In Silico Designing of an Industrially Sustainable Carbonic Anhydrase Using Molecular Dynamics Simulation

Sachin Kumar Bharati, Mousumi Hazra, Manish Paul, Swati Mohapatra, Deviprasad Samantaray, Ramesh Chandra Dubey, Shourjya Sanyal, Saurav Datta, and Saugata Hazra

1Department of Biotechnology and Centre for Nanotechnology, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India
2Department of Botany and Microbiology, Gurukula Kangri University, Haridwar 249404, Uttarakhand, India
3Department of Microbiology, Orissa University of Agriculture and Technology, Bhubaneswar 751003, Odisha, India
4Complex and Adaptive System Laboratory, School of Physics, University College Dublin, Dublin 4, Ireland

ABSTRACT: Carbonic anhydrase (CA) is a family of metalloenzymes that has the potential to sequester carbon dioxide (CO2) from the environment and reduce pollution. The goal of this study is to apply protein engineering to develop a modified CA enzyme that has both higher stability and activity and hence could be used for industrial purposes. In the current study, we have developed an in silico method to understand the molecular basis behind the stability of CA. We have performed comparative molecular dynamics simulation of two homologous α-CA, one of thermophilic origin (Sulfurihydrogenibium sp.) and its mesophilic counterpart (Neisseria gonorrhoeae), for 100 ns each at 300, 350, 400, and 500 K. Comparing the trajectories of two proteins using different stability-determining factors, we have designed a highly thermostable version of mesophilic α-CA by introducing three mutations (S44R, S139E, and K168R). The designed mutant α-CA maintains conformational stability at high temperatures. This study shows the potential to develop industrially stable variants of enzymes while maintaining high activity.

INTRODUCTION

Ever-increasing concentration of CO2 is a pertinent threat, whose major contributor is flue gas from coal-fired power plants. Because of its low cost and high availability, there is a rapid increase in the usage of coal, especially in power plants. Hence, there is an urgent demand for developing techniques that would capture the generated carbon dioxide to prevent its excessive emission to the atmosphere. Current procedures, including amine absorption, calcium hydroxide absorption, gas separation, and so forth, are of limited or no success, due to very high energy consumption. Amine absorption, which is the most energy-efficient CO2 capture technology to date, is associated with around 85% “cost of electricity”. This indicates that 85% of the energy generated by a coal-fired power plant will be consumed in capturing the CO2 emitted by the same power plant, making development of novel CO2 capture technology absolutely essential for the carbon-constraint world in the 21st century. The ultimate goal of our research is to develop a biological method using enzymes to capture CO2.

In nature, bacteria transform CO2 into bicarbonate at physiological pH, using the carbonic anhydrase (CA) class of enzymes as catalysts. CA is a family of zinc-containing metalloenzymes that stimulates the interconversion of carbon dioxide (CO2) and water into bicarbonate and proton and hence, this enzyme helps in transporting CO2 in the form of water-soluble bicarbonate.

CO2 + H2O ⇌ H2CO3

H2CO3 ⇌ H+ + HCO3−

Therefore, the above-mentioned enzymatic mechanism can be used in tackling the pollution problem caused by the emission of CO2 from industries. CA is found ubiquitously in prokaryotic microorganisms (bacteria), algae, plants as well as in higher eukaryotic animals. There are five classes of evolutionary distinct CAs, namely, α, β, γ, δ, and ζ. All of these classes are found in bacteria, except for the δ class, which is found in diatoms. These different classes of CAs share very low sequence identity and have different structural folds (Figures 1 and S1). Differences in sequence homology and structure along with identity in function make CA an ideal example of convergent evolution. To adapt with environ-
mental changes, different classes of CAs change their overall secondary structure without altering their active site. All of these classes are of ancient origin and appear to have evolved independently from one another, having a zinc metal ion in their active site. α-CA has three conserved histidine residues in the active site (Figure S2).19,20

