Improving thermostability of papain through structure-based protein engineering

Debi Choudhury, Sampa Biswas, Sumana Roy and J.K.Dattagupta

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Kolkata 700064, India

1To whom correspondence should be addressed. E-mail: jibank.dattagupta@saha.ac.in

Received September 23, 2009; revised February 11, 2010; accepted February 22, 2010

Edited by: Stephen Mayo

Papain is a plant cysteine protease of industrial importance having a two-domain structure with its catalytic cleft located at the domain interface. A structure-based rational design approach has been used to improve the thermostability of papain, without perturbing its enzymatic activity, by introducing three mutations at its interdomain region. A thermostable homologue in papain family, Ervatamin C, has been used as a template for this purpose. A single (K174R), a double (K174RV32S) and a triple (K174RY32SV36S) mutant of papain have been generated, of which the triple mutant shows maximum thermostability with the half-life ($t_{1/2}$) extended by 94 min at 60°C and 45 min at 65°C compared to the wild type (WT). The temperature of maximum enzymatic activity ($T_{max}$) and 50% maximal activity ($T_{50}$) for the triple mutant increased by 15 and 4°C, respectively. Moreover, the triple mutant exhibits a faster inactivation rate beyond $T_{max}$ which may be a desirable feature for an industrial enzyme. The values of $t_{1/2}$ and $T_{max}$ for the double mutant lie between those of the WT and the triple mutant. The single mutant however turns out to be unstable for biochemical characterization. These results have been substantiated by molecular modeling studies which also indicate highest stability for the triple mutant based on higher number of interdomain H-bonds/salt-bridges, less interdomain flexibility and lower stability free-energy compared to the WT. In silico studies also explain the unstable behavior of the single mutant.

Keywords: interdomain H-bonds/molecular modeling/papain/structure-based protein engineering/thermostability

Introduction

Enzymes are catalysts that often provide unsurpassed fidelity and selectivity under ambient and near ambient conditions of pH, temperature and solvent composition and therefore, many of them are widely used in the industry. However, despite their favorable qualities, insufficient protein stability in reaction media has hampered their implementation in various industrial processes. Higher stability is generally considered an economic advantage because of reduced enzyme turnover. In addition, stable enzymes permit the use of high process temperatures, which usually have beneficial effects on reaction rates, reactant solubility and the risk of microbial contamination. A number of approaches have been employed to develop stable proteins, either by isolating and expressing novel proteins with exceptional stability (Haki and Rakshit, 2003) or engineering the protein environment through the addition of co-solutes (Davis-Searles et al., 2001; Broering and Bommarius, 2005), chemical modifications and immobilization of the protein in solid or gel-like matrices (Chan et al., 2007). Alternatively, protein engineering methods have been used to modify the protein scaffold to improve protein stability (Burley and Petsko, 1985; Lim and Sauer, 1989; Serrano and Fersht, 1989; Shirley et al., 1992; Fujinaga et al., 1993; Hardy et al., 1994; Kirino et al., 1994; Yip et al., 1998). Protein engineering approaches can be broadly subdivided into three main methods: (i) random mutagenesis followed by selection (Giver et al., 1998; Tamakoshi et al., 2001), (ii) rational design based on three-dimensional (3D) structure of the protein of interest (Mooers et al., 2003) and (iii) consensus approaches using statistics and sequence databases (Steipe et al., 1994). Each of these approaches has been successful in the design of thermally stabilized mutant proteins. However, all the abovementioned methods of stabilizing proteins have demonstrated that there is a trade-off between the rigidity required for stability and the flexibility necessary for activity in most enzymes. Hence, it is most important that engineering industrial enzymes with enhanced stability should also aim at retaining its enzymatic activity so as to make the enzyme industrially viable.

