J Mol Model (2013) 19:4865–4875
DOI 10.1007/s00894-013-1930-9

ORIGINAL PAPER

Insight into the structural stability of wild type and mutants of the tobacco etch virus protease with molecular dynamics simulations

Yu Wang · Guo-Fei Zhu · Si-Yan Ren · Yong-Guang Han · Yue Luo · Lin-Fang Du

Received: 26 March 2013 / Accepted: 18 June 2013 / Published online: 17 September 2013

Abstract The efficiency and high specificity of tobacco etch virus protease (TEVp) has made it widely used for cleavage of recombinant fusion proteins. However, TEVp suffers from a few intrinsic defects such as self-cleavage, poorly expressed in *E. coli* and less soluble. So some mutants were designed to improve it, such as S219V, T17S/N68D/I77V and L56V/S135G etc. MD simulations for the WT TEVp and its mutants were performed to explore the underlying dynamic effects of mutations on TEVp. Although the globular domains are fairly conserved, the three mutations have diverse effects on the dynamics properties of TEVp, including the elongation of β-sheet, conversion of loop to helix and the flexibility of active core. Our present study indicates that the three mutants for TEVp can change their secondary structure and tend to form more helixes and sheets to improve stability. The study also helps us to understand the effects of some mutations on TEVp, provides us insights into the change of them at the atomic level and gives a potential rational method to design an improved protein.

Keywords Molecular dynamics simulations · Mutation · Stability · Tobacco etch virus protease

Abbreviations

TEVp	Tobacco etch virus protease
MD	Molecular dynamics
WT	Wild type
SASA	Solvent accessible surface area
DCCM	Dynamic cross correlation map

Y. Wang · G.-F. Zhu · S.-Y. Ren · Y.-G. Han · Y. Luo · L.-F. Du (✉)
Key Laboratory of Bio-resources and Eco-environment of the Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610064, People's Republic of China
e-mail: dulinfang@scu.edu.cn

Introduction

Tobacco etch virus protease (TEVp) is a 27 kDa catalytic domain of the nuclear inclusion a (NIa) protein targeting the recognition specific sequence ENLYFQG/S. Owing to its stringent sequence specificity, overproduction in E. *coli*, and retaining activities in different temperature and composition, TEVp is frequently utilized for cleavage at the designed TEVp cut between the tag and target protein to produce native protein in vivo and in vitro [1–6].

As a tool for removing tags from recombinant fusion proteins, TEVp suffers from a few intrinsic defects. One is that wild type TEVp cleaves itself between Met218 and Ser219 in the TEVp protease catalytic domain to generate a truncated enzyme with greatly diminished activity [1]. Replacing serine at 219 with glutamic acid, valine or proline conferred proteases with greater resistance to auto-inactivation [7]. The S219V mutant was highly resistant to auto-inactivation with ~100-fold slower than the wild type, and improved catalytic activity in vitro compared with the wild type. The other is that TEVp is poorly expressed in E. *coli* and less soluble. Mutations of certain amino acid residues in TEVp also enhance the protein solubility. The soluble mutants containing T17S/N68D/I77V termed TEVSH was generated by directed evolution using the enhanced GFP (EGFP) as a C-terminal solubility reporter. It displayed more than fivefold increase in production of the purified protease [8]. So far, its solubility and activity in vivo have not been reported. The folding mutants containing L56V/S135G was constructed by rational design based on the predicted values of folding free energy upon mutation. This double mutant remained soluble at concentration more than 40 mg/ml and displayed the improved catalytic activity in vitro compared with WT TEVp [9].

Crystal structures of the catalytically inactive and catalytically active TEV protease in complex with a peptide substrate and product respectively, provided the first insight into

the three-dimension structure of this enzyme at atomic resolution [10]. Owing to self-cleavage of TEVp, this two crystal structures were truncated and without the C-terminal. Subsequently, by replacing Cys151 of TEVp with Ala to make it inactivation, the crystal structures consisted of the whole TEVp had been obtained to provide an insight into the outline of the C-terminal [11]. TEV protease crystal structures confirm that, as expected, this enzyme adopts the characteristic two-domain anti-parallel β-barrel folds that typify the trypsin-like serine proteases with the catalytic triad residues His46, Asp81 and Cys151 located at the interface between the domains (Fig. 1a). And the C151A crystal structure displays that the C-terminal binds with the active sites, suggesting the role of the C-terminal on the catalytic activity of TEVp [11]. Although TEVp formed dimer as was shown in the crystal structure, its native conformation was in the form of monomer [10].

As a complimentary approach, molecular dynamics (MD) simulations have become a powerful tool to provide atomic level and time-dependent information for protein structure and functions that otherwise is difficult to obtain from experiments [12–15]. The main advantage of this technique is that it can model aspects of protein structure in solution [16].

The structure of TEVp could be observed directly because of the success resolution of crystal structure, but the changes of various mutant types at the atomic scale were not understood and why mutations at these sites could influence functions was not clear. For this purpose, as mentioned above, we selected wild type of TEVp as control, and three mutants that are S219V, T17S/N68D/I77V and L56V/S135G mutants to obtain the mechanism of them with molecular dynamics simulations.

Materials and methods

Structure preparation

The initial 3D structure of TEVp was taken from the Protein Data Bank (PDB ID: 1Q31) obtained by x-ray [11]. The globular domain (Fig. 1a) contains two-domain antiparallel β-barrel folds that typify the trypsin-like serine proteases with the catalytic triad residues His46, Asp81 and Cys151 located at the interface between the domains.

Residue Cys151 of this structure was replaced by Ala, moreover, the three N-terminal residues of each monomer, residues 223-234 of chain A and residues 222-234 of chain B were not visible in electron density maps. So we selected the chain A of PDB 1Q31, built loops of the defective residues, replaced Ala151 by Cys, and generated WT (the wild type) by using Modeller program [17–20]. The structures of the studied mutants T17S/N68D/I77V, L56V/S135G and S219V were also constructed using Modeller program by replacing the corresponding residues based on WT.