Figure 1. a. Cartoon representation and sequence alignment of Ng α-CA and Ssp α-CA. b. Structural alignment of Ng α-CA and Ssp α-CA. These two alignments clearly show the presence of similarity in the protein tertiary structure in spite of differences in the amino acid sequence.
In this study, we have chosen prokaryotic α-CAs, as they are inexpensive as well as easily available for experimental validation and concurrent production using cell lines. One of the major requirements to make the project successful is the productivity of the enzyme. To develop the protein, we depend on an *Escherichia coli*‐based bacterial expression system for simple reasons: *E. coli* (1) is well characterized, (2) has a faster growth rate, and (3) requires optimized growth conditions both in shake flasks and bioreactors. Thus, there is a need for selecting a CA of mesophilic origin. We have selected α-CA from *Neisseria gonorrhoeae* because of its mesophilic origin, considering that it would be better to express the protein in the *E. coli* expression system. In addition to this, α-CA from *N. gonorrhoeae* has a high sequence identity and a structure nearly similar to that of *E. coli* α-CA (Figure 2). Moreover, there are also some other attributes for choosing a mesophilic model compared to a thermophilic one. Besides the similarity of α-CA between *N. gonorrhoeae* and *E. coli*, both organisms are Gram-negative and inhabit similar types of environments, which further supports our choice of mesophilic systems.

α-CAs are among the fastest catalyzing enzymes; however, one of the biggest challenges of using them in industries is that they are not stable enough to withstand the harsh environment found in the power plant exhaust. We have taken a highly interdisciplinary approach including in silico techniques followed by genetic engineering to develop an industrially stable, engineered α-CA that can tolerate harsh operating conditions (in terms of temperature and pH) typically encountered in CO₂ capture processes for power plant flue gas. The engineered enzyme could be immobilized to a membrane, and the enzyme-tagged membrane could be used as a nanodevice to sequester carbon dioxide from the flue gas outlet of the power plant.

In this study, our goal is to design a mesophilic α-CA (*N. gonorrhoeae*) with higher thermal stability while retaining its enzyme activity. To gain insight into the unfolding mechanism prior to the rationally designing the mutations, comparative molecular dynamics (MD) simulations were performed on the mesophilic α-CA (*N. gonorrhoeae*) and a thermophilic α-CA (*Sulfurohydrogenibium sp.*) at temperatures of 300, 350, 400, and 500 K, each for 100 ns. Using MD simulation trajectories, different stability-determining variables such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (*Rₕ*), solvent-accessible surface area (SASA), hydrogen bonds, salt bridges, unfolding pathway, principal component analysis (PCA), and free-energy landscape (FEL) were analyzed. On the basis of the analyzed results, by comparing differences in those analyses of stability-determining factors, a mutant variant of mesophilic α-CA was designed, containing the variant residues: S44R, S139E, and K168R (Figure 3). This was followed by MD simulations at the above-mentioned temperatures for the designed mutant. Analysis not only reveals a sharp increase in the stability of the engineered variant of α-CA but also predicts the stability to be comparable to that of the thermophilic homologue. This study would be a good platform for...
Figure 3. continued
thermostability compared to Ng α-CA, and the structure with β sheets in blue is thermophilic Ssp α-CA. On the basis of our comparative analysis of the distribution of salt bridges in these two structures and their effect on stability, we have designed a mutant variant of mesophilic Ng α-CA, whose β sheets are shown in brown.

RESULTS AND DISCUSSION

Salt Bridge Analysis. The bonds formed between the negatively charged side chains of aspartic or glutamic acid and the positively charged side chain of lysine or arginine are salt bridges, which stabilize the secondary structure of proteins. The stability increases with both increased number of salt bridges and the reduced distance between two salt bridge-forming side chains.

In Ssp α-CA, there are five salt bridges: Lys36 − Glu99, Lys41 − Glu223, Lys39 − Glu223, Lys41 − Glu335, and Arg165 − Glu223. In place of these salt bridges, Ng α-CA has four salt bridges: Arg136NH1 − Asp194OD1, Arg136NH2 − Asp194OD2, Arg136NH2 − Asp194OD1, and Arg136NH2 − Asp194OD2. In Ng α-CA, these salt bridges are responsible only for the local stabilization of Ng α-CA; however, in the case of Ssp α-CA, four salt bridges are shown to stabilize a specific domain of the protein and the remaining one stabilizes another domain. In Ssp α-CA, there are a total of eight residues involved in salt bridge formation. MD simulations show that the salt bridges of Ssp α-CA maintain an overall fixed length throughout all of the temperature simulations. But in the case of Ng α-CA, the salt bridge lengths vary in a broad range and they are very unstable. The stable salt bridges in Ssp α-CA are majorly responsible for its greater thermostability compared to Ng α-CA.

