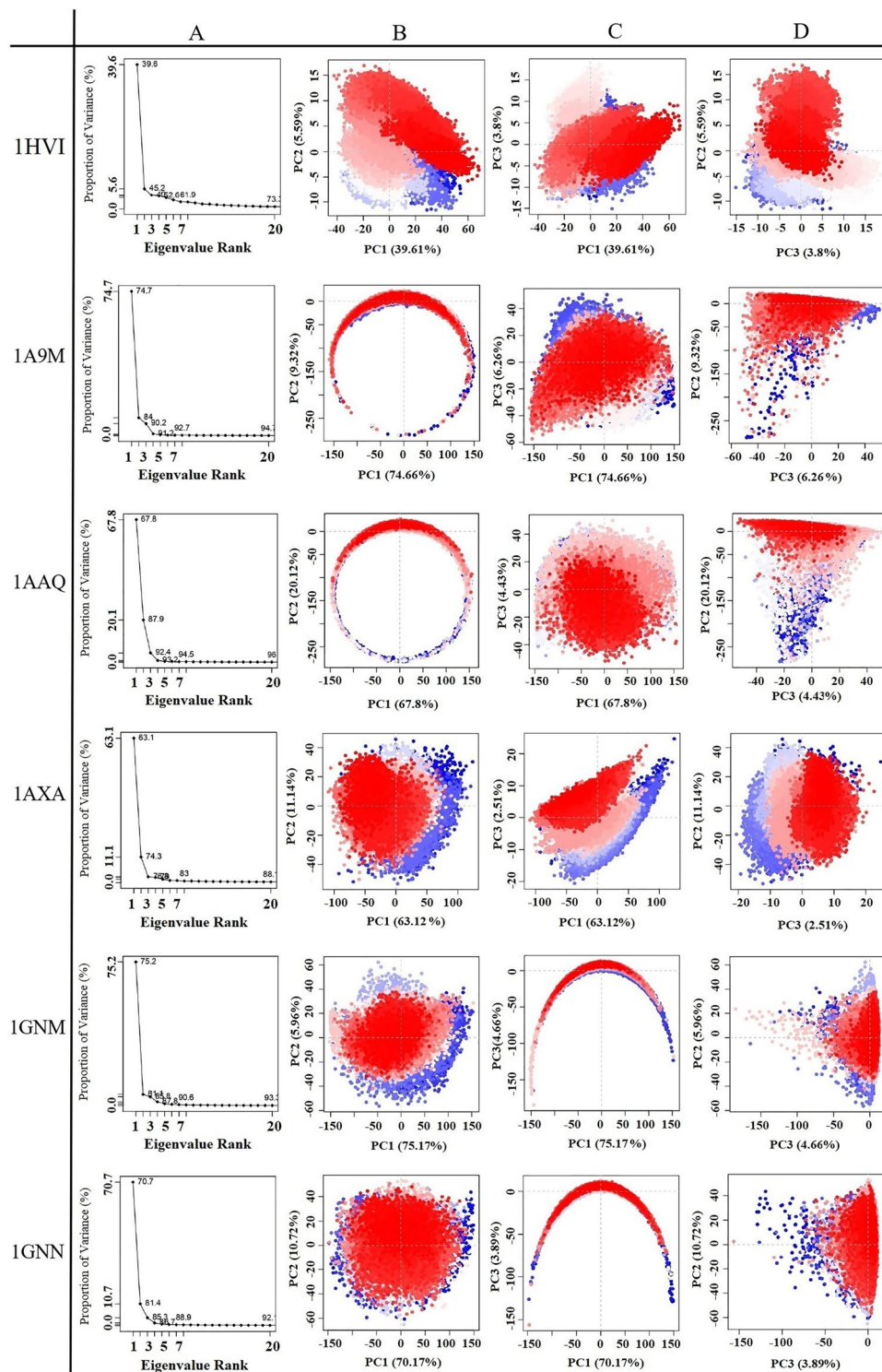
**Solvent-accessible surface area (SASA).** The SASA analysis is used to measure the degree to which an amino acid is exposed to its environment. A higher SASA value denotes a diffused protein structure, while a lower SASA value represents a compact structure. An increase or decrease in SASA value represents a change in the structural conformation of the protein<sup>37</sup>. The SASA values of the WT and ten mutated proteins were analyzed for predicting how the mutations affect the structure of the native protein. The SASA values calculated for the WT and ten mutated proteins with time are shown in Fig. 3D, and average SASA values are depicted in Table 1.

The rank of collected average SASA values are listed as: 1GNN (107.363 nm<sup>2</sup>) > 1MET (107.140 nm<sup>2</sup>) > 1AAQ (106.254 nm<sup>2</sup>) > 1A9M (106.222 nm<sup>2</sup>) > 1PRO (106.057 nm<sup>2</sup>) > 1AXA (105.925 nm<sup>2</sup>) > 1ODX (105.898 nm<sup>2</sup>) > 1ODY (105.717 nm<sup>2</sup>) > 1HVI (105.677 nm<sup>2</sup>) > 1GNM (105.518 nm<sup>2</sup>) > 1MES (105.314 nm<sup>2</sup>). According to the Rg analysis, no significant difference was found between the wild and mutated protein, and a similar effect was also observed in the case of the SASA profile. Thus, the SASA results are also in settlement with the RMSD, RMSF, and Rg results.

**Principal component analysis (PCA).** To gain deeper insight into the large-scale collective motions associated with conformation in the selected proteins, we performed PCA for analyzing the dominant protein conformational patterns in a principal components (PCs) phase space during 50 ns the MD simulations. In fact, we investigated the conformational behavior of the Ca atoms of the proteins by projecting them along the direction of the first three eigenvectors (PC1, PC2, and PC3).