Papain, a cysteine protease of plant origin with strong proteolytic activity against a broad range of natural and synthetic substrates, is a widely studied enzyme because of its extensive use in various industrial processes. Thermal denaturation studies of papain by both differential scanning calorimetry and spectrophotometric methods (Sumner et al., 1993) indicated that papain starts losing its activity from 55°C. Thus it may be worthwhile to increase the stability of this important enzyme without compromising its proteolytic efficiency for its various commercial applications. There have been attempts at stabilizing the papain molecule either by immobilizing the protein onto various supports (Axen and Embark, 1971; Ganapathy et al., 2001; Chen et al., 2009) or by chemical modification of the protein (Rajalakshmi and Sundaram, 1995; Sangetha and Abraham, 2006). In the present study, we have explored the possibility of enhancing the stability of papain, by structure-based rational design essentially by two approaches: (i) experimentally incorporating the mutations (structure-based) by site-directed mutagenesis (SDM) and (ii) in silico methods through molecular dynamics (MD) simulations.

Three-dimensional structures of three isozymes ervatamia-A, -B and -C, isolated from the latex of the tropical plant *Ervatamia coronaria*, were determined in our laboratory and from their structural and functional analysis it
was found that although the ervatamins belong to the papain-family, they exhibit higher thermal stability than papain (Biswa et al., 2003; Guha Thakurta et al., 2004; Ghosh et al., 2007, 2008). Moreover, three amino acids in the inter-domain region at positions 32, 36 and 174 (papain numbering) of ervatamins were identified which play a vital role in the gradual increase of their thermostabilities by forming additional interdomain H-bond network. Since, among the three ervatamins, Ervatamin C (Erv C) was found to be most thermostable, it was used as a template to improve thermostability of papain. The targets for mutagenesis in papain therefore were the three residues Val32, Gly36 and Lys174 in the interdomain region to be mutated to Ser, Ser and Arg, respectively, as found in the corresponding positions in Erv C. Since the catalytic site is also located at this domain interface, care had to be taken so that the chosen positions for substitution were sufficiently away (not within 10 Å sphere around catalytic cysteine) from the catalytic region so as not to disturb the proteolytic efficiency of the enzyme.

We started with the single mutation K174R, since the replacement K→R is known to stabilize protein structure (Argos et al., 1979; Menéndez-Arias and Argos, 1989), and continued with the second (K174RV32S) and the third (K174RV32SG36S) mutations one after another to understand whether the effects of these mutations on thermostability are cumulative or intertwined. All the three mutants of papain thus formed were characterized for their enzymatic properties and thermal stabilities.

The thermal stability of these three mutants and wild type (WT) were compared in terms of \( T_{1/2} \), the temperature at which 50\% of the maximal activity of an enzyme is retained, and \( T_{50} \), the half-life of an enzyme at a particular temperature. Simultaneously, MD simulations, normal-mode analyses and energy calculations were carried out to model the three mutants together with the WT to elucidate the possible structural changes and the dynamic behavior of the four proteins and to assess their thermal stabilities. Of the three designed variants of papain, two showed more stability than the WT protein and retained their enzymatic activities at levels comparable with the original protein. The experimental data were corroborated with the MD simulation results.

### Materials and methods

**Construction, expression, purification and activation of papain mutants**

The wild-type (WT) propapain clone, pET30 propap (Choudhury et al., 2009), was used for introducing the mutations. SDM was carried out using QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA). In this method, double-stranded dam-methylated plasmid WT propapain gene isolated from Escherichia coli strains was used with a pair of complementary primers containing the desired point mutation. The mutagenesis primers were extended by pfuTurbo DNA polymerase in a thermostable, it was used as a template to improve thermostability of papain. The targets for mutagenesis in papain therefore were the three residues Val32, Gly36 and Lys174 in the interdomain region to be mutated to Ser, Ser and Arg, respectively, as found in the corresponding positions in Erv C. Since the catalytic site is also located at this domain interface, care had to be taken so that the chosen positions for substitution were sufficiently away (not within 10 Å sphere around catalytic cysteine) from the catalytic region so as not to disturb the proteolytic efficiency of the enzyme.