On the basis of these observations, we have introduced three mutations in mesophilic Ng α-CA. From sequence and structure alignment, we have found that the residues Lys41, Glu136, and Arg165 form salt bridges in thermophilic α-CA, whereas they are absent in its mesophilic counterpart. Mesophilic Ng α-CA contains Ser44, Ser139, and Lys168 in place of those salt bridge-forming residues. From this observation, we have designed a mutant variant consisting of S44R, S139E, and K168R. As a result, the mutant variant contains five new salt bridges. Glu226 in the mutant variant is involved in forming four salt bridges with Arg44 and Arg168. Another salt bridge is formed between Glu139 and Arg136. So, the mutant variant now contains a total of nine salt bridges, resulting in enhanced stability compared to its thermophilic counterpart.

We have analyzed the salt bridge length variation of all of the CA systems at four different simulation temperatures (300, 350, 400, and 500 K) (Figure 4, Table 1). We have found that the lengths of all of the salt bridges for mesophilic α-CA fluctuate considerably, and at high temperatures, the salt bridges become destabilized. At 300 K, the salt bridge-forming residues (Arg136 and Asp194) of Ng α-CA are located at α3 and β11. With increasing temperature, the β11 secondary structure disintegrates, making Asp194 flexible. This causes weakening of the Arg136 − Asp194 salt bridge. In addition, it stimulates the loss of an α3 secondary structure imposing instability to Arg136. These cumulative effects make the Arg136 − Asp194 salt bridge fully unstable at 500 K.

In the case of thermophilic α-CA, all of the five salt bridges remain stabilized and the lengths of the four salt bridges, Lys36 − Glu99, Lys39 − Glu223, Arg165 − Glu223, and Lys136 − Glu133, remain unchanged for all four temperature simulations. It has also been observed that three salt bridge-forming residues of Ssp α-CA, Lys39, and Glu223 remain in the loop at 300 K but occupy β1 and β11 secondary structures at higher temperatures (350, 400, and 500 K). It can be concluded that it is this strong interaction and the salt bridge network between these three residues that help in the formation of the secondary structure in the thermophilic protein. Also the two residues, Glu133 − Lys136, which form a salt bridge, maintain the stability of the α-helix in which they reside.

In the mutant variant, three modified residues, S44R, S139E, and K168R, resulted in five new salt bridge interactions (Arg44NH1 − Glu226OE2, Arg44NH2 − Glu226OE2, Arg168NH1 − Glu226OE1, Arg168NH2 − Glu226OE1, and Arg136NH1 − Glu133OE2). These five new salt bridges show extra
stability at all four temperatures. In particular, the salt bridges formed between mutant residues Arg44 and Arg168 with Glu226 remain significantly stable even at 400 and 500 K. Salt bridges between Arg136 and Asp194, which are common in both the wild-type and mutant variants, show differences in bond length (6.6 and 4.0 Å respectively). Another stabilized salt bridge in mutant α-CA (Arg136-Glu39) is found to maintain its length within 4.0–4.9 Å throughout all temperature simulations.