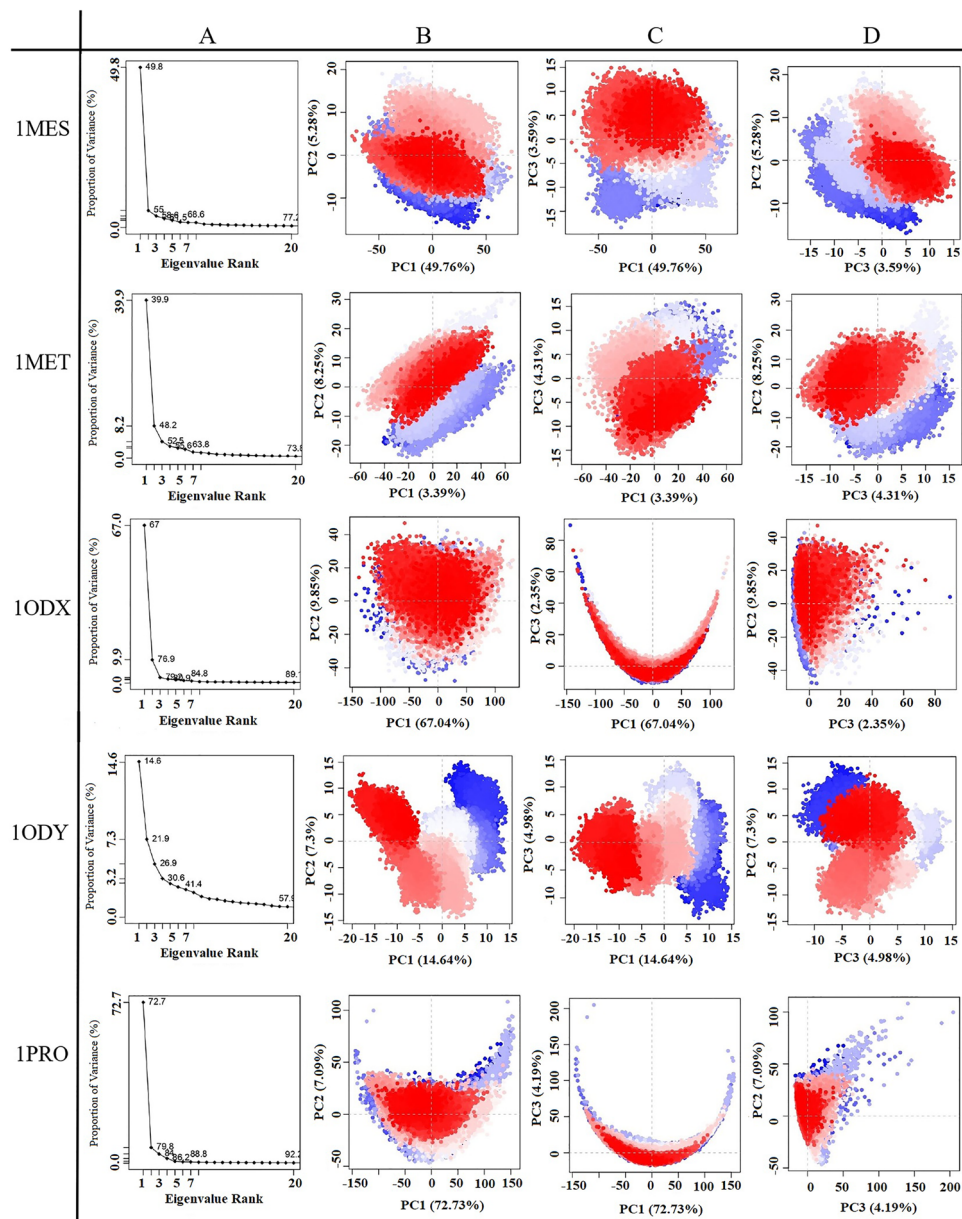
Figures 5 and 6 exhibit the first three PCs for selected proteins extracted from respective 50 ns MD simulation trajectories in the form of cluster groups. The 2D principal component plot between eigenvectors 1, 2, and 3 was drawn to compare acceptable conjoined motions. This 2D plot indicates the variations in the ensemble distribution for each conformation during the simulation interval, where each dot represents one conformation of the trajectory at a time *t*. The uninterrupted color representation (from blue to white to red) indicates the presence of substantial periodic bounces between conformers during MD simulations. There were three conformational states in all proteins, including the unstable conformational states (blue dots), the intermediate states (white dots), and the stable conformational states (red dots). These observations supported the compact and cluster motions for all selected proteins in their respective trajectory. Convincingly, dynamic motions of clusters in each extracted PC for the respective protein structure suggested the induction of collective fluctuation by point mutations as a function of the 50 ns MD simulation interval.

Principal components or eigenvectors are ranked according to decreasing eigenvalues, which directly correspond to their contribution to the overall conformational variance. Figures 5A and 6A show the scree plot of



**Figure 5.** Principal component analysis of the test proteins. This picture shows the proportion of variance (scree) plot (A), the projection of PC2 versus PC1 (B), PC3 versus PC1 (C), and PC2 versus PC3 (D) of the WT, 1A9M, 1AAQ, 1AXA, 1GNM and 1GNN, proteins during the simulation period.

eigenvalues for the first 20 PC, which indicates the percentage of the total variance (mean-square fluctuation) captured by each PC based on their eigenvalue rank. In this diagram, tags on each point indicate the cumulative sum of variance accounted that by a specific eigenvector and its previous eigenvectors. Indeed, cumulative variance is shown as a function of the number of PCs. This Figure shows the first few eigenvalues at the beginning are associated with larger concerted motions but decline rapidly and attain more localized fluctuations.



**Figure 6.** Principal component analysis of the test proteins. This picture shows the proportion of variance (scree) plot (A), the projection of PC2 versus PC1 (B), PC3 versus PC1 (C), and PC2 versus PC3 (D) of the 1MES, 1MET, 1ODX, 1ODY, and 1PRO proteins during the simulation period.

Percent of the variance and cumulative variance for each of the proteins during the 50 ns of MD simulations are presented in Table 2. According to Table 2, PCA showed that the top 20 PCs could capture 73.3%, 94.7%, 96.4%, 88.7%, 93.3%, 92.7%, 77.2%, 73.8%, 89.7%, 57.9% and 92.2% of total variance during the 50 ns of MD simulations in 1HVI, 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY and 1PRO proteins, respectively. From this result, also it was observed that the first three PCs were 49.0%, 90.2%, 93.2%, 76.7%, 85.7%, 85.3%, 58.6%, 52.4%, 79.2%, 26.9% and 84.0% of the total variance in the motion observed in the trajectories of 1HVI, 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY and 1PRO, respectively. According to Table 2, PCA suggests that the properties of the motions described by the first three PCs are clearly different for all the proteins. The scree plot of all proteins indicates that the elbow point is located at the third PC, meaning the first three PCs appear to be significant (Figs. 5 and 6). As shown in Figs. 5A and 6A, after the third PC, there are no momentous variations in the eigen fraction till 20 eigenvalues, indicating a state of convergence in the respective proteins. These observations suggested that significant flexibility was produced in all proteins during the initial phase of 50 ns MD simulation which eventually diminished to attain a stable system. Moreover, a steady decrease in the amplitude of an eigen fraction further indicates an additional localized fluctuation in the protein structure to attain a favorable conformation. Comparing all proteins, the highest and lowest PC1 magnitude is 75.1% and 14.6% for 1GNM and 1ODY, respectively. Except for 1ODY and 1MET, the magnitude



Protein	Principle component (PC)	Eigenvalue	Variance (%)	Cumulative variance (%)
1HVI (WT)	PC1	265.012	39.611	39.611
	PC2	37.407	5.591	45.203
	PC3	25.440	3.802	49.005
1A9M	PC1	2704.054	74.664	74.664
	PC2	337.502	9.319	83.983
	PC3	226.812	6.263	90.246
1AAQ	PC1	4278.441	74.335	74.335
	PC2	831.677	14.450	88.784
	PC3	255.046	4.431	93.216
1AXA	PC1	996.035	63.120	63.120
	PC2	175.818	11.142	74.261
	PC3	39.660	2.513	76.775
1GNM	PC1	2080.934	75.166	75.166
	PC2	164.879	5.956	81.121
	PC3	128.989	4.659	85.781
1GNN	PC1	1780.544	70.707	70.707
	PC2	269.991	10.722	81.428
	PC3	98.061	3.894	85.322
1MES	PC1	413.644	49.757	49.757
	PC2	43.914	5.282	55.040
	PC3	29.853	3.591	58.631
1MET	PC1	297.357	39.901	39.901
	PC2	61.484	8.250	48.151
	PC3	32.091	4.306	52.457
1ODX	PC1	1104.702	67.038	67.038
	PC2	162.356	9.852	76.890
	PC3	38.706	2.349	79.239
1ODY	PC1	50.291	14.642	14.642
	PC2	25.083	7.303	21.946
	PC3	17.118	4.984	26.929
1PRO	PC1	1766.330	72.730	72.730
	PC2	172.274	7.094	79.824
	PC3	101.688	4.187	84.011

**Table 2.** The eigenvalue, percent of the variance and cumulative variance for three principal components of each of the proteins during the 50 ns of MD simulations.

of PC1 in other proteins is significantly increased, which might correlate with the increased flap movement in the mutant proteins. The mutation in the 1MET protein did not alter the PC1 contribution significantly, but in the 1ODY protein, a lesser PC1 contribution was observed. These observations validated the result of higher flexibility of the mutated proteins compared to the native protein.