Molecular dynamics (MD) simulations

Molecular dynamics simulations were performed with NAMD simulation package [21] and the CHARMM27 force field [22]. Each form of TEVp structure was solvated in a cubic box by using TIP3P [23] water molecular with at least 10 Å between protein and box-boundary and simulated at neutral PH. In particular, here Cl^- ions were added to keep the system neutral. First, we minimized each system energetically to remove bad contact with the protein atoms positions constrained and then relaxed them by using conjugate

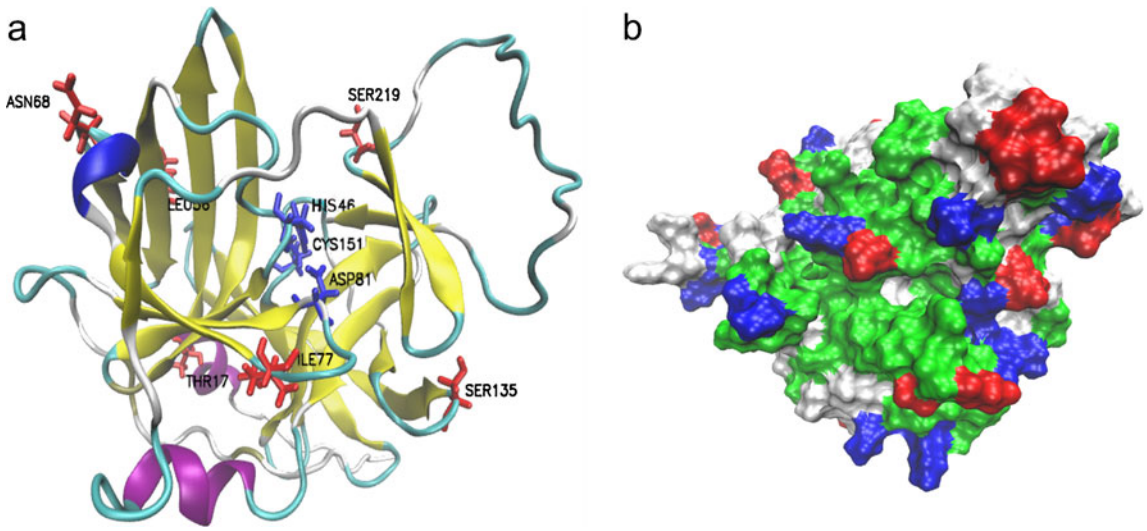

Fig. 1 Structures of the WT TEVp. **a** The backbone is shown in new-cartoon pattern and α-helix, β-sheet, turn and coil regions are shown in *pink*, *yellow*, *cyan* and *white*, respectively. The catalytic triad (His 46, Asp81, Cys151) is shown in licorice and colored by *blue*. The mutant residues sites is also shown in licorice and colored by *red*. **b** Surface electrostatic potential representation of the WT type. Positive, negative, neutral polar and non-polar residues are colored *blue*, *red*, *green* and *gray*, respectively

gradient method. Subsequently each system was subjected to 60 ps and then heated from 0 to 300 K followed by an equilibration run. Each system was then simulated for 30 ns and the trajectories were saved at 10 ps intervals for further analysis. All simulations employed periodic boundary and multiple time stepping [24] wherein local interactions were calculated every 2 fs time step and full electrostatic evaluations were performed every two time steps. The non-bonded cutoff distance was set to 12 Å and the particle mesh Ewald (PME) [25] method was used to calculate long-range electrostatics interactions. All covalent bonds involving hydrogen were kept rigid and the others using STTLE [26] or RATTLE [27] algorithms. All equilibration and subsequent MD stages were carried out without any restrains in the NPT ensemble at 300 K by using the Langevin thermostat control [28] and a target 1 atm by using Langevin piston pressure control [29].

Analysis methods

The trajectories were analyzed using VMD program [30]. The root-mean-square deviation (RMSD) value provides a measure of the deviation of the aggregate from the initial structure and is a useful tool to quantify the conformational changes between the mutants and WT. In this study, the RMSDs of backbone without H atoms were calculated for the whole domain and the partial domain (residues 10-242) from the structure after equilibration. To evaluate the compactness of the studied proteins, the size of protein is roughly estimated by radius of gyration (Rg), which is defined as the mass-weighted positional mean of the distances of atoms from the center of mass. The dictionary of secondary structure of proteins (STRIDE) program developed by Frishman and Argos [31] was used to assign secondary structures. STRIDE implements a knowledge-based algorithm that makes combined use of hydrogen bond energy and statistically derived backbone torsional angle information and is optimized to return resulting assignments in maximal agreement with crystallographers' designations. To evaluate the local structure fluctuation of helixes and β-sheets, the hydrogen bond populations were calculated. A hydrogen bond was defined by a hydrogen–acceptor distance ≦3.5 Å and a donor–hydrogen–acceptor angle less than 35°. Our aim to analyze the hydrogen bonds was to explore the reason of the change of secondary structure, and only backbone hydrogen bonds influenced the secondary structure [32], thus, backbone hydrogen bonds were mainly considered here.

Solvent-accessible surface area (SASA) is the area of the surface traced by the center of a probe sphere, whose radius is the nominal radius of the solvent, as it rolls over the van der Waals surface of the molecule [33]. SASA was calculated for the entire simulation period using SurfVol, a Plugin for Visual Molecular Dynamics (VMD) software [30] for measuring the surface area of proteins (http://www.compbiochem.org/Software/SurfVol/Home.html). For the four cases studied in this work, SASA was calculated using a probe radius of 1.4 Å. To evaluate the stability of the globular domain of TEVp, we mainly calculated the SASA value of a core consisting of hydrophobic residues [34, 35].

Simulation trajectories were used to compute various properties of the protein, including the correlation of motions among its residues in various regions. The correlation analysis was obtained by examining the dynamic cross correlation map (DCCM) of the Cα atoms. The matrix element C_{ij} in DCCM read as:

$$C_{ij} = \frac{< \Delta\vec{r}_i(t)\Delta\vec{r}_j(t) >}{\left(< \Delta\vec{r}_i(t)^2 >< \Delta\vec{r}_j(t)^2 >\right)^{1/2}}$$

where $< >$ denotes an MD-averaged quantity $\Delta\vec{r}_i(t) = r_i(t) - < \vec{r}_i >$ and the displacement from the average MD position $< \vec{r}_i >$ of atom i during a generic MD step. C_{ij} varies from −1.0 for completely anti-correlated motions to + 1 for completely correlated motions. A value close to +1 reflects a high correlation between the motions of a pair of Cα atoms. The largest values are obviously found for Cα atoms belonging to residues i and i±a with a=0, 1, 2 (diagonal elements in the map) [36]. Here we selected the last 10 ns trajectories to calculated DCCM.