Figure 4. Salt bridge length comparison of a. Ng α-CA, b. Ssp α-CA, and c. mutant Ng α-CA. a. Ng α-CA: salt bridge interaction happens between two residues; Arg136—Asp194. Lengths increase significantly with increasing temperature. b. Ssp α-CA: Different sets of residues forming salt bridges all over the protein. The length of bridges remains intact or minor changes happen even in higher temperature simulations. c. mutant Ng α-CA: Change in three residues results in several new salt bridges; as in Ssp-CA, the length remains intact at higher temperatures.
hydrogen bonds in mutant bridge is the crucial reason here for the stability of those thermophilic decreases, and the protein becomes unstable. The average helps in increasing the protein thermostability. With an increase bond is responsible for maintaining the protein integrity and extra stability due to the induced mutations (Table S1). A salt intramolecular hydrogen bonds suggest that the mutant gains the number of average intramolecular hydrogen bonds in the nearly maintained up to 350 and 400 K (162.78 and 161.64). The importance of intramolecular hydrogen bonds in maintaining the protein thermostability. Mutant CA shows an even higher average number of hydrogen bonds at 300 K (170.18), which is 350 K and 400 K (168.20, 168.84, 160.62, and 145.50). This clearly depicts the protein thermostability. In such cases, fewer protein residues remain free for interacting with the solvent and hence they are able to retain more secondary content even at a high temperature. But in the case of mesophilic protein such a rapid decrease in intramolecular H-bonding in a protein–solvent system even when the protein gradually unwraps from its native state to an unfolded state. The number of protein–solvent intramolecular hydrogen bonds rapidly decreases in the case of mesophilic CA compared to its thermophilic counterpart across the entire range of temperatures. Increased temperature also causes the loss of intermolecular H-bonding in a protein–solvent core region of the protein gradually becomes unwrapped with its unfolding. Surfacing of the hydrophobic region of protein increases the affinity of the protein core to the solvent molecule. The average number of intermolecular hydrogen bonds in mesophilic and thermophilic α-CA are 465.94, 432.88, 388.79, and 303.38 and 510.85, 457.73, 412.06, and 319.65, respectively. The decrease in the number of intermolecular hydrogen bonds is not usual. So, residues in the relatively less stable mesophilic Ng α-CA, as supported by other analyses such as RMSD, RMSF, PCA, and FEL.

Hydrogen Bonding Pattern. Like salt bridges, hydrogen bonds are also important for maintaining the protein secondary structure. Hydrogen bond from two different residues remains connected with each other within 3.5 Å. In this study, we have considered two different types of hydrogen bonds: (a) intramolecular hydrogen bonds (protein–protein) and (b) intermolecular hydrogen bonds (protein–solvent) (Table S1).

Intramolecular Hydrogen Bonds. This type of hydrogen bond is responsible for maintaining the protein integrity and helps in increasing the protein thermostability. With an increase in temperature, the number of intramolecular hydrogen bonds decreases, and the protein becomes unstable. The average number of intramolecular hydrogen bonds in mesophilic and thermophilic α-CA are 168.20, 168.84, 160.62, and 145.50 and 163.20, 164.19, 163.65, and 154.07, respectively at 300, 350, 400, and 500 K. The number of hydrogen bonds that maintain the protein integrity decreases rapidly in the case of mesophilic CA compared to the thermophilic one. This clearly depicts the importance of intramolecular hydrogen bonds in maintaining the protein thermostability. Mutant CA shows an even higher average number of hydrogen bonds at 300 K (170.18), which is nearly maintained up to 350 and 400 K (162.78 and 161.64). The number of average intramolecular hydrogen bonds in the mutant also decreases to 148.72, which is higher than that in the case of mesophilic CA at 500 K (145.50). These values of intramolecular hydrogen bonds suggest that the mutant gains extra stability due to the induced mutations (Table S1). A salt bridge is the crucial reason here for the stability of those hydrogen bonds in mutant α-CA. Five additional salt bridges were observed in the mutant variant compared to mesophilic Ng α-CA (Arg44 NH1−Glu226 OE2, Arg65 NH2−Glu226 OE2, Arg66 NH1−Glu226 OE1, Arg66 NH2−Glu226 OE1, and Arg156 NH1−Glu159 OE2). But a more important factor is the position of those salt bridges. They are distributed in the protein rather than being on one particular side and bringing the loops together (Figure S9). This distribution helps the protein to hold the entire structural architecture, especially at higher temperatures, enhancing stability.