**Table I. Names of the mutants and oligonucleotides used for site-directed mutagenesis**

<table>
<thead>
<tr>
<th>Mutation sites</th>
<th>Name of the mutants (pro-form)</th>
<th>Oligonucleotide primers(^a)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>K174R</td>
<td>Propap-R (single mutant)</td>
<td>5′-GGACCAAAATTACATACTCATACGTAATTCTAGGGGTACAGGATG-3′</td>
<td>SDM performed on pET30-propap</td>
</tr>
<tr>
<td>K174RV32S</td>
<td>Propap-RS (double mutant)</td>
<td>5′-GGTCAGCAATTACATACTCATACGTAATTCTAGGGGTACAGGATG-3′</td>
<td>SDM performed on ‘Propap-R’</td>
</tr>
<tr>
<td>K174RV32SG36S</td>
<td>Propap-RSS (triple mutant)</td>
<td>5′-CGAAGCAATTACATACTCATACGTAATTCTAGGGGTACAGGATG-3′</td>
<td>SDM performed on ‘Propap-RS’</td>
</tr>
</tbody>
</table>

\(^a\)Underlined sequences represent the designated mutation for the target amino acid residue.
forms of the single, double and the triple mutants were designated as Pap-R, Pap-RS and Pap-RSS, respectively.

**Characterization of enzymatic properties of the mutants**

The refolding efficiency was monitored by gelatin SDS–PAGE, for which a 12% gel containing 0.1% gelatin was prepared by mixing gelatin with the resolving gel solution. The activated enzyme was diluted with non-reducing SDS–PAGE sample buffer, incubated for 20 min at 37°C, and separated by SDS–PAGE at 4°C. After electrophoresis, the gel was washed with 2.5% Triton X-100 in 20 mM Tris/HCl, pH 8.0 for 30 min. The gel was then washed extensively with distilled water to remove the detergent completely and incubated in activation buffer (100 mM Na-acetate, pH 5.5, 1 mM EDTA, 2 mM DTT) at 37°C for 30 min. Following this, the gel was incubated overnight at 37°C in the activation buffer without DTT to allow refolding and protein digestion. The gel was then stained with Coomassie Brilliant Blue R-250 and proteinase activity detected by appearance of clear bands on a dark blue background.

The proteolytic activity of the refolded purified wild-type and mutant papain was measured individually with substrate azocasein. For comparison, the pro-proteins were first activated to their mature forms, and then the hydrolysis of azocasein was determined by absorbance increase at 366 nm against time in 100 mM Na-acetate, pH 6.5, containing 2 mM EDTA. Assays were initiated by addition of 25 μl activated protein to 975 μl of reaction mixture containing 20 μl of 2% azocasein, 2 mM EDTA and 20 mM cysteine. After incubating the solution at 37°C for 30 min, the reaction was stopped by addition of 500 μl of ice-cold 5% trichloroacetic acid. Released azo-peptides were detected by measuring absorbance at 366 nm in a Thermo Nicolat Evolution 100 spectrophotometer using the specific absorption coefficient (ε366 = 40) for azocasein solution (Chen et al., 2003). One enzyme unit is defined as the amount of protease required to release 1 μg of soluble azopeptides per minute.

**Thermal stability of the mutants**

The kinetic thermal stability of the wild-type and the mutant proteins were determined by assaying the residual activity after heat treatment of the purified enzymes in a temperature-controlled water bath. Protein samples (50–100 μg of each pro-protein) in 100 mM Tris–HCl (pH 8.6) were incubated over a temperature range of 40–80°C at 5°C intervals for 10 min each and cooled immediately in an ice bath. After cooling, the pro-proteins were activated to their respective mature forms by incubating at 50/60°C for specific time for each protein (WT or mutant). Initial and residual enzyme activities (at each temperature) were determined using azocasein assay at 37°C as described previously. The temperature of incubation at which the enzymes showed maximum activity was termed $T_{\text{max}}$ and the temperature of incubation at which 50% of the maximum activity of the purified enzyme was retained was expressed as $T_{50}$. The protease activities were compared with that of the wild-type recombinant papain. The $t_{1/2}$ (half-life) values of the WT and the mutants were estimated by incubating the proteins at 60 and 65°C for 0 min to 4 h and then assaying the residual activity by azocasein assay. Similar experiments were done for mature Erv C purified from plant latex to determine the thermal stability of the protein except that the $t_{1/2}$ was determined at 70°C.