Results and discussion

Overall dynamics behavior of the WT and mutated TEVp

As well known, conformation and flexibility of the structure can be altered by mutations. Therefore Cα root-mean-square fluctuations (RMSFs) around the average structure for the relevant time interval were calculated since they can give an indication of backbone mobility. In Fig. 2c, Cα RMSFs fluctuated in a similar manner for all systems. Cα RMSFs of the residues around the active sites (HIS46, ASP81 and CYS151) were relatively small, indicating that the catalytic center was stable. The N-terminal, residues around Ser120 and the C-terminal, were more flexible, and had relatively larger RMSF values than the other domains. It was also observed that the RMSF values of L56V/S135G were obviously larger than others, and those of S219V were relatively smaller, especially in the domain of residues 81 to 120. The C-terminal bound with the active sites but there was no clear evidence indicating the connection between the N-terminal and the function of TEVp.

Backbone RMSDs were calculated for the globular domain and the partial domain (residues 10 to 242) for all simulations because the N-terminal was very flexible as described previously. In Fig. 2a, the backbone RMSD values

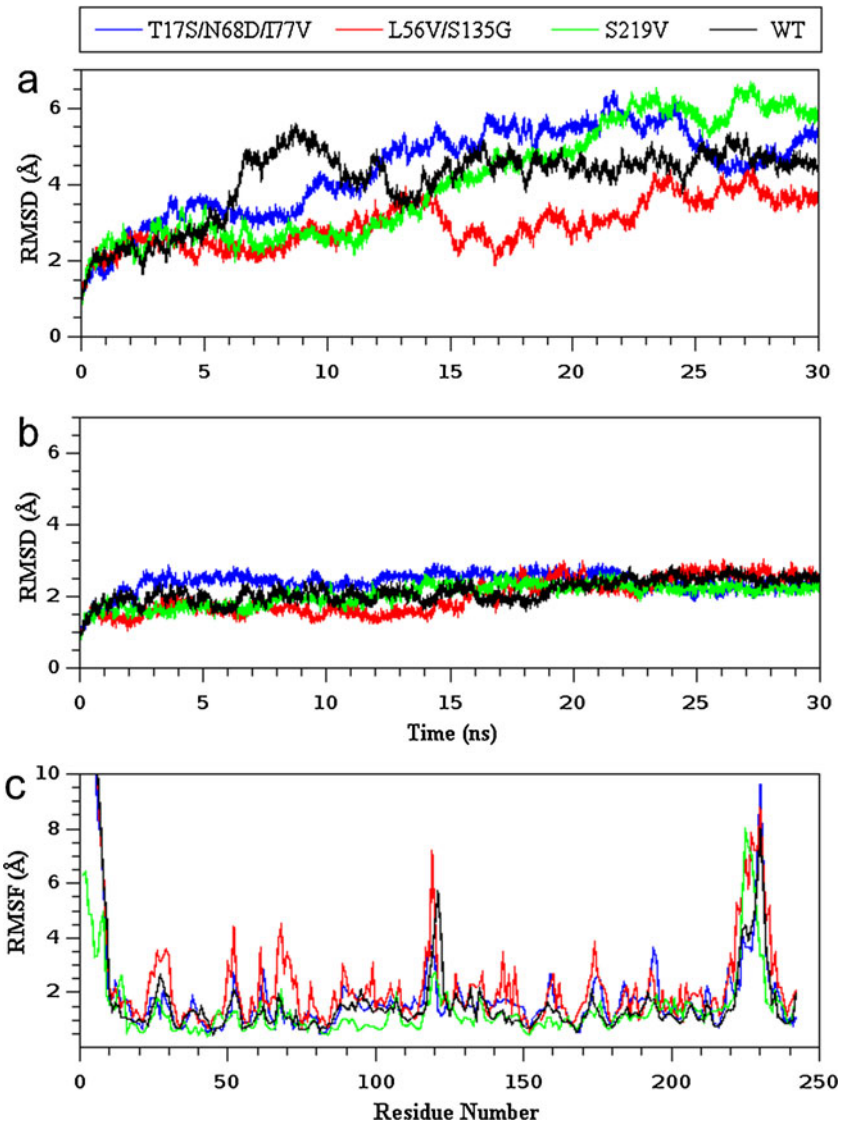

Fig. 2 Structure characteristics of TEVp calculated by MD simulations. **a** Backbone RMSD from the starting structure of the globular domain of all systems. **b** Backbone RMSD from the starting structure of the partial domain (residues without the N-terminal nine amino acids) of all systems. **c** Cα RMSF as a function of residue number

of the globular domain were generally high and larger than 3 Å during the last 5 ns, and seemed unstable. This phenomenon might be caused by the restless N-terminal and thus we calculated the backbone RMSD values of the partial domain without N-terminal 9 residues. After running for 20 ns, every system tended to be stable; meanwhile, the backbone RMSD values of partial domains fluctuated in a similar manner for all systems and were smaller than 3 Å during the last 5 ns (Fig. 2b).

Table 1 showed the mean values of the radius of gyration (Rg) of WT and mutants of TEVp during the last 5 ns. Rg values essentially provide an insight about the size of the protein. The slightly lower Rg values for the three mutants implied a possible compact of TEVp structure due to mutations. Correlations of the motion among various regions in TEVp can be obtained by examining the dynamic cross correlation map (DCCM) of the Cα atoms. The DCCMs of WT TEVp and its three mutants were displayed in Fig. 3. Mutants showed a general decrease in residue-residue correlations compared to WT. Moreover, mutations were seen to affect the correlation of residues that were sequentially and spatially apart. In the system of L56V/S135G, the nearby domains of residues 56 and 135 respectively, were obviously anti-correlation due to mutations at residues 56 and 135. The decreased residue-residue correlations indicated that the changes arising due to mutations around the mutated sites were coupled with other motions in the structure.

Stability of the hydrophobic core

To evaluate the stability of the globular domain of TEVp generally, a core of interacting hydrophobic side chains was defined in TEVp, which contained residues 4-5, 8, 13-14, 18, 21, 32, 35, 37, 39-42, 47-48, 55-57, 60, 63-64, 66, 76-77, 82-

Table 1 Calculated mean values for various properties, their standard deviations, and the differences between the mean values of wild type and mutated TEVp

System	Mean	Standard deviation	Δ(Mut-WT)
Radius of gyration of TEVp (Å)			
WT	19.39	0.23	–
T17S/N68D/I77V	18.53	0.27	−0.86
L56V/S135G	17.74	0.16	−1.65
S219V	18.39	0.21	−1.00
SASA of TEVp ($Å^2$)			
WT	18089.45	134.50	–
T17S/N68D/I77V	17625.98	132.76	−463.47
L56V/S135G	15647.53	125.01	−2441.92
S219V	17426.54	132.01	−662.91
Hydrophobic SASA of TEVp ($Å^2$)			
WT	3904.18	62.48	–
T17S/N68D/I77V	3752.25	61.26	−151.93
L56V/S135G	3627.68	60.23	−276.50
S219V	3759.65	61.32	−144.53
Hydrogen bonds in backbone			
WT	47.51	4.40	–
T17S/N68D/I77V	51.14	4.62	3.63
L56V/S135G	51.20	4.42	3.69
S219V	49.13	4.38	1.62