Intermolecular Hydrogen Bonds. Intermolecular hydrogen bonds maintain the interactions between the protein and solvent. The interaction between the protein and solvent molecule increases when the hydrophobic core region of the protein gradually becomes unwrapped with its unfolding. Surfacing of the hydrophobic region of protein increases the affinity of the protein core to the solvent molecule. The average number of intermolecular hydrogen bonds in mesophilic and thermophilic α-CA are 465.94, 432.88, 388.79, and 303.38 and 510.85, 457.73, 412.06, and 319.65, respectively. The decrease in the number of intermolecular hydrogen bonds is less in the case of mesophilic CA compared to its thermophilic counterpart across the entire range of temperatures. Increased temperature also causes the loss of intermolecular H-bonding in a protein–solvent core region to the solvent molecule. The interaction between the protein and solvent increases when the hydrophobic core region of the protein gradually becomes unwrapped with its unfolding. Surfacing of the hydrophobic region of protein increases the affinity of the protein core to the solvent molecule. The average number of intermolecular hydrogen bonds in mesophilic and thermophilic α-CA are 465.94, 432.88, 388.79, and 303.38 and 510.85, 457.73, 412.06, and 319.65, respectively. The decrease in the number of intermolecular hydrogen bonds is less in the case of mesophilic CA compared to its thermophilic counterpart across the entire range of temperatures. Increased temperature also causes the loss of intermolecular H-bonding in a protein–solvent system even when the protein gradually unwraps from its native state to an unfolded state. The number of protein–solvent intramolecular hydrogen bonds rapidly decreases in the case of a comparatively stable protein, as we can see in the case of Ssp α-CA and mutant Ng α-CA. In such cases, fewer protein residues remain free for interacting with the solvent and hence they are able to retain more secondary content even at a high temperature. But in the case of mesophilic protein such a rapid decrease in intramolecular H-bonds is not usual. So, residues in the relatively less stable mesophilic Ng α-CA remain bound to solvent molecules through intramolecular H-bonds. This phenomenon causes structural instability of mesophilic Ng α-CA at high temperatures. In the case of the mutant, average values of intermolecular hydrogen bonds are 467.05, 434.24, 396.30, and 301.46, respectively, for 300, 350, 400, and 500 K. So, the decrease in the number of average intermolecular hydrogen bonds observed in thermophilic and mutant proteins

| salt bridge          | mesophilic α-CA | thermophilic α-CA | meso_mutant α-CA |
|----------------------|-----------------|-------------------|------------------|
| Arg168 NH2−Glu226 OE1 | 4.6             | 4.6               | 4.6              |
| Lys39 NZ−Glu99 OE1   | 9.6             | 9.6               | 9.6              |
| Arg44 NH2−Glu139 OE2 | 4.8             | 4.5               | 4.5              |
| Glu223 OE1−Arg136 NH1| 3.4             | 3.4               | 4.5              |
| Glu223 OE2−Arg136 NH2| 3.4             | 3.4               | 4.5              |

Table 1. Salt Bridges in α-CA and Their Length Variation at Different Simulation Temperatures

From the MD study of average salt bridge length of all of the α-CA systems, it is evident that more number of strong and comparatively stable salt bridges in both thermophilic and mutant α-CA are responsible for their greater thermal adaptability compared to the mesophilic one. These stable salt bridges are also important for retaining the local secondary structure content in thermophilic and mutant α-CA. Modified residues in mutant α-CA form extra salt bridges, resulting in a further increase in stability compared to the mesophilic α-CA, as supported by other analyses such as RMSD, RMSF, PCA, and FEL.
Figure 5. continued
Figure 5. continued

| (b)  | 0 ns | 300K | 350K | 400K | 500K |
|------|------|------|------|------|------|
| 10 ns|      |      |      |      |      |
| 20 ns|      |      |      |      |      |
| 30 ns|      |      |      |      |      |
| 40 ns|      |      |      |      |      |
| 50 ns|      |      |      |      |      |
| 60 ns|      |      |      |      |      |
| 70 ns|      |      |      |      |      |
| 80 ns|      |      |      |      |      |
| 90 ns|      |      |      |      |      |
| 100 ns|     |      |      |      |      |

| (c)  | 0 ns | 300K | 350K | 400K | 500K |
|------|------|------|------|------|------|
| 10 ns|      |      |      |      |      |
| 20 ns|      |      |      |      |      |

indicates that they have less unwrapping of their hydrophobic core region compared to the mesophilic protein. Thus, thermophilic and mutant proteins make fewer contacts with the outer solvent molecule with increasing temperature and maintain their higher stability (Table S1).