**In silico mutation, molecular dynamics and analysis**

The high resolution (1.65 Å) crystal structure of papain (pdb code: 9pap) was used as the starting model on which all the calculations were performed. Three oxygen atoms in the crystal structure, which irreversibly oxidizes the active site cysteine SG were not considered in this calculation and the catalytic residues were kept in their neutral form. Other ligands except crystallographic water molecules were omitted from this structure. Hydrogen atoms were generated to fill the unoccupied valencies of heavy atoms at the neutral state and the generated H atom positions were optimized by the Builder module of InsightII/Discover package (MSI Inc., USA) keeping the heavy atom positions fixed. This molecule was used as the template for subsequent calculations. All the simulations were carried out with the consistent valence force field (CVFF), and the cell multipole method with a dielectric constant value of 1 was used for non-bonded calculations.

The three mutants (K174R, K174RV32S and K174RV32SG36S) were generated by the Biopolymer module of the InsightII package and the lowest energy rotamer for the mutated amino acid(s) was chosen. Any crystallographic water molecule interfering with this rotamer was deleted from the structure. The atomic positions were further refined by 50 cycles Conjugate gradient (CG) minimization in the presence of the crystallographic water molecules keeping the backbone of the protein fixed. The proteins were then solvated by a 10 Å water shell using the Soak utility of InsightII. The positions of these waters along with the crystallographic waters were optimized by 500 cycles CG keeping the entire protein fixed, then 1000 cycles CG keeping the backbone fixed. The entire system was then simulated for a short period of 50 ps at 300 K and minimized using CG till the derivative reached <1 kcal/mol/Å. The mutant models were then used as starting models for MD simulations. The system was equilibrated for 1.4 ns at 300 K using the same protocol described by us previously (Ghosh et al., 2008), of which the last 1 ns was considered the product run. Coordinates were saved at 1 ps interval for the 1 ns product-run for the analysis. The 1 ns trajectory in each case was analyzed by the Analysis tool of the InsightII/Discover package. The last 100 ps average structures were used for analysis of the WT, single and double mutants. For the triple mutant, the amino acid R174 took two different rotamer conformations and both were quite stable in the trajectory. Out of the two, the rotamer similar to Erv C was chosen for analyses (averaged over 750–850 ps). All the averaged structures were optimized by 100 cycles steepest descent (SD) minimization and were used for subsequent calculations and structural analyses. The root mean square fluctuations of all the non-hydrogen atoms of the mutants with respect to their starting structures along the trajectory indicate that the systems achieve acceptable equilibrium state in 1 ns. However, we performed simulation for further 4 ns on the three mutants, Pap-R, Pap-RS and Pap-RSS, and the plateau in the graph of time evolution of root mean square deviations (RMSD) indicate that the structures continue to be stable beyond 1 ns with their RMSD values well within 1 Å (Supplementary data available at PEDS online, Fig. S1).
Therefore, the simulation length (1.4 ns including heating and equilibration) used in this study to generate the model structure of the three mutants was considered long enough to allow rearrangement of side chains of mutated residues in the mutant structures.

**Normal-mode analysis**

Normal-mode analysis (NMA) is a powerful tool for predicting collective molecular motions of biological macromolecules (Tama, 2003). This method is now widely applied to understand domain motions in multidomain proteins (Delarue and Sanejouand, 2002) which are quite often related to biological functions (Gerstein and Krebs, 1998). In the present study, we measured the interdomain stability of the two interacting domains of the WT and the three mutants of papain by analyzing the extent of domain movement using the low-frequency normal modes. The web-based program ElNemo (http://igs-server.cnrs-mrs.fr/elnemo/index.html), an interface to the elastic network model was used for this purpose to generate the low-frequency mode(s). We have calculated five normal modes (NMODE = 5) and the perturbation was given from −100 (DQMIN = −100) to +100 (DQMAX = +100) in 20 steps (DQSTEP = 20). Other parameters were used as default. From the five normal modes given by ElNemo (modes 7, 8, 9, 10 and 11) the lowest frequency mode, i.e. mode 7, was chosen for the analysis. The pdb files for mode 7 containing 11 structures for each of the mutants and the WT were downloaded. The animated view of these structures from pdb files was observed by DSVisualizer2.5 (Accelrys Inc.) software package to study the domain movement. Finally, the distance between the catalytic dyad comprising of residues Cys25 and His159 from the two domains and located at the opening of the interdomain cleft was calculated for each structure present in mode 7. The calculation was done to study how the distance fluctuates in the mutant proteins with respect to the wild-type protein during each lowest frequency mode of vibration.