85, 87-88, 91-95, 100, 103, 111-112, 114, 116, 121, 124-125, 132-133, 135, 138-140, 143-144, 154-156, 162-164, 166, 169, 172, 178, 182-183, 186-187, 189-190, 195, 198-199, 202, 204, 206, 209-211, 215-219, 221, 224-225, 227-228, 231, 234-235, 238-239. To begin with, we monitored the solvent accessible surface area (SASA) of the globular domain averaged during the final 5 ns MD data (Table 1). As can be seen, the WT experienced more solvent exposure than the three mutants. This observation was in accordance with the decreased Rg values of the three mutants. Secondly, the SASA of the hydrophobic core were monitored during the last 5 ns in order to assess its compactness (Table 1). It was found that the changes of the four cases were in accordance with the total SASA values. At last, as the large changes in SASA could be caused by some residues that exposed hydrophobic surfaces, we also measured SASA of the related residues averaged over the last 5 ns simulations (Fig. 4) to get further detailed information. As Fig. 4 showed, the hydrophobic surface differences mainly lay in the residues Pro13, Leu60, Pro92, Pro95, Phe116, Met121, Met124, Val25, Phe132, Mut135, Phe162, Ala195, Trp198, Pro224, Phe225 and Val228, especially in the domains of residues 114-125, 132-135 and 224-235. These sites were mostly located in β-hairpin and C-terminal loops. As is well known (Fig. 1a), the residue Leu60 was located around the mutant site Leu56, and β-hairpin was the neighbor of the mutant site Ser135. Furthermore, the residue Leu60 was also located near the β-hairpin structure at the dimensional structure. It is not necessary to consider N-terminal and C-terminal loops because they exposed in the water absolutely.

On the one hand, the β-hairpin structure had higher flexibility. Thus the SASA values of the related hydrophobic residues located in these domains were variable. On the other hand, mutations introduced different side chains, and their different steric and electrostatic effects influenced the conformations and properties of the nearby residues. As a result, their SASA values in four proteins were diverse. The TEVp has a large number of charged residues on the protein surface (Fig. 1b), for example, the WT TEVp has 28 positive and 24 negative charged residues on the surface. Hence electrostatic interactions are very important for the stability of TEVp. T17S/N68D/I77V variant introduced an extra negative charged amino acid, while L56V/S135G and S219V mutants shared the same whole protein net charge with WT. WT had larger SASA values of the hydrophobic core (Table 1) and hydrophobic residues near residues 56 and 135 (Fig. 4), suggesting the mutants were more stable since the hydrophobic residues which if exposed in water might decrease stability of proteins and even cause denaturation. This phenomenon also declared that the fixed point mutation not only affected the adjacent conformational changes of TEVp but also the distant.

Secondary structure stability

The analysis of secondary structures is very useful for further understanding of how the three mutants influence the

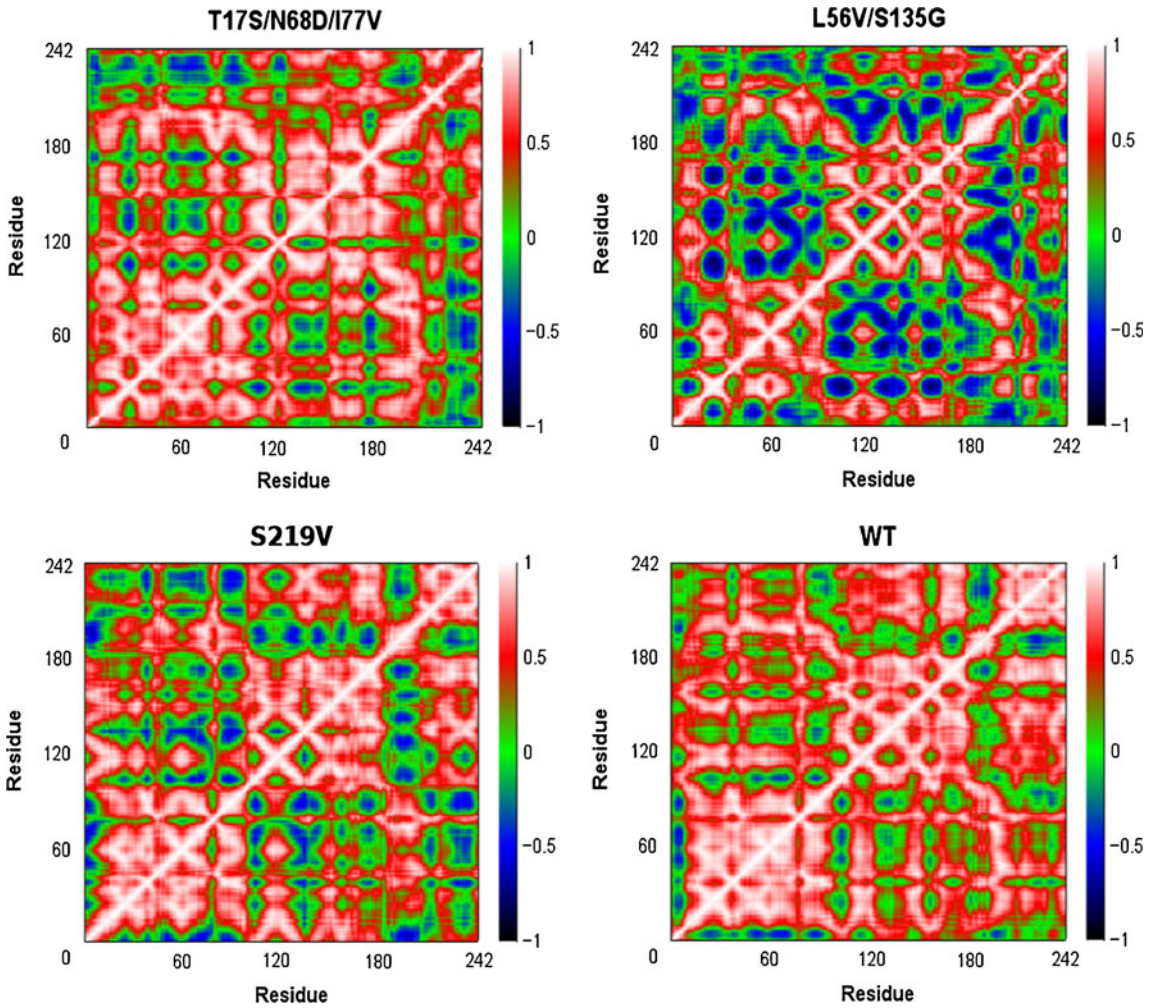

Fig. 3 Correlations of the motions of various regions in TEVp. Two dimensional cross-correlation maps of the TEVp of WT and mutated TEVp. *Red patches* indicate the positively correlated motions, whereas *blue patches* indicate anti-correlated motion. The maps have been calculated for Cα atoms from the final 10 ns MD data