**Unfolding Pathway.** Analysis of the unfolding pathway of a protein predicts the pattern in which it unwraps from its native fully folded form to a completely unfolded state. Different protein conformations obtained from successive time scales throughout the MD simulation trajectory present specific stable and unstable domains in the protein structure. Different secondary structure contents, which stabilize protein’s three-dimensional (3D) conformations, can also be determined by this method. In this study, the detection of unfolding pathway suggests that the induced mutations in mesophilic α-CA have a stabilizing influence on their corresponding secondary structure elements, which in turn contribute to the overall rigidity of the conformation (Figure 5). In short, we can look through a protein at the intramolecular level to find important interactions responsible for maintaining its stability during its proper folding. Unfolding pathway of all three (mesophilic, thermpophilic and mutant) α-CA were studied according to successive time scales of increased temperature through their respective MD simulation trajectory.

Mesophilic α-CA shows significant loss in its secondary structure contents along its unfolding pathway, mostly at all different temperatures. The α4 helix gradually decreases in size for the 300 K temperature simulation and is eliminated from the 20 to 80 ns time scale. This is probably caused by a certain increase in RMSF values of the residues Ala159–Leu162 in this region. Moreover, the N-terminus of the largest β-sheet in the core hydrophobic twisted β-sheet region becomes shorter at 50 ns and onward during the 300 K temperature simulation. This N-terminal region of the β-sheet in mesophilic α-CA shows visible loss of its contents also at 10, 30, 40, 80, and 100 ns time scale of the 350 K temperature simulation. This extended fragment of core β-sheet is so unstable that it turns into a loop at the 30–100 ns time scale at 400 K. At higher temperatures like 500 K, no trace is found for this region between the time scale of 40 and 100 ns. The residue, Arg84, at the tip of the N-terminus of the above-mentioned β-sheet has a salt bridge interaction with Asp116 of its next β-sheet. This salt bridge becomes distorted between these time scales at 500 K, causing the deformation of the β-sheet part. The N-terminus of the α3 helix disappears and transforms into a loop in mesophilic α-CA at 50–100 ns at 350 K. This part of the α3 helix also disappears at 40 and 80 ns at 500 K. Arg136 of α3 forms a salt bridge with Asp194 from β11 at the beginning time scale of the 350 K simulation. This salt bridge breaks up between 50 and 100 ns at 500 K, causing the disappearance of both α3 and β11. Both α1 and α2 helices disappear at 40 and 50 ns of the 400 K temperature simulation in mesophilic α-CA and transform into a loop. The α1 helix is shown to reappear after 50 ns and exist till the end of the 400 K simulation. Lys25 of α1 has a salt bridge interaction with Glu12 from the N-terminus loop of the protein. This salt bridge remains stable up to the 1–30 ns time scale of the 400 K simulation but after that, the salt bridge breaks down, resulting in the deformation of α1. Lys37 of α2...
makes two salt bridges with Asp35 from the adjacent N-terminus loop. These salt bridges remain stable up to the 1–30 ns time scale of the 400 K simulation. After that, the amine side chain of Lys37 moves toward a direction in which the salt bridges deform and α2 transforms into a loop. Between the time range of 20 and 60 ns of the 500 K simulation, a short 310 helix appears at the C-terminus of mesophilic α-CA. Another 310 helix appears adjacent to the previously mentioned 310 helix at the end of the 500 K simulation. Frequent formation of such 310 helices indicates weak interatomic interactions between the backbone residues at different local regions of mesophilic α-CA, leading to the massive loss of secondary structure contents. All of the α-helices gradually decrease in size and finally disappear at 70 ns at 500 K. The β-sheets in the core of the twisted hydrophobic region also become very short in size, imparting structural instability to mesophilic α-CA at 500 K.