**Solvent accessibility analysis and folding free-energy calculations**

Solvent accessible surface area (SASA) (Wang and Levinthal, 1991) for all the mutants and the wild-type protein was calculated by the web-based tool GETAREA1.0 (http://curie.utmb.edu/getarea.html) with a solvent probe of radius 1.4 Å. Folding free-energy calculations for the four structures were performed by the program FoldX (Schymkowitz et al., 2005) and compared.

**Results**

**Rational design of mutations**

Three-dimensional structure of papain molecule (Kamphuis et al., 1985) with 212 amino acid residues shows that the polypeptide chain is folded into two domains (L and R) of comparable sizes. The L-domain, the mainly α-helical domain, is composed of residues from 10 to 111 and 208 to 212, while the R-domain, a predominantly β-sheet domain, contains residues from 1 to 9 and 112 to 207 (Fig. 1). The longest helix LI (residues 25–42), in the L-domain spanning almost the entire interdomain cleft, contains the catalytically important residue Cys25 at its N-terminus. The central β-strand (SI) of the R-domain runs from residues 158 to 167 and contains the residue His159, the second residue of the catalytic dyad. The V-shaped active site cleft is situated at the interface of the L and R domains and is connected by a hinge region at the bottom of the cleft (Fig. 1). The proteolytic activity of papain is governed by the catalytic dyad, situated at the opening of the cleft, that exists as a Cys25−...His159+ zwitterion—a prerequisite for enzyme catalysis (Dardenne et al., 2003).

It was shown earlier from the structural and biochemical/biophysical studies on Erv C that three substitutions in the interdomain region together with an extra disulfide bond and shorter surface loops (compared to papain) impart thermostability to the enzyme (Guha Thakurta et al., 2004). In the present work, we focused on the three mutations at the domain interface. Of these three mutations, all in the interdomain region, two (V32S and G36S) belong to the helix LI domain interface. Of these three mutations, all in the interdomain cleft was calculated for each structure present in mode 7. The calculation was done to study how the distance fluctuates in the mutant proteins with respect to the wild-type protein during each lowest frequency mode of vibration.

![Fig. 1. Ribbon diagram of the structure of papain. Two domains L and R and the hinge region are represented by blue, red and green colors, respectively. Catalytic dyad residues (C25 and H159) and target residues for mutation (V32, G36 and K174) are indicated by sticks.](https://academic.oup.com/peds/article/23/6/457/1539896)
Propap-RSS, although the required pH and reducing agent remained unaltered, the time and temperature of conversion from pro to mature form were 40 and 45–50 min at 60°C, respectively. We could not perform this experiment on Propap-R as there was no detectable activity for this protein even after trying different conditions of activation. This is probably due to the low efficiency of refolding of this mutant. We also observed that the mutant Propap-R was not as stable as the WT and the other two mutants as assessed from SDS–PAGE at different times of storage of the refolded proteins after refolding (data not shown). Hence, further characterizations were carried out only on the double (Pap-RS) and the triple (Pap-RSS) mutants and compared with the wild-type enzyme.

**Enzymatic characterization and thermal stability of the mutants.** The proteolytic activities of the two mutants were comparable to that of the wild-type papain, with specific activities being 128.1 ± 0.6, 121.3 ± 1.3 and 130.2 ± 1.5 U/mg for the WT, Pap-RS and Pap-RSS, respectively, indicating that the amino acid substitutions did not affect the enzymatic property of papain. Determination of the effect of temperature on enzyme stability showed that the maximum proteolytic activity ($T_{\text{max}}$) of the WT, Pap-RS and Pap-RSS were at 50, 60 and 65°C, respectively (Fig. 3, Table II). This indicates that the thermostolerance of the mutants was more than that of the WT and it is the highest for Pap-RSS, the triple mutant. It was also observed that although $T_{\text{max}}$ for Pap-RSS was higher than Pap-RS, the fall in activity of Pap-RSS beyond $T_{\text{max}}$ was sharper than Pap-RS and the WT (Fig. 3). The $T_{50}$ values for the WT, Pap-RS and Pap-RSS were determined to be 68, 73 and 72°C, respectively (Table II). For mature Erv C, the template enzyme, the corresponding values for $T_{\text{max}}$ and $T_{50}$ were 70 and 76°C, respectively. For the thermal stability studies, the pro-forms of the WT and the variants were used for incubation at different temperatures. This is because, papain being a highly active protease, we were unable to purify sufficient amounts of the mature protein, even in the presence of reversible inhibitors, because of autocatalysis. It is also known that the mature part of these proteases already attain the 3D structure when they are in their zymogen form (pro-form) and no conformational change is observed on activation from pro to mature forms of the proteases (Groves et al., 1996; Cygler and Mort, 1997). Since thermal stability is known to depend on 3D structures and the mutations are in the mature part of the enzyme, we presume that the effect of temperature of incubation on the structure would be the same whether the enzymes are in the pro-form or in the mature form.