function of TEVp. As we know, the mutation has effects on the three dimensional conformation of proteins. From the analysis of the secondary structure contents of the whole simulation time shown in Table 2, we found the helix and sheet contents of the three mutants had increased. Furthermore, T17S/N68D/I77V mutant had the highest percent of helix especially the α-helix while L56V/S135G mutant had the highest percent of anti β-sheet. As Table 2 showed, the average α-helical contents of T17S/N68D/I77V, L56V/S135G and S219V were 5.40 %, 5.37 % and 5.28 % respectively and all higher than that of WT (5.14 %). The average β-sheet contents of T17S/N68D/I77V, L56V/S135G and S219V were 40.44 %, 40.70 % and 37.64 % respectively and all higher than that of WT (36.94 %). The higher α-helix and β-sheet contents

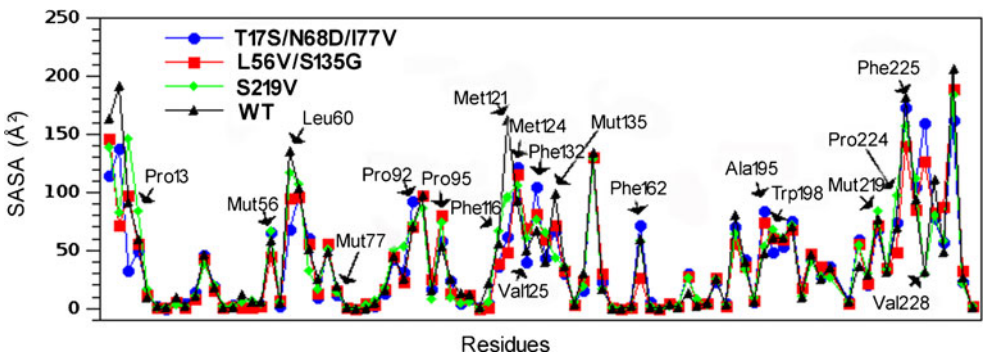

Fig. 4 The solvent-accessible surface area (SASA) of the TEVp. The SASA of the hydrophobic core averaged during the last 5 ns simulations data for WT and mutated TEVp

Table 2 Average secondary structure contents and standard errors of WT and the mutants

System	Helix (%)			β-sheet (%)	β-bridge (%)	Coil (%)	Turn (%)
	α-helix	3_{10}-helix	Sum				
T17S/N68D/I77V	5.40±0.45	1.06±0.93	6.46±1.07	40.44±1.34	1.54±0.30	23.70±2.03	27.86±2.21
L56V/S135G	5.37±0.37	0.79±0.90	6.16±0.94	40.70±1.24	1.56±0.28	21.85±1.89	29.73±2.17
S219V	5.28±1.31	1.08±1.02	6.36±1.49	37.64±1.29	1.66±0.30	24.32±2.07	30.02±2.52
WT	5.14±0.82	0.89±0.98	6.03±1.12	36.94±1.57	1.61±0.34	25.34±2.22	30.08±2.67

indicated these mutants were more stable than WT. The midpoint of denaturation, which is a measure of the thermostability of the structure, was found to be 52.1 °C for WT and 53.7 °C for L56V/S135G mutant [9]. This experiment indicated the L56V/S135G mutant was more stable at thermodynamics. According to Kapust RB and coworkers, the S219V mutant had a marginal effect on stability. The experiment data obtained from the two mutants kept with our study. Although there was no definite experiment data about the stability of T17S/N68D/I77V, our results declared this mutant had improved its stability at thermodynamics.

We calculated the number of hydrogen bonds in the backbone of each system along with the simulation time, because the forming and maintaining of the secondary structures of protein needs the formation of hydrogen bond, and the unfold protein is accompanied by the damage of hydrogen bonds. As Table 1 showed, we found the average number of backbone hydrogen bonds of WT, T17S/N68D/I77V, L56V/S135G and