Except for 1ODY, the PC1 accounts for more than one-third of the total variance and strongly overcomes the total variance showing the global dynamics<sup>38</sup>. Interestingly, all the selected proteins, except 1ODY, showed a sharp increase in the percentage of variance corresponding to the first three PCs and covered more than > 50% of the total proportion of variance of atom positional fluctuations in each simulated protein. After that, the individual component contributions fall below 3%. In fact, these first three PCs account for a large proportion of the overall protein conformation and capture the most significant dominant motions, in other words, the fluctuations of the highest amplitude that are generally biologically relevant motions. These results showed that point mutations caused significant changes in the conformational motions of the selected proteins. Therefore, PCA dots images generated from the first three eigenvectors are used to observe the conformational transitions of these proteins<sup>39,40</sup>.

On a 2D principal component plot, the larger the cumulative variance on the two considered principal components, the more significant the distance between the points. It means that diverse conformations will have diverged while similar conformations will have grouped on the 2D PCA plot<sup>38</sup>. Comparing the 2D scatter plots of all proteins, it could be seen that the conformational states of the mutated proteins compared with wild protein had changed significantly. As shown in Figs. 5 and 6, except 1ODY, the PC2 versus PC1 plots for all proteins clearly indicated the conformers visit a large conformational space.

Evaluation of internal motions through the first three principal components shows that PC1 and PC2 are prominently related. As shown in Figs. 5 and 6, the PC2 versus PC1 plots obtained from the MD trajectories are

almost varied for all proteins, which display differences in motion across the two first principal components. These observations clearly indicated the differences in protein motion and the conformational landscape between the proteins. This indicates that mutations in these proteins have caused conformational changes. However, the presence of overlap between blue and red colors indicates that the protein does revisit the same state during the simulation, even though it undergoes conformational changes. Also is an indication of the quality of sampling, which illustrates the simulation time is sufficient. As shown in PC2 versus PC1 plots, in the case of comparing proteins with each other, the 1ODY protein has a smaller phase space and lower performance flexibility than other proteins. In the 1ODY protein, the contribution of PC1 and PC2 to the variance is 14.6% and 7.3%, respectively, while other PCs contributed no more than 5.0%.

It can be found from Figs. 5 and 6 that in PC2 versus PC1 diagrams, the points of all proteins are almost evenly distributed near the midline, while in PC2 versus PC3 diagrams, the points of proteins are differently distributed. The closer the point distribution in Figs. 5 and 6 indicates that the protein system is more stable, so the 1ODX and 1AXA proteins are in a more stable state. In another word, 1ODX and 1AXA proteins exhibited the most favorable converged conformations and limited variation against other proteins during MD simulation; suggesting considerable stability as noted from the respective RMSD and RMSF values.

In some PC2 versus PC1 and PC3 versus PC1 diagrams, the projection of the points of proteins reveals a semicircle or U-shape pattern. This type of pattern has been attributed to random diffusion of motion in proteins, allowing only to inform on more accessible degrees-of-freedom for thermal motion along our studied time scale<sup>41</sup>.

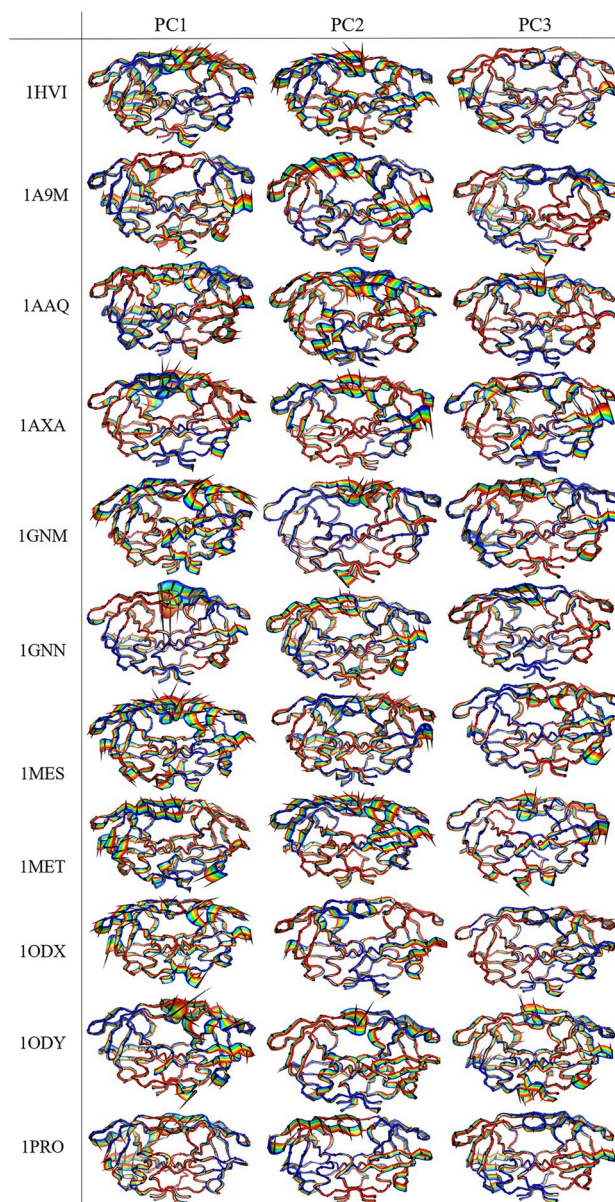
Point mutations increase overall collective motions in selected proteins because these mutations effectively increase the Ca movements of proteins. In fact, these mutations have increased the flexibility of proteins. Except for 1ODY, the conformational space covered in other proteins proved to be broader than that of 1HVI. These results mean that after point mutation, the dynamics of the mutated proteins change, which may be required for proper protein function. Also, these fluctuations registered in each protein may be regarded as a requirement for the stability of the relevant protein during MD simulation as a function of time<sup>41</sup>. The PCA results are consistent with the RMSD, RMSF, Rg, and SASA results.

**Porcupine plot.** The main motions of protein residues can be better observed and analyzed by displaying eigenvectors as porcupine plots<sup>42</sup>. Porcupine plots are drawn using the custom-made program PyMOL<sup>43</sup> to visualize the movements of the first three PCs obtained from the principal component analysis. It corresponds to an outline of protein motion, suggesting what part of the protein moves in concert and in which direction. In fact, the first and last eigenvectors from any PC were generated using the PyMOL tool and presented as a porcupine plot. The extent and direction of the most dominant motions of all proteins were visualized through porcupine plots using the 'modevectors.py' script (written by Sean M. Law) in PyMOL version 1.7. The plot shows a cone for each Ca atom reflecting the direction of its motion, where the length of the cone indicates the motion amplitude and the size of the cone specifies the number of such Ca atoms. The linear interpolations between the first and last eigenvectors are shown with the color transition from blue to red to highlight the conformational differences between them.

The porcupine plots showing the motions of all selected proteins, along the directions of PC1, PC2, and PC3 are presented in Fig. 7. The cones in black represent the direction of the concerted motion, and the length of the cones represents the extent of the motion. Based on the porcupine plots, it is evident that point mutation increases the overall movements of all mutant proteins as compared to the wild-type protein. All proteins showed anti-symmetric movement patterns in the flap, flap elbow, fulcrum, and cantilever regions. In all proteins, the flap and flap elbow regions experienced more flexibility compared to the different segments of the protease, as indicated by the length of the vectors. Thus, it is obvious that in all proteases the flaps and flap elbow motions are highly correlated. Flexible regions of the proteins are not simply the result of loose packing or instability but have been evolutionarily selected<sup>44</sup>. Considering that the conformational flexibility of the flap region is necessary for the activity of the HIV-1 protease, it can be concluded that the flap regions are one of the most important signs of the evolution of conformational dynamics in HIV-1 protease.