Kinetic stabilities (duration for which an enzyme remains active before undergoing irreversible inactivation) of the WT, Pap-RS and Pap-RSS were determined at 60°C ($T_{\text{max}}$ for Pap-RS) and 65°C ($T_{\text{max}}$ for Pap-RSS) under standard assay conditions mentioned in the previous section. The half lives ($t_{1/2}$) of the WT, Pap-RS and Pap-RSS at 60°C were 77, 114 and 171 min, respectively, and $t_{1/2}$ values at 65°C were 35, 45 and 80 min, respectively (Fig. 4, Table II). This shows
that the triple mutant, Pap-RSS, has the highest stability among all the three proteins. It was also observed that both the mutants Pap-RS and Pap-RSS retained more than 70% activity at 70°C (Fig. 3). The $t_{1/2}$ value for Erv C was found to be 120 min at 70°C ($T_{\text{max}}$ for Erv C).

### Computational analysis

**Analysis of the catalytic dyad and the S2 subsite.**

The effect of mutation(s) on the catalytic dyad residues (Cys25 and His159) has been examined by measuring the distance between the two atoms Cys25SG and His159ND1 (the atoms which are ionized and take part in the acylation/deacylation reaction) along the simulation trajectory of 1 ns. This distance has remained almost the same for the WT, double and triple mutants (Fig. 5) throughout the trajectory. However, this distance is more for the single-mutant K174R than the rest of the three, along the trajectory as seen in Fig. 5, indicating a destabilizing effect on the catalytic dyad in this case.

### Comparison of interdomain H-bonds/salt-bridges.

To explore the stabilization mechanism in the interdomain region, the H-bond/salt-bridge profiles of the protruding amino acid side chains in the modeled structures of the three mutants and the wild-type papain were analyzed and compared (Fig. 7). The modeled structure of the single-mutant Pap-R shows that an additional H-bond/salt-bridge, compared to the WT, has been formed due to the substitution (Fig. 7A and B). However, total number of H-bonds in the region remains the same like that of the WT (Table II). In the double-mutant Pap-RS, a second residue (V32), situated on the opposite wall of R174, has been mutated to Ser. As a result, an integrated H-bond network bridging the two domains has been formed between the guanidium group of R174 (from R-domain) with T14, P15, S32 (from L-domain) and Y184 (from R-domain) (Fig. 7C). A similar network (Fig. 7D) is also observed in the triple-mutant Pap-RSS. The triple-mutant Pap-RSS, with the third replacement G36S, has an additional water-mediated interdomain H-bond network formed by S36 (L-domain) with Q128 and P129 (R-domain) (Fig. 7E). So in both Pap-RS and Pap-RSS, the number of H-bonds significantly increases (9 H-bonds in Pap-RS and 10 in Pap-RSS against 3 in the WT) due to their respective substitutions (Table II). It is also to be noted that the wild-type papain contains two other H-bonds (Q19NE2...S176OG and S29OG...S131OG) in the same region.
which do not involve the three targeted residues and these H-bonds are retained both in Pap-RS and Pap-RSS mutants. The number of H-bonds in the inter-domain region of Erv C structure (Guha Thakurta et al., 2004) is also the same as that of Pap-RSS.