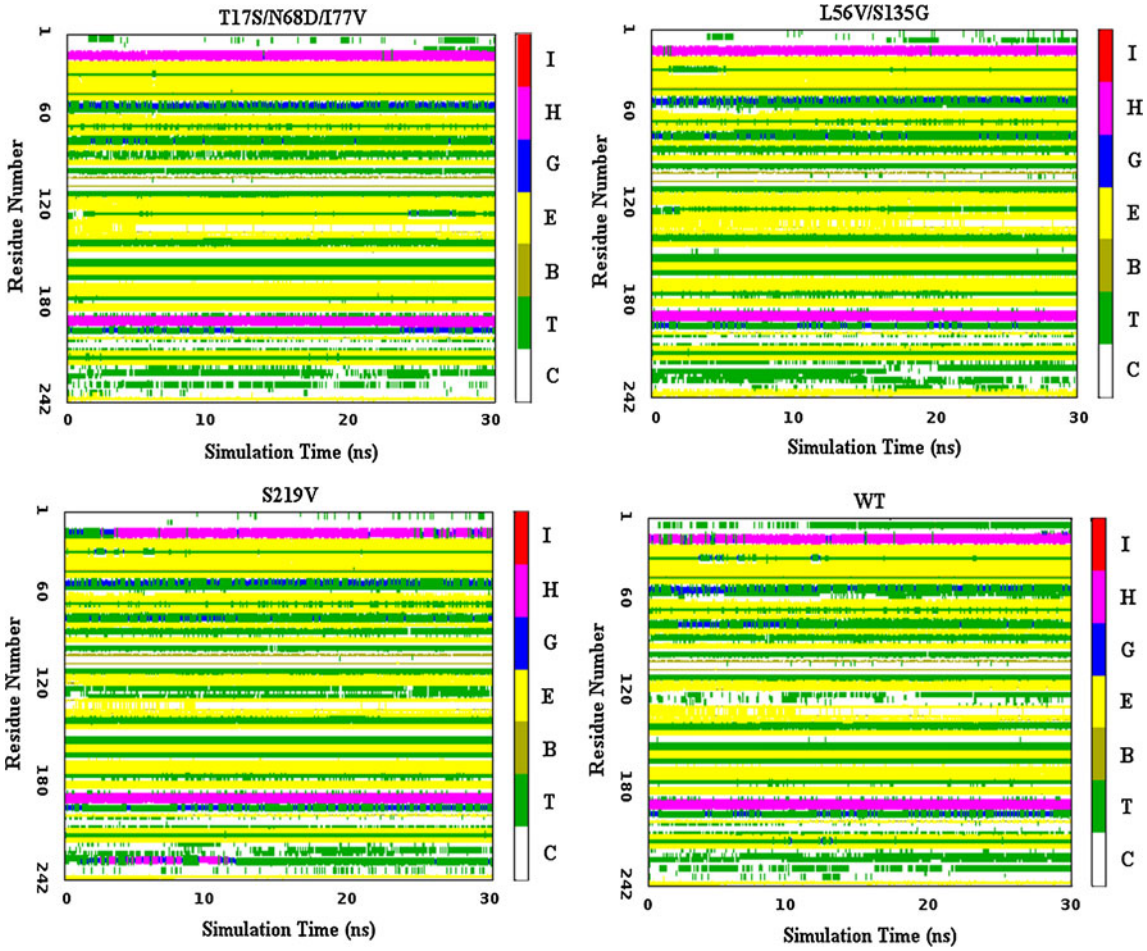

Fig. 5 Graphical regulation of secondary structure analysis for the WT and the three mutants based on the STRIDE algorithm. Here, we labeled a coil by "C", a turn by "T", a β-sheet by "E", a β-bridge by "B", a 310-helix by "G", a α-helix by "H" and a pi-helix by "I"

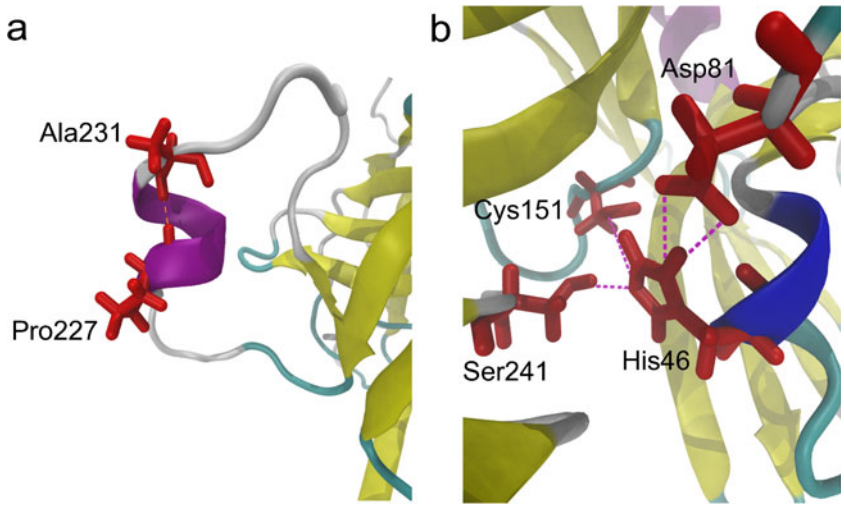

Fig. 6 Structures of the TEVp. **a** The α-helix is formed at the loop domain of the C-terminal in the S219V mutant. **b** The structure of active core. Residue 46, 81 151 and 241 are drown by licorice, colored by *red* and the others are drawn by new cartoon, shows the hydrogen bond between catalytic triad residues and the C-terminal bound with active site

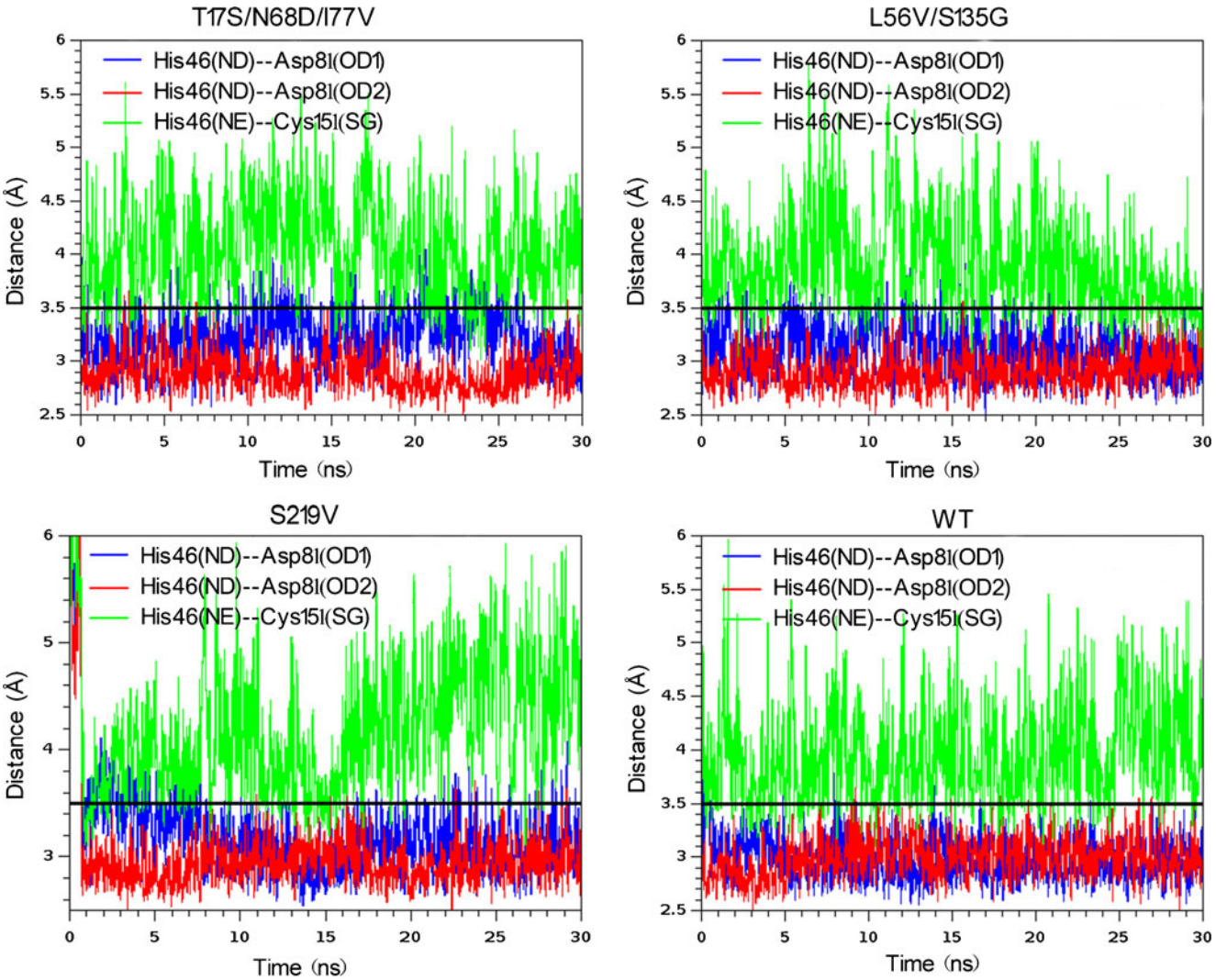

Fig. 7 Distance between catalytic triad residues. Distance between hydrogen bond acceptor-donor pairs between catalytic triad residues His46, Asp81 and Cys151, and the *black line* represent the cur-off distance of hydrogen bond as 3.5 Å