**Dynamical cross-correlated map (DCCM) analysis.** To investigate further the effects of point mutation on the conformational dynamics, DCCM analysis was used for Ca atom fluctuations during the 50-ns MD simulation for analyzing the correlation motions of each protein. The cross-correlation analysis could show the relationships between residues and between various regions by quantifying their relative motions. The correlation was normalized and varied from -1.0 and +1.0 (from dark blue to white to red) (Fig. 8). The positive correlation (red regions) represented residues moved in the same direction and the negative correlation (blue regions) meant residues moved in the opposite direction<sup>45</sup>. The deeper color indicated a stronger positive correlation or negative correlation. The white regions ranging from -0.25 to 0.25 were considered as a low correction. A diagonal point represents the Ca atom of the same residue along both axes; therefore, the diagonal elements show the maximum correlation.

The cross-correlation matrix of the C-alpha displacement indicated complex correlated and anti-correlated motions in the wild-type protein and all mutant proteins except 1ODY. The diagonal parts show obvious correlated movements. In the cross-correlation matrix, the  $C_{ij}$  elements of the matrix were found to be symmetrical about the diagonal. Interestingly, in all mutant proteins, correlated and anti-correlated motions increase. The only exception is the 1ODY mutant in which correlated and anti-correlated motions highly decrease. For the mutated proteins, we found deeper shades of red and blue, distinguishing regions of high correlation and anti-correlation that corroborate the fact that mostly global motions are seen for these proteins. In most mutant proteins, except 1ODY, almost all residues have concerted motions and they move in an anti-correlated fashion with the rest



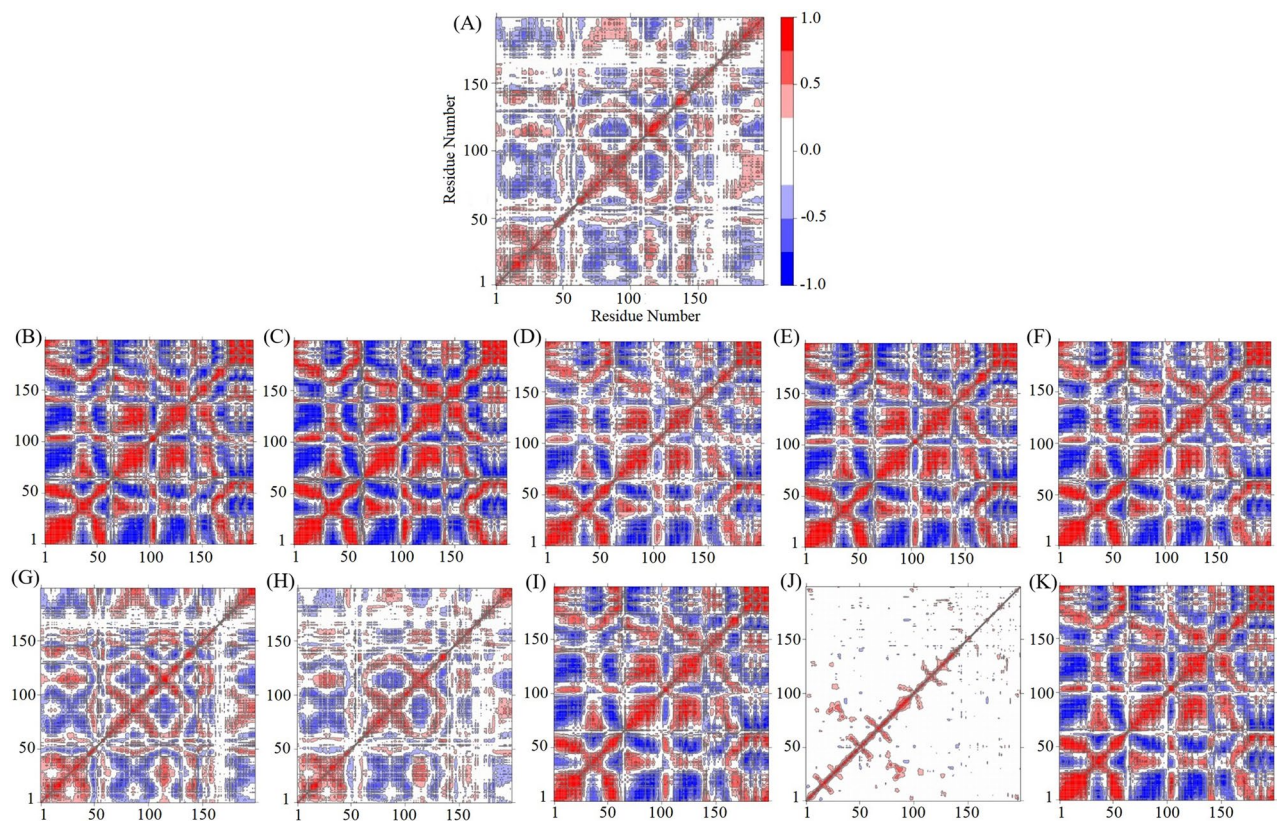
**Figure 7.** Porcupine plots corresponding to PC1, PC2, and PC3 obtained from PCA on the MD trajectories of the wild-type protein and the mutant proteins. The cones in black represent the direction of the concerted motion, and the length of the cones represents the magnitude of the movement.

of the structure (Fig. 8). The point mutation has significantly changed the direction and the cooperativity of motion in 1ODY.

It can also be observed that the motion of atoms in all mutant proteins as compared to the wild-type protein is more compact, while the 1ODY mutated protein represented a very deformed form of the backbone atomic motion. The 1MES and 1MET proteins have approximately the same correlation and anti-correlation motions as the 1HVI protein, however, in these two mutant proteins, the intensity of the correlated and anti-correlation motions is slightly higher than that of the 1HVI protein. Thus, our results show that variations in sequence can lead to changes in dynamics without altering structure.

As seen from principal components and the dynamic cross-correlation maps, the point mutations change the major motions of the proteins and may improve the dynamic behavior of proteins. These results indicate that point mutations may have a pronounced effect on the conformational flexibility of proteins.

The conformational dynamics in the flaps of HIV-1 protease plays a critical role in the mechanism of substrate or inhibitor binding. Opening of the flaps is essential for the entrance of substrate into the binding cleft and flaps in the closed conformation sturdy the substrate for catalysis. The protease dimer is in dynamic equilibrium between the closed conformation and different open conformational states. These conformational changes are highly associated with the flexibility of the flaps of the protease<sup>46</sup>. Molecular dynamics studies show that



**Figure 8.** Comparison of cross-correlation matrices of wild-type and mutant proteins. Cross correlation matrix of C-alpha atoms during 50 ns simulation for wild-type protein (A) and mutant proteins of 1A9M (B), 1AAQ (C), 1AXA (D), 1GNM (E), 1GNN (F), 1MES (G), 1MET (H), 1ODX (I), 1ODY (J), and 1PRO (K). The range of motion is shown by different colors in the panel. Red color denotes positive correlation while blue color denotes anti-correlation.

mutations in the flap regions may affect the dynamics of the flap and thus the binding of protease inhibitors<sup>47</sup>. The results revealed that the mutations caused increased movement in the flap, and flap elbow regions. The presence of a region with such conformational dynamic diversity in the protein that also provides the necessary flexibility to perform the proper function of the protein indicates that the protein has also evolved dynamically in addition to structural evolution. Indeed, since the flexible regions of proteins are selected evolutionarily, it can be inferred that the flap regions are the most significant sign of the evolution of conformational dynamics in the HIV-1 protease.