LI helix and backbone structure. The length of the helix LI (containing C25) in Pap-RS and Pap-RSS remains the same like that of the wild-type papain (Fig. 8A, C and D). However, a significant change is observed in the single-mutant Pap-R where the helix is shortened because of the absence of the two turns of the helix at C-terminus (Fig. 8B). The substituted residue Arg174 has a bulkier side chain compared with Lys and therefore in Pap-R, it has a different interaction pattern compared to the WT. It pushes the extended loop in the region A12-P15. The residue P15 takes a different rotamer conformation and a series of changes occur down stream. The T14 side chain reorients and makes a salt-bridge with the side chain of E35 resulting in a local strain in the backbone of the helix containing E35. The next residue being G36, a known helix destabilizing residue, rest of the helix beyond E35 (E35-T42) breaks (Fig. 8B). Though this kind of interaction between E35 and T14 is also observed in the triple-mutant Pap-RSS and in the crystal structure of Erv C (Guha Thakurta et al., 2004), full length of the helix L1 (C25-T42) is sustained in both the cases which is probably due to the presence of a Ser residue instead of Gly at position 36.

Analysis of domain motion. For papain and its mutants, which are two-domain proteins, NMA has been used for calculating the extent of domain movement as a measure of stability, since less flexibility (less domain movement in this case) would indicate higher stability. The maxima and minima of openness between the two domains, usually known as ‘opening’ and ‘closing’ of the domains, were calculated using NMA. For this purpose, we have chosen to calculate the distance between the two representative Co atoms (C25A and H159A) from the two domains situated close to the open end and away from the hinge region of the interface (Fig. 1) to monitor the extent of domain motion. Observations in this study indicate that the triple-mutant Pap-RSS has the least interdomain motion (Table II) and thus may be considered to have the highest interdomain stability among the three mutants and the WT.

Discussion
In the present study, we have explored the feasibility of improving the thermostability, without disturbing the catalytic efficiency, of papain by introducing multiple H-bonds/salt-bridges through substitution of three amino acids in the
interdomain region based on the structure of a naturally occurring thermostable protein Erv C from the same family. A number of earlier studies on other proteins have demonstrated that higher electrostatic and hydrogen bonding interactions in thermostable proteins are responsible for the increased stability compared to their lesser stable counterparts (Vogt and Argos, 1997; Vogt et al., 1997; Ladenstein and Antranikian, 1998; Karshikoff and Ladenstein, 1998; Szilagyi and Zavodszy, 2000; Acharya et al., 2004; Tigerstrom et al., 2004). Importance of interdomain region for cysteine proteases in this context has also been discussed earlier by Vernet et al. (1992).

The single mutant, Lys174Arg (Pap-R), was prepared first which did not impart any thermostability to papain. This single substitution in Pap-R in fact led to a destabilization of the protein as observed from inefficient refolding and subsequent failure to characterize the protein kinetically under conditions normally used for the WT. When the second mutation (Val32Ser) was added to the single mutant to generate the double mutant (K174RV32S—Pap-RS), we observed an increase in thermostability of the protein. The addition of the third mutation, Gly36Ser, in the series to create the triple mutant (K174RV32SG36S—Pap-RSS) further increased the thermal stability. The $T_{\text{max}}$ of Pap-RS and Pap-RSS was found to increase by 10–15°C when compared to the WT (Fig. 3). The $T_{50}$ values of Pap-RS and Pap-RSS, though higher than the WT, do not vary much among themselves. Apart from a general enhancement of thermostability, it was observed that Pap-RS retains its maximum proteolytic activity for almost 50–60 min at the temperature of its peak activity (60°C) and Pap-RSS does that for 40 min at 65°C. Also, Pap-RSS showed steady activity profile for a longer range of temperature (40–65°C) compared to the WT and Pap-RS both of which started with a lower activity value at 40°C till they reached their maximum activity (Fig. 3). So, on the basis of highest $T_{\text{max}}$ (65°C) and longest $t_{1/2}$ (80 min at 65°C), we conclude that Pap-RSS is the most thermostable among the three. Descent of proteolytic activity of both the WT and Pap-RS was comparatively slower than Pap-RSS whose activity fell quickly to less than 18% at 75°C (100% at 65°C) compared to 33% (100% at 60°C) for Pap-RS (Fig. 3). Faster inactivation rate is probably a desirable feature for an industrial enzyme where the enzyme is required to act at a desired temperature for a limited span of time only.