Table 3 The occupancy of hydrogen bonds forming at the active sites

Donor	Acceptor	Occupancy (%)			
		T17S/N68D/I77V	L56V/S135G	S219V	WT
His46:N	Asp81:OD1	79.96	69.51	59.79	63.78
His46:ND	Asp81:CG	65.65	70.31	54.33	74.49
His46:ND	Asp81:OD1	25.57	41.94	47.14	54.19
His46:ND	Asp81:OD2	47.80	23.97	20.77	18.77
Cys151:SG	His46:NE	4.79	6.13	2.4	4.39
Cys151:SG	Ser241:OG	2.26	5.19	9.05	2.53
Ser241:OG	His46:NE	50.73	72.57	58.19	74.83
Ser241:OG	Asp81:OD2	0.13	–	0.8	0.4

S219V were 47.51, 51.14, 51.20 and 49.13 respectively. The number of backbone hydrogen bonds of the three mutants was all higher than that of the WT type. This phenomenon was in concord with the changes of helix and sheet in the three mutants, which was similar with experiments by Kapust [7] and Cabrita LD [9] as described previously. This phenomenon also suggested the three mutants were more stable than the WT type.

As described previously, the C-terminal of the TEVp could be cut off by itself and generated a truncated enzyme with greatly diminished activity [1], and the crystal structure of TEVp displayed the C-terminal bound with the active sites [11], suggesting the relativity between C-terminal and its activity. However, it was a pity that the loop located in the C-terminal could not directly be observed because of its possible higher flexibility. As Fig. 5 showed, we compared the transformation of the secondary structures of all four systems along with the simulation time. The phenomenon that the loop domain of the C-terminal in the S219V mutant was folded to form 3_{10}-helix or α-helix could be observed (Fig. 6a). A similar phenomenon did not appear in the other systems. Thus this phenomenon might be caused by the nearby residue Ser219 replaced by Val. As described previously, the S219V mutant not only prevented TEVp from cleaving itself, but also could increase its activity obviously [7]. This property might be caused by the appearance of the helix in the C-terminal of the S219V mutant. As Kapust RB reported, the Kcat/Km of S219V was 4.36 $mM^{-1}S^{-1}$, and obviously higher than that of WT, which was 2.62 $mM^{-1}S^{-1}$ [7]. This experiment data was in accordance with that our study.

The hairpin structure of residues 114 to 124, can form anti-sheet with the same domain of another subunit at the crystallize process [10]. This domain and the nearby sheet of residues 132-135 were very flexible, and had higher RMSF values. In addition, the two domains were located on the surface of protein, and there were many hydrophobic residues with larger variable SASA values located in these domains as described previously. As Fig. 5 showed, the sheet structure domains of residues 124 to 135 in the T17/N68D/I77V and L56V/S135G mutants increased obviously, while those in the S219V mutant and WT had no significant changes. This phenomenon suggested that the mutations of the nearby residues Gln68, Ile77, Leu56 and Ser135 of these domains, not only influenced the SASA values of these domains, but also changed their secondary structure to form more β-sheets. As a result, their stabilities increased.

The RMSF values of catalytic triad residues His46, Asp81 and Cys151 were low but those of the nearby loop domains of residues 44 to 54 and residues 143 to 152, which were usually regarded as the areas bound with substrate, were flexible as described previously (Fig. 2c). The results showed that there were unstable 3_{10}-helix appearing at the loop domain of residues 44 to 54 which was the neighbor of His46 (Fig. 6b) and the frequency of this phenomenon observed was higher in the three mutants, especially in the T17/N68D/I77V mutant than that in the WT type during the simulation time (Fig. 5). The secondary structures around residue Asp81 were changeful, and there was 3_{10}-helix appearing sometimes but the secondary structures around residue Cys151 were quite stable and mostly had no changes.

We calculated the changes of distances between partial atom pairs along with the simulation time (Fig. 7). The His46: ND could form a hydrogen bond with Asp81:OD1 or Asp81:OD2 as donor, and His46: NE could also form a

Table 4 Average distance and stand errors of active core of WT and the three mutants (Å)

	T17S/N68D/I77V	L56V/S135G	S219V	WT
His46(ND):Asp81(OD1)	3.22±0.08	3.10±0.07	3.23±0.27	2.98±0.04
His46(ND):Asp81(OD2)	2.90±0.03	2.91±0.03	3.00±0.17	2.99±0.04
His46(ND):Cys151(SG)	4.06±0.21	3.90±0.20	4.20±0.40	3.96±0.21

hydrogen bond with Cys151:SG as acceptor. In TEV crystal structure and its high similar cysteine proteases (hepatits A virus and rhinovirus), the distance between SG atom in catalytic cysteine and NE atom in histidine is in the range of 4.0 Å. The distance between the O atom in the third member of catalytic triad residues and ND atom in the histidine is 2.95–3.08 Å [10]. It was consistent with these that our results indicated the distance between His46:NE and Cys151:SG was longer than 3.5 Å mainly (Fig. 7). Therefore, the hydrogen bond lifetime between His46:NE and Cys151:SG is very short (Table 3). Furthermore, the distances of these two atoms were the most variable in the S219V mutant. The distances between His46:ND and Asp81:OD1 or Asp81:OD2 were mainly shorter than 3.5 Å, suggesting that the hydrogen bond between His46 and Asp81 could be formed stably. As Table 4 showed, the catalytic triad residues and the C-terminal could form hydrogen bond networks at the active sites. His46, Asp81 and Ser241 were connected firmly by hydrogen bonds with high occupancy. In Fig. 7 and Table 4, we find the average distances of these three atomic pairs of S219V mutant are longer than the others, indicating the substrate-bind package of S219V was bigger than that of WT, and the stand errors were also larger than WT, indicating the active core was more flexible in the S219V mutant. Flexibility, one of the key modulators of enzyme specificity, appears to play an important role in the entry and accommodation of substrates in the active site, and in the release of the respective products of reactions catalyzed by enzyme, as a result of possibly influencing the activity [37]. So the increase of flexibility of active core might be one of the reasons why S219V could improve catalytic activity in vitro compared with the wild type.