Previous studies show that dynamic structural regions exist in various proteins whose conformational flexibility is essential for the function and allostery of proteins (Table 3). Thus, it is time to group these dynamic structural regions and choose a specific and unique name for these areas. Our suggestion for naming these regions is dynamozones. Dynamozones are dynamic structural regions in some proteins that contribute to the biological function and allostery of proteins via their convenient flexibility. These regions provide the flexibility needed for proteins to suitable function, and their presence is completely essential for protein function. One of the characteristics of these regions in most proteins is the presence of the amino acid glycine in their structure. Dynamozones are a subset of three main groups: loops, hinges, and linkers. One of the clear signs of the evolution of conformational dynamics of proteins is the presence of dynamozones in proteins because these dynamic regions have evolved in such a way that they workable the function and allostery of proteins. These regions via different mechanisms such as performing opening-closing movements for ligand binding to the active site, appropriate motions of the loops for placement of the catalytic residues in the suitable position for catalysis, switching from an inactive "out" conformation to an active "in" conformation to create the catalytically active form of the enzyme, proper motions of linkers in proteins to accept compact and extended conformations, movements of hinges to connect antibodies to antigens, etc. help various proteins to perform their proper function. Thus, we can say that dynamozones are a footprint of the evolution of the conformational dynamics of proteins. The details of some of the know dynamozone members are listed in Table 3.

## Conclusions

Protein dynamics, structure, and function are highly correlated. In the case of the HIV-1 protease, the dynamics of conformational changes are very critical for enzyme function. In the present study, we have explored the effect of mutations on the dynamic behavior of the WT and mutant HIV-1 proteases variants using a combination of MD simulations, cross-correlation analysis, and PCA. Sequence similarities between proteins were strikingly

Structure	Dynamozones	Proteins	Role or Function	References
Loop	Flap	<ul style="list-style-type: none"> <li>• HIV protease</li> <li>• Plasmepsins</li> <li>• Beta-secretase</li> <li>• Cathepsin</li> <li>• Pepsins</li> </ul>	These regions control the entrance and stabilization of ligands in the active site	67–70
Loop	Loop (Residues 166–176)	<ul style="list-style-type: none"> <li>• Triosephosphate isomerase</li> </ul>	In the ligand-bound state, the loop moves for $\sim 7 \text{ \AA}$ as a rigid lid toward the active site and accepts a “closed” conformation. These motions of the rigid lid close to the active site are essential for the catalytic mechanism of the enzyme	71–73
Loop	Loop	<ul style="list-style-type: none"> <li>• Enolases</li> <li>• Aldolases</li> </ul>	Movements of the loop permit the catalytic residues to be oriented in a suitable position for catalysis	74–76
Loop	WPD loop	<ul style="list-style-type: none"> <li>• Protein tyrosine phosphatases (PTPs)</li> </ul>	This loop closes over the active site upon binding of the substrate, and loop closure permits the correct orientation of catalytic residues around the ligand	77,78
Loop	Met20 loop (Active site loop (	<ul style="list-style-type: none"> <li>• Dihydrofolate reductase (DHFR)</li> </ul>	This loop acts as a lid that closes on the cofactor, thereby allowing DHFR to adopt occluded and closed conformations	79–81
Loop	Helical loop	<ul style="list-style-type: none"> <li>• Lipases</li> </ul>	This loop is important for the enzyme function, acting as a lid to open or close the hydrophobic active site	82,83
Loop	Long loop	<ul style="list-style-type: none"> <li>• <math>\beta</math>1,4-galactosyl transferase</li> </ul>	A displacement of more than $20 \text{ \AA}$ this long loop in protein provides binding sites for various ligands	84
Loop	Omega loop	<ul style="list-style-type: none"> <li>• Cdc34-like E2 enzymes</li> </ul>	This loop can act as a lid that regulates the accessibility of the catalytic site and disturbs the charging activity of ubiquitin until a conformational change toward an open state is promoted by phosphorylation	85–87
Linker	Flexible linker	<ul style="list-style-type: none"> <li>• Calmodulin (CaM)</li> </ul>	CaM has two globular domains connected by a short and flexible linker that permits the protein to accept a wide variety of extended and compact conformations	88–90
Hinge Loop Linker	<ul style="list-style-type: none"> <li>• Hinge region</li> <li>• P-loop</li> <li>• hydrophobic “spines”</li> <li>• A-loop</li> <li>• <math>\alpha</math>C helix</li> <li>• DFG motif</li> <li>• <math>\alpha</math>B helix</li> </ul>	<ul style="list-style-type: none"> <li>• Protein kinases (PKs)</li> <li>• Tyrosin kinase</li> <li>• Src protein kinase</li> </ul>	<p>Hinge region: The hinge motion is necessary for the opening and closure of the kinase catalytic domain (CD)</p> <p>P-loop (<math>\beta</math>1-<math>\beta</math>2 loop or G-loop or Gly-rich loop): This flexible loop is very important for the coordination of ATP phosphates</p> <p>Hydrophobic “spines”: Two hydrophobic “spines” link the two lobes of protein kinase and dynamically connect all the elements important for catalysis</p> <p>A-loop (Activation loop): In the inactive state of the enzyme, the A-loop is folded onto itself, and its opening is required to create the catalytically active form</p> <p><math>\alpha</math>C helix: This helix in the “in” active conformation forms a hydrogen bond with the <math>\beta</math>3 strand for creating the catalytically active form of the enzyme</p> <p>DFG motif: This motif in the active site switches from an inactive (DFG-out) conformation to an active (DFG-in) conformation, which is necessary to create the catalytically active form of the enzyme</p> <p><math>\alpha</math>B helix: This helix creates a cavity, the so-called PIF pocket, which is very important for allosteric regulation of the protein kinases belonging to the AGC family</p>	91–96
Hinge	Hinge region	<ul style="list-style-type: none"> <li>• Lactoferrin</li> </ul>	The hinge motions permit the formation of the complete iron-binding site in the closed states of lactoferrin	97–100
Hinge	Hinge region	<ul style="list-style-type: none"> <li>• Immunoglobulins</li> </ul>	The hinge region is connecting the Fab (Fragment antigen binding) region to the Fc (Fragment crystallizable) region	101,102

**Table 3.** List of some known members of Dynamozone.

similar and confirmed the assumption that proteins with greater sequence similarity behave similarly. The results indicate that mutations not only produce important effects on the correlated motions and flexibility of HIV-1 protease but also increase the stability of HIV-1 protease during the simulation period. These positively selected mutations introduce significant flexibility in important regions such as the flap and flap elbow. Further, the parameters such as RMSF, DCCM, porcupine plot, and the PCA analysis revealed increased fluctuation/motion in the mutated proteases in comparison to the wild-type structure. Our main results were as follows: (1) The flap regions are the most evident indication of the evolution of conformational dynamics in HIV-1 protease and are an excellent case for investigating the evolution of conformational dynamics; (2) Dynamozones are dynamic structural regions in some proteins that donate to the biological function and allostery of proteins through their proper flexibility; (3) Because of the presence of other members of dynamozones in various proteins, we offer to consider dynamozones as an indication of the evolution of the conformational dynamics of proteins. The flap regions are one of the most significant dynamozone members that are crucial for HIV-1 protease function.