Molecular flexibility of the active site of an enzyme is essential for substrate binding, catalysis and product release. On the other hand, higher thermostability necessitates an increase in the rigidity of the structure. As a result, stabilization of an enzyme is quite often associated with a loss of catalytic activity (Shoichet et al., 1995). Most of the earlier attempts to stabilize papain through immobilization of the enzyme or through other methods have resulted in reduction of the enzyme activity (Axen and Ernback, 1971; Rajalakshmi and Sundaram, 1995; Afaq and Iqbal, 2001; Lei et al., 2004; Sangeetha and Abraham, 2006; Srinivasrao et al., 2006). But in our present study, we have generated papain variants with higher thermostability almost without compromising its activity. This has been achieved by taking care, during the design of the mutations, that the catalytic centre and the substrate binding subsites of papain are away from the mutation sites.

While carrying out the molecular modeling studies of the three designed mutants, the shortening of helix L1, less number of H-bonds/salt-bridges, higher interdomain flexibility and higher value of stability free-energy (Table II) indicate that the stability of the single mutant (Pap-R) may even be less than that of the WT. Since the negative charge of the active site Cys thiol group is maintained substantially by the dipole moment of the $\alpha$-helix L1, the reduction in the length of L1 in the single mutant (Fig. 8) leads to a reduction of dipole moment value which in turn is likely to affect the Cys$\ldots$His$^+$ ion pair and thus the catalytic activity of the mutant (Doran and Carey, 1996; Dardenne et al., 2003). Moreover, the distance between the catalytic dyad residues is higher in single mutant than what is observed for the WT in the MD trajectory (Fig. 5). This also explains the similar experimental results for the single mutant which showed no detectable proteolytic activity.

In case of double and triple mutants Pap-RS and Pap-RSS, Val32→Ser substitution, in addition to K174R, not only provide the extra room needed for the bulkier R174 side chain (avoiding the steric clash with Val32), but also create a favorable electrostatic potential for the Arg residue and intensifies interdomain salt-bridges in the region. The G→S mutation at position 36 in triple mutant Pap-RSS provides an
additive effect on stability of the double-mutant Pap-RS due to two reasons: it creates a water-mediated interdomain H-bonding and probably it also enhances the entropy contribution of the helix since Gly → Xaa substitution is known for such an effect (Matthews et al., 1987). Our molecular modeling calculations indicate that the double-mutant Pap-RS has more interdomain stability compared to the WT, though the overall stability of Pap-RS is less than the WT (Table II). The experimental results on kinetic stability show that Pap-RS retains its activity (Table II) for a longer time (compared to the WT) at 60 and 65°C, which indicate the importance of interdomain stability on activity. Both the biochemical and molecular modeling studies therefore establish that two thermostable variants of papain, Pap-RS and Pap-RSS, have been designed and generated of which Pap-RSS is more thermostable. A comparison with the kinetic stability values of the template enzyme Erv C shows higher values for $T_{\text{max}}$, $T_{50}$ and $t_{1/2}$. This is not unexpected.

**Fig. 7.** Comparison of interdomain H-bonds/salt-bridges involving the residue at position 174 in (A) wild type, (B) Pap-R, (C) Pap-RS and (D) Pap-RSS. (E) Water (red ball) mediated H-bonds generated in Pap-RSS due to substitution of G36 → S.
because Erv C not only has a similar interdomain environment as that of Pap-RSS but also has an additional S-S linkage and a more compact molecular structure (Guha Thakurta et al., 2004). The success of the methodology also suggests that it may be applied to other enzymes of this family and/or other multidomain proteins either on its own or in combination with other strategies based on redesign of the interactions in the interdomain region of an enzyme.

Fig. 8. Influence of K174→R substitution on LI helix and the preceding loop region in (A) wild type, (B) Pap-R (having the LI helix of reduced length), (C) Pap-RS and (D) Pap-RSS. Residue 174 is represented as ball and stick model in each diagram.
Supplementary data
Supplementary data are available at PEDS online.

Funding
The work was partially supported by Council of Scientific & Industrial Research (No. 21/0653/06/EMR-II), Department of Science and Technology (No. SR/WOS-A/LS-215/2008) and Department of Biotechnology (No. BT/PRO0139/R&D/15/11/96), Government of India.

References