Conclusions

By performing all-atom MD simulations, we investigated the atomic-level structural variations in WT TEVp and three improved mutants to understand the molecular origin for their structural transition. Although the globular domain was fairly conserved between WT and mutants, the changes in their local structures were observed upon mutations. Our present study illustrated that the three mutants for TEVp might undergo different tactics to improve quality. As for S219V mutant, its active core was more flexible and its C-terminal loop could be folded to form a helix. So the S219V mutant avoided cleaving itself and had higher activity and stability than the WT type. This result was in accordance with the experiment data reported by Kapust [7]. The Kcat/Km of S219V mutant was obviously higher than that of WT, hence S219V displayed the improved catalytic activity compared with WT. The ΔG_{H2O} of S219V was not as low as that of WT, indicating that the stability of S219V had slightly improved compared to WT. As for L56V/S135G and T17S/N68D/I77V mutants, the stronger polarity residues introduced in these mutants, made the microenvironment near mutant sites change, as a result, the β-sheet extended and the SASA values of hydrophobic residues located here decreased. Thus their stabilities had improved. The midpoint of denaturation, which is a measure for the thermostability of the structure, was found to be 52.1 °C for WT and 53.7 °C for L56V/S135G mutant. So the double variant showed an increase in the thermal stability compared with WT. This experiment data was in accordance with our results. There was no explicit experiment data about the thermal stability of T17S/N68D/I77V, but we inferred its stability could improve from our results. Overall, this paper showed that the realistic atomistic simulations can give insights into the effects of mutations on TEVp dynamics and stability, which in turn adds to our understanding of how these mutations cause the changes of property of TEVp.

Acknowledgments This work was financially supported by the National Key Technology R&D program of China (2009BAK61B04, 2006BAF07B01) and Science & Technology Foundation of Sichuan Province (2011JTD0026). NAMD and VMD were developed by the Theoretical Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign, USA.

References

1. Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG (1995) Expression and purification of a recombinant tobacco etch virus NIa proteinase: biochemical analyses of the full-length and a naturally occur- ring truncated proteinase form. Virology 210:194–201
2. Kapust RB, Waugh DS (2000) Controlled intracellular processing of fusion proteins by TEV protease. Protein Expr Purif 19:312–318
3. Shih YP, Wu HC, Hu SM, Wang TF, Wang AHJ (2005) Self-cleavage of fusion protein in vivo using TEV protease to yield native protein. Protein Sci 14:936–941
4. Chen X, Pham E, Truong K (2010) TEV protease-facilitated stoichiometric delivery of multiple genes using a single expression vector. Protein Sci 19:2379–2388
5. Waugh DS (2011) An overview of enzymatic reagents for the removal of affinity tags. Protein Expr Purif 80:283–293
6. Wei L, Cai X, Qi Z, Rong L, Cheng B, Fan J (2012) In vivo and in vitro characterization of TEV protease mutants. Protein Expr Purif 83:157–163
7. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14:993–1000
8. Van den Berg S, Löfdahl PA, Härd T, Berglund H (2006) Improved solubility of TEV protease by directed evolution. J Biotechnol 121:291–298
9. Cabrita LD, Gilis D, Robertson AL, Dehouck Y, Rooman M, Bottomley SP (2007) Enhancing the stability and solubility of TEV protease using in silico design. Protein Sci 16:2360–2367
10. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK III, Kapust RB, Li M, Wlodawer A, Waugh DS (2002) Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 277:50564–50572

11. Nunn CM, Jeeves M, Cliff MJ, Urquhart GT, George RR, Chao LH, Tscuchia Y, Djordjevic S (2005) Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. J Mol Biol 350:145–155
12. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9:646–652
13. Sotomayor M, Schulten K (2007) Single-molecule experiments in vitro and in silico. Science 316:1144–1148
14. Lin YW (2011) Structural insights into a low-spin myoglobin variant with bis-histidine coordination from molecular modeling. Proteins 79:679–684
15. Lin YW, Wu YM, Liao LF, Nie CM (2012) Molecular modeling of cytochrome b5 with a single cytochrome c-like thioether linkage. J Mol Model 18:1553–1560
16. Amorim HLN, Netz PA, Guimarães JA (2010) Thrombin allosteric modulation revisited: a molecular dynamics study. J Mol Model 16:725–735
17. Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815
18. Martı-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Biophys Biomol Struct Rev 29:291–325
19. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9:1753–1773
20. Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. Methods Mol Biol 426:145–159
21. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
22. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
24. Batcho PF, Case DA, Schlick T (2001) Optimized particle-mesh ewald/multiple-time step integration for molecular dynamics simulations. J Chem Phys 115:4003–4018
25. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593
26. Miyamoto S, Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962
27. Andersen HC (1983) RATTLE: a velocity version of the SHAKE algorithm for molecular dynamics calculations. J Comput Phys 52:24–34
28. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
29. Martyna GJ, Tobias DJ, Klein ML (1994) Constant-pressure molecular dynamics algorithms. J Chem Phys 101:4177–4189
30. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
31. Frishman D, Argos P (1995) Knowledge-based secondary structure assignment. Proteins 23:566–579
32. Chong SH, Lee C, Kang G, ParK M, Ham S (2011) Structural and thermodynamic investigations on the aggregation and folding of acylphosphatase by molecular dynamics simulations and solvation free energy analysis. J Am Chem Soc 133:7075–7083
33. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379–400
34. Guo J, Ning L, Ren H, Liu H, Yao X (2012) Influence of the pathogenic mutations T188K/R/A on the structural stability and misfolding of human prion protein: insight from the molecular dynamics simulations. Biochim Biophys Acta 1820:116–123
35. Padhi AK, Kumar H, Vasaikar SV, Jayaram B, Gomes J (2012) Mechanisms of loss of functions of human angiogenin variants implicated in amyotrophic lateral sclerosis. PLoS One 7(2):e32479
36. Natarajan K, Senapati S (2012) Understanding the basis of drug resistance of the mutants of αβ-Tubulin dimer via molecular dynamics simulations. PLoS One 7(8):e42351
37. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. Science 324:203–207