## Computational methods

**Protein structures preparation.** A set of eleven HIV-1 protease proteins with experimentally determined structures chosen from the Protein Data Bank (PDB), were used in this study<sup>48</sup>. The selected proteins’ structures have been determined using X-ray crystallography, and their resolution and R-factor are less than 3.0 and 0.19, respectively. Molecular dynamics simulation approaches pave the way for an in-depth analysis of the effects of mutations on protein structure and dynamics. In the present study, mutations G48H, L63I, A28S, V82D, V82N, (I3V; I84V), (I3V; V82F), (A71T; V82A), (I3V; R41A) and S37N, which are related to structures 1A9M, 1AAQ, 1AXA, 1GNM, 1GNN, 1MES, 1MET, 1ODX, 1ODY and 1PRO, respectively, were analyzed to investigate the effect of mutations on the conformational dynamics of HIV-1 protease using molecular dynamics simulations (Fig. 1B). The PDB code: 1HVI was also selected as a wild-type protein. We selected a suitable mutation from

each of the different parts of the HIV-1 protease to study the dynamic behavior of the proteases. The ligands, ions and water molecules were removed from the protein structures. The details about the selected HIV-1 protease variants are listed in Table 4.

**Molecular dynamics simulation.** In this study, all of the MD simulations were performed using the GROMACS 2016.3<sup>49</sup>. The force field parameters were assigned according to the Amber99SB\*-ILDN force field<sup>50</sup>. Protonation states were assigned to each structure using PDB2PQR<sup>51</sup> through ProPKa<sup>52</sup> at pH 7.0. The TIP3P water model<sup>53</sup> was used to solvate the system, which was generated as a cubic box like area with a side of 1 nm such that the protein is covered appropriately with water molecules. All proteins were electrostatically neutralized by adding chlorine ions around the molecules. Each protein was minimized by the steepest descent algorithm up to a maximum of 50,000 steps and a convergence tolerance of 10 kJ mol<sup>-1</sup> nm<sup>-1</sup>. The energy-minimized structure of the native protein and ten mutant proteins were used as the starting points for the MD simulations. Equilibration of the proteins was conducted in two phases NVT and NPT ensembles each for 100 ps. Particle-Mesh-Ewald (PME) method<sup>54</sup> with a 1.0 nm cutoff was used to calculate the electrostatic interactions. During the MD simulation, the LINCS algorithm<sup>55</sup> was used to constrain all the bonds. The temperature was kept constant (300 K) and pressure was maintained at 1 bar using the V-rescale thermostat<sup>56</sup> and the Parrinello-Rahman barostat<sup>57</sup>, respectively. Following the equilibrium methods, MD simulations were performed for each of the native and mutant proteins with 3 repetitions for 50 ns.

**Analysis of trajectory files.** The resulting trajectory files of the simulations were studied employing different parameters existing in GROMACS utilities. From the refined trajectories, various parameters, such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), and solvent-accessible surface area (SASA) for all proteins were calculated. The sequence alignments were created with T-Coffee<sup>58</sup> and ESPript<sup>59</sup>. PCA was carried out using the Bio3D package<sup>60</sup> and used to reveal the changes in the motion patterns of the protein systems.

To exclude the possibility of stochasticity and to confirm the first simulation performed, all 11 selected structures were simulated with different initial velocities with three repeats for 50 ns.

PCA analysis which is explained in detail below is one of the main analyzes performed for this research. To confirm that our simulation time is sufficient to investigate the overall dynamic behavior of the protein, we performed the cosine content analysis, which is part of the PCA method. As a result of this analysis, we can determine whether the protein exhibits random diffusion dynamics during the simulation. A cosine content close to 1 indicates random motion in the protein and non-convergence of the simulation time in the selected time interval, so it cannot be considered for PCA analysis. It has been reported that the cosine content close to 0.2 and sometimes up to 0.5 indicates the non-random diffusion dynamics of the protein during the simulation time and is reliable for further analysis, such as the free energy landscape (FEL)<sup>61,62</sup>.

In this study, we reported the replicates that had cosine content values of the first two eigenvectors (PC1 and PC2) close to 0. Based on chosen principal components (PC1 and PC2), we generated the FEL to calculate Gibbs's free energy value for selected structures. The cosine content values, Gibbs's free energy values, and RMSD of the simulated structures are reported in the supplementary data file (Figures S1-S11, Table S1).

**Principal component analysis.** A principal component analysis was performed to investigate conformational flexibility and the collective motions of the selected proteins using the Bio3D package installed in the R program. This method is based on diagonalization of the covariance matrix of atomic fluctuations to obtain

Simulated proteins (PDB ID)	Residue position	Amino acid change	Property change
1HVI (WT)			
1A9M	48	G → H	Hydrophilic/+ *
1AAQ	63	L → I	Hydrophobic/Hydrophobic
1AXA	28	A → S	Hydrophobic/Hydrophilic
1GNM	82	V → D	Hydrophobic/-**
1GNN	82	V → N	Hydrophobic/Hydrophilic
1MES	3	I → V	Hydrophobic/Hydrophobic
	84	I → V	Hydrophobic/Hydrophobic
1MET	3	I → V	Hydrophobic/Hydrophobic
	82	V → F	Hydrophobic/Hydrophobic
1ODX	71	A → T	Hydrophobic/Hydrophilic
	82	V → A	Hydrophobic/Hydrophobic
1ODY	3	I → V	Hydrophobic/Hydrophobic
	41	R → A	+ /Hydrophobic
1PRO	37	S → N	Hydrophilic/Hydrophilic

**Table 4.** Overview of the amino acid changes in the HIV-1 protease families. \* + Positively charged amino acids; \*\*– negatively charged amino acids.

orthogonal eigenvectors and the corresponding eigenvalues. The eigenvectors are the principal components that represent the directions of the coordinated motions of atoms. The eigenvalues indicate the magnitude of the motions along the movement direction<sup>63</sup>. The ensemble formula used to obtain a covariance matrix with elements  $C_{ij}$  for coordinates  $i$  and  $j$  is given as:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (i, j = 1, 2, 3, \dots, 3N) \quad (1)$$

where  $x_i$  and  $x_j$  are the mass-weighted Cartesian coordinates of the  $i$ th and  $j$ th Ca atoms,  $N$  is the number of the Ca atoms considered, and  $\langle x_i \rangle$  and  $\langle x_j \rangle$  represent the time average over all the configurations obtained in MD simulation. In this study, by the Bio3D package, Ca atoms from 50,000 frames obtained through 50 ns trajectory were superimposed on the initial pose to minimize the root mean square variations between the equivalent residues using fit.xyz function<sup>64</sup>.

**Dynamic cross-correlation map analysis.** The cross-correlation analysis can provide information about the impact of mutations on protein dynamics by analyzing how atomic displacements were coupled<sup>65</sup>. The extent of correlative motion of two atoms (or two residues) can be denoted by the cross-correlation coefficient,  $C_{ij}$ . It is defined by:

$$C_{ij} = \frac{\langle \Delta x_i \cdot \Delta x_j \rangle}{(\langle \Delta x_i \rangle^2 \langle \Delta x_j \rangle^2)^{1/2}} \quad (i, j = 1, 2, \dots, 3N) \quad (2)$$

where  $i$  ( $j$ ) means  $i$ th ( $j$ th) residue (or atom),  $\Delta x_i$  and  $\Delta x_j$  are the displacements from the mean position of  $i$ -th and  $j$ -th residues (or atoms), with respect to time, respectively, and  $N$  represents the number of Ca atoms. The angular brackets “ $\langle \rangle$ ” illustrate the time average on the whole trajectory. The value of the cross-correlation coefficient is from  $-1$  to  $+1$ . The positive value implies positively correlated movement (moving in the same direction), and the negative value implies anti-correlated movement (moving in the opposite direction). Higher values of the absolute value of  $C_{ij}$  show two residues (or atoms) are more correlated (or anti-correlated)<sup>66</sup>. In this article, we calculated the cross-correlations for all Ca atomic fluctuations extracted from the MD trajectory by using the Bio3D packages of R.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 21 February 2023; Accepted: 17 August 2023

Published online: 30 August 2023

## References

- Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
- Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci.* **102**, 6679–6685 (2005).
- Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: Signaling and other functions. *Curr. Opin. Struct. Biol.* **20**, 142–147 (2010).
- Wlodawer, A. Rational approach to AIDS drug design through structural biology. *Annu. Rev. Med.* **53**, 595–614 (2002).
- Schweizer, L. & Mueller, L. *Biased Signaling in Physiology, Pharmacology and Therapeutics* (Elsevier, 2014).
- Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
- Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
- James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
- Khersonsky, O., Roodveldt, C. & Tawfik, D. S. Enzyme promiscuity: Evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
- Fornili, A., Pandini, A., Lu, H.-C. & Fraternali, F. Specialized dynamical properties of promiscuous residues revealed by simulated conformational ensembles. *J. Chem. Theory Comput.* **9**, 5127–5147 (2013).
- Kimura, M. & Ohta, T. On some principles governing molecular evolution. *Proc. Natl. Acad. Sci.* **71**, 2848–2852 (1974).
- Marsh, J. A. & Teichmann, S. A. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *BioEssays* **36**, 209–218 (2014).
- Campitelli, P., Modi, T., Kumar, S. & Ozkan, S. B. The role of conformational dynamics and allostery in modulating protein evolution. *Annu. Rev. Biophys.* **49**, 267–288 (2020).
- Tang, Q.-Y. & Kaneko, K. Dynamics-evolution correspondence in protein structures. *Phys. Rev. Lett.* **127**, 098103 (2021).
- Amadei, A., Linssen, A. B. & Berendsen, H. J. Essential dynamics of proteins. *Proteins Struct. Function Bioinf.* **17**, 412–425 (1993).
- Wlodawer, A. & Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).
- Navia, M. A. *et al.* Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**, 615–620 (1989).
- Tomasselli, A. G. & Heinrikson, R. L. Targeting the HIV-protease in AIDS therapy: A current clinical perspective. *Biochim. Biophys. Acta BBA Protein Struct. Mol. Enzymol.* **1477**, 189–214 (2000).
- Perryman, A. L., Lin, J. H. & McCammon, J. A. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **13**, 1108–1123 (2004).
- Harte, W. Jr. *et al.* Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci.* **87**, 8864–8868 (1990).
- Cui, Q. & Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (CRC Press, 2005).
- Batista, P. R. *et al.* Free energy profiles along consensus normal modes provide insight into HIV-1 protease flap opening. *J. Chem. Theory Comput.* **7**, 2348–2352 (2011).

23. Batista, P. R. *et al.* Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis—application to the HIV-1 protease. *Phys. Chem. Chem. Phys.* **12**, 2850–2859 (2010).
24. Tóth, G. & Borics, A. Closing of the flaps of HIV-1 protease induced by substrate binding: A model of a flap closing mechanism in retroviral aspartic proteases. *Biochemistry* **45**, 6606–6614 (2006).
25. Tóth, G. & Borics, A. Flap opening mechanism of HIV-1 protease. *J. Mol. Graph. Model.* **24**, 465–474 (2006).
26. Sadiq, S. K., Noé, F. & De Fabritiis, G. Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc. Natl. Acad. Sci.* **109**, 20449–20454 (2012).
27. Nicholson, L. K. *et al.* Flexibility and function in HIV-1 protease. *Nat. Struct. Biol.* **2**, 274–280 (1995).
28. Ding, F., Layten, M. & Simmerling, C. Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments. *J. Am. Chem. Soc.* **130**, 7184–7185 (2008).
29. Lesk, A. M. & Chothia, C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225–270 (1980).
30. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826 (1986).
31. Landau, M. *et al.* ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
32. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98**, 861–871 (2010).
33. Knapp, B., Frantal, S., Cibena, M., Schreiner, W. & Bauer, P. Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible?. *J. Comput. Biol.* **18**, 997–1005 (2011).
34. Kumar, K., Anbarasu, A. & Ramaiah, S. Molecular docking and molecular dynamics studies on  $\beta$ -lactamases and penicillin binding proteins. *Mol. Biosyst.* **10**, 891–900 (2014).
35. Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **103**, 915–920 (2006).
36. Khan, S. A., Zia, K., Ashraf, S., Uddin, R. & Ul-Haq, Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *J. Biomol. Struct. Dyn.* **39**, 2607–2616 (2021).
37. Mazola, Y. *et al.* A comparative molecular dynamics study of thermophilic and mesophilic  $\beta$ -fructosidase enzymes. *J. Mol. Model.* **21**, 1–11 (2015).
38. Hayward, S. & De Groot, B. L. Normal modes and essential dynamics. *Mol. Model. Proteins* **443**, 89–106 (2008).
39. Maisuradze, G. G., Liwo, A. & Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **385**, 312–329 (2009).
40. Chen, J. Drug resistance mechanisms of three mutations V32I, I47V and V82I in HIV-1 protease toward inhibitors probed by molecular dynamics simulations and binding free energy predictions. *RSC Adv.* **6**, 58573–58585 (2016).
41. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E* **62**, 8438 (2000).
42. Haider, S., Parkinson, G. N. & Neidle, S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J.* **95**, 296–311 (2008).
43. DeLano, W. L. *The PyMOL Molecular Graphics System Version, 1.8* (Schrödinger, 2002).
44. Sinha, N. & Smith-Gill, S. J. Protein structure to function via dynamics. *Protein Pept. Lett.* **9**, 367–377 (2002).
45. Yuan, Y., Zhan, W., Yi, H., Zhao, Y. & Song, S. Molecular dynamics simulations study for the effect of cations hydration on the surface tension of the electrolyte solutions. *Colloids Surf. A* **539**, 80–84 (2018).
46. Freedberg, D. I. *et al.* Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* **11**, 221–232 (2002).
47. Meher, B. R. & Wang, Y. Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): Molecular dynamics simulation and binding free energy studies. *J. Phys. Chem. B* **116**, 1884–1900 (2012).
48. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
49. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
50. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinf.* **65**, 712–725 (2006).
51. Dolinsky, T. J. *et al.* PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
52. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Protein Struct. Funct. Bioinf.* **61**, 704–721 (2005).
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
54. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
55. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
56. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
57. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
58. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
59. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
60. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
61. Topno, N. S., Kannan, M. & Krishna, R. Interacting mechanism of ID3 HLH domain towards E2A/E12 transcription factor—An insight through molecular dynamics and docking approach. *Biochem. Biophys. Rep.* **5**, 180–190 (2016).
62. Pandey, B. *et al.* Alanine mutation of the catalytic sites of Pantothenate Synthetase causes distinct conformational changes in the ATP binding region. *Sci. Rep.* **8**, 903 (2018).
63. Ichiye, T. & Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct. Funct. Bioinf.* **11**, 205–217 (1991).
64. Gedeon, P. C., Thomas, J. R. & Madura, J. D. Accelerated molecular dynamics and protein conformational change: A theoretical and practical guide using a membrane embedded model neurotransmitter transporter. *Mol. Model. Proteins* **1215**, 253–287 (2015).
65. Wan, H., Hu, J.-P., Tian, X.-H. & Chang, S. Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d. *Phys. Chem. Chem. Phys.* **15**, 1241–1251 (2013).
66. Chillemi, G. *et al.* Thr729 in human topoisomerase I modulates anti-cancer drug resistance by altering protein domain communications as suggested by molecular dynamics simulations. *Nucleic Acids Res.* **36**, 5645–5651 (2008).



67. Louis, J. M., Ishima, R., Torchia, D. A. & Weber, I. T. HIV-1 protease: Structure, dynamics, and inhibition. *Adv. Pharmacol.* **55**, 261–298 (2007).
68. Lee, J. & Goodey, N. M. Catalytic contributions from remote regions of enzyme structure. *Chem. Rev.* **111**, 7595–7624 (2011).
69. Karubi, W., Bhakat, S., McGillewie, L. & Soliman, M. E. Flap dynamics of plasmeprin proteases: Insight into proposed parameters and molecular dynamics. *Mol. Biosyst.* **11**, 1061–1066 (2015).
70. Kumalo, H. M., Bhakat, S. & Soliman, M. E. Investigation of flap flexibility of  $\beta$ -secretase using molecular dynamic simulations. *J. Biomol. Struct. Dyn.* **34**, 1008–1019 (2016).
71. Joseph, D., Petsko, G. A. & Karplus, M. Anatomy of a conformational change: Hinged" lid" motion of the triosephosphate isomerase loop. *Science* **249**, 1425–1428 (1990).
72. Wade, R. C., Davis, M. E., Luty, B. A., Madura, J. D. & McCammon, J. A. Gating of the active site of triose phosphate isomerase: Brownian dynamics simulations of flexible peptide loops in the enzyme. *Biophys. J.* **64**, 9–15 (1993).
73. Wang, Y., Berlow, R. B. & Loria, J. P. Role of loop–loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry* **48**, 4548–4556 (2009).
74. Zgiby, S., Plater, A., Bates, M., Thomson, G. & Berry, A. A functional role for a flexible loop containing Glu182 in the class II fructose-1, 6-bisphosphate aldolase from *Escherichia coli*. *J. Mol. Biol.* **315**, 131–140 (2002).
75. Pegan, S. D. *et al.* Active site loop dynamics of a class IIa fructose 1, 6-bisphosphate aldolase from *Mycobacterium tuberculosis*. *Biochemistry* **52**, 912–925 (2013).
76. Hakobyan, D. & Nazaryan, K. Molecular dynamics study of interaction and substrate channeling between neuron-specific enolase and B-type phosphoglycerate mutase. *Proteins Struct. Funct. Bioinf.* **78**, 1691–1704 (2010).
77. Kamerlin, S. C. L., Rucker, R. & Boresch, S. A molecular dynamics study of WPD-loop flexibility in PTP1B. *Biochem. Biophys. Res. Commun.* **356**, 1011–1016 (2007).
78. Hu, X. & Stebbins, C. E. Dynamics of the WPD loop of the *Yersinia* protein tyrosine phosphatase. *Biophys. J.* **91**, 948–956 (2006).
79. Francis, K., Stojković, V. & Kohen, A. Preservation of protein dynamics in dihydrofolate reductase evolution. *J. Biol. Chem.* **288**, 35961–35968 (2013).
80. Boehr, D. D. *et al.* A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry* **52**, 4605–4619 (2013).
81. Tuttle, L. M., Dyson, H. J. & Wright, P. E. Side chain conformational averaging in human dihydrofolate reductase. *Biochemistry* **53**, 1134–1145 (2014).
82. Jiang, Y., Li, L., Zhang, H., Feng, W. & Tan, T. Lid closure mechanism of *Yarrowia lipolytica* lipase in methanol investigated by molecular dynamics simulation. *J. Chem. Inf. Model.* **54**, 2033–2041 (2014).
83. Rehm, S., Trodler, P. & Pleiss, J. Solvent-induced lid opening in lipases: A molecular dynamics study. *Protein Sci.* **19**, 2122–2130 (2010).
84. Gunasekaran, K., Ma, B., Ramakrishnan, B., Qasba, P. K. & Nussinov, R. Interdependence of backbone flexibility, residue conservation, and enzyme function: A case study on  $\beta$ 1, 4-galactosyltransferase-I. *Biochemistry* **42**, 3674–3687 (2003).
85. Papaleo, E. *et al.* Loop 7 of E2 enzymes: An ancestral conserved functional motif involved in the E2-mediated steps of the ubiquitination cascade. *PLoS ONE* **7**, e40786 (2012).
86. Papaleo, E. *et al.* An acidic loop and cognate phosphorylation sites define a molecular switch that modulates ubiquitin charging activity in Cdc34-like enzymes. *PLoS Comput. Biol.* **7**, e1002056 (2011).
87. Das, R. *et al.* Allosteric activation of E2-RING finger-mediated ubiquitylation by a structurally defined specific E2-binding region of gp78. *Mol. Cell* **34**, 674–685 (2009).
88. Anthis, N. J. & Clore, G. M. The length of the calmodulin linker determines the extent of transient interdomain association and target affinity. *J. Am. Chem. Soc.* **135**, 9648–9651 (2013).
89. Krishnamurthy, V. M., Semetey, V., Bracher, P. J., Shen, N. & Whitesides, G. M. Dependence of effective molarity on linker length for an intramolecular protein–ligand system. *J. Am. Chem. Soc.* **129**, 1312–1320 (2007).
90. Shewmake, T. A., Solis, F. J., Gillies, R. J. & Caplan, M. R. Effects of linker length and flexibility on multivalent targeting. *Bio-macromolecules* **9**, 3057–3064 (2008).
91. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
92. Cox, S., Radzio-Andzelm, E. & Taylor, S. S. Domain movements in protein kinases. *Curr. Opin. Struct. Biol.* **4**, 893–901 (1994).
93. Dölker, N. *et al.* The SH2 domain regulates c-Abl kinase activation by a cyclin-like mechanism and remodulation of the hinge motion. *PLoS Comput. Biol.* **10**, e1003863 (2014).
94. Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci.* **106**, 139–144 (2009).
95. Biondi, R. M. *et al.* Identification of a pocket in the PDK1 kinase domain that interacts with PIF and the C-terminal residues of PKA. *EMBO J.* **19**, 979–988 (2000).
96. Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126 (2001).
97. Gerstein, M. *et al.* Domain closure in lactoferrin: Two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.* **234**, 357–372 (1993).
98. Mizutani, K., Mikami, B. & Hirose, M. Domain closure mechanism in transferrins: new viewpoints about the hinge structure and motion as deduced from high resolution crystal structures of ovotransferrin N-lobe. *J. Mol. Biol.* **309**, 937–947 (2001).
99. Baker, E. & Baker, H. Lactoferrin: Molecular structure, binding properties and dynamics of lactoferrin. *Cell. Mol. Life Sci.* **62**, 2531–2539 (2005).
100. Nezlín, R. Internal movements in immunoglobulin molecules. *Adv. Immunol.* **48**, 1–40 (1990).
101. Kim, H., Matsunaga, C., Yoshino, A., Kato, K. & Arata, Y. Dynamical structure of the hinge region of immunoglobulin G as studied by <sup>13</sup>C nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **236**, 300–309 (1994).
102. Price, N. E., Price, N. C., Kelly, S. M. & McDonnell, J. M. The key role of protein flexibility in modulating IgE interactions. *J. Biol. Chem.* **280**, 2324–2330 (2005).

## Acknowledgements

The authors would like to acknowledge the Department of Biophysics of Tarbiat Modares University.

## Author contributions

M.R. and M.T. conceived and designed the research, and M.R. and F.A.J. performed the simulations and prepared figures. M.R. Writing—original draft. M.T. supervised the project. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40818-x>.

**Correspondence** and requests for materials should be addressed to M.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed 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/>.

© The Author(s) 2023