For thermophilic α-CA, no significant loss or deviation of the secondary structure content from the initial conformation has been noticed. All of the β-sheets, especially the largest sheet in the central region of the hydrophobic twisted β-sheet region, remain structurally intact and stable all along the temperature simulations. A 310 helix in the structure of thermophilic α-CA frequently appears at 10, 50, 70, and 100 ns time scale of the 300 K simulation. The salt bridges between Lys37 and Asp35 remain stable at those time scale of the 300 K simulation, which might be the cause of the repetitive appearance of this 310 helix. Shortening of the N-terminal of α1 and α2 helices has been observed at 350 and 400 K, but later these helices, especially α1, reappear with greater stability and remain intact for the rest of the simulations, even at the end of the 500 K temperature simulation. The stabilization of both α1 and α2 helices is mainly caused by the formation of salt bridges between Gln16 of α1 and Lys25 of α2.

In the case of mutant α-CA, most of its secondary structure contents are maintained in a compact form for an extended time scale during all of the temperature simulations. There are no visible changes or significant loss in its secondary structure up to 80 ns at 300 K. Specifically small structural fluctuation has been observed in the core twisted β-sheet region and the surface β/β region. At 90 ns of the 300 K simulation, the N-terminus of the largest β-sheet in the central hydrophobic core becomes reduced in size. The reduced part of this β-sheet reappears at 100 ns at the same temperature. The β8 sheet of mutant α-CA disappears at 20 ns in the 300 K simulation. The reduction of the N-terminus of the largest β-sheet occurs again at 10, 20, 50, and 80 ns of the 350 K temperature simulation. But this part remained fully compact for the rest of the time scale at 350 K. At 400 K, this part of the largest core β-sheet reduces at 30, 80, and 90 ns. At 500 K, the N-terminus of the β-sheet remains obsolete from 10 to 40 ns. But after that time scale, this β-sheet remains intact in full length for the rest of the simulation time at 500 K. The stability and reappearance of the core β-sheet in all of the above-mentioned cases are mainly because of the existence of a strong salt bridge between Arg104 and Glu129. Interestingly, at the end of the simulation at 500 K, an extra β-sheet is formed. The appearance of the small fragment of this β-sheet might be due to the formation of an adjacent 310 helix. Asp194 of the small β-sheet fragment has a salt bridge interaction with Arg136 of the 310 helix. On the other hand, the 310 helix is formed due to the generation of a salt bridge between the mutant residue Glu139 and Arg136. The contributions from both this β-sheet and 310 helix probably have some impact in retaining the other secondary structure of the mutant α-CA, mostly intact at higher temperatures.

The protein secondary structure loses its integrity with increasing temperature. Increasing temperature also results in a high rate of dynamism in the structure. This leads to the possibility of formation of newer bonds, which is not possible at normal temperature. If two amino acids have side chains far apart while forming some interaction, then with increasing temperature, the kinetic energy increases and the side chains come together. That would lead to the formation of some stronger bonds, which is not possible at lower temperatures.

Comparative unfolding pathway depicts that Ng α-CA lost most of its secondary structure significantly from 70 to 100 ns at 500 K. Notably, it can be found that helical contents are totally absent at the 70 ns time scale of the 500 K simulation in Ng α-CA, and all of them transform into extended and flexible loops. On the other hand, both Ssp α-CA and mutant Ng α-CA retain all their secondary contents at those high temperatures. Moreover, we have also reported that, during the time scale of 70, 80, and 100 ns of the 500 K temperature simulation, Ssp α-CA and Ng α-CA show some small β-plated sheets on their surface, which seem to be responsible for their higher thermostability.

The energy compensation for the formation of such secondary structures with increasing temperature has already been discussed in the section of FEL. A protein has to cross a high-energy barrier for the deformation of those types of temporarily stable secondary structures. In our study, for thermophilic Ssp α-CA and mutant Ng α-CA, the frequency of occurrence of such structures is more than that in its mesophilic counterpart, which make them more thermostable.

**Constraint Network Analysis (CNA).** CNA is a graph-theory-based rigidity analysis approach that determines global and local flexibility and rigidity characteristics of any protein from the trajectory of thermal unfolding simulations. By this method, relative differences between a structurally weak and a stable region of a protein can be detected with respect to thermal stability. We have used this method to predict the thermostability of proteins and to identify structural weak spots, that is, residues that upon mutation would improve a protein’s thermostability. In this approach, a protein is modeled as a constraint network where atoms are connected by sets of bars representing different covalent and noncovalent interactions. A rigidity analysis performed on this network results in the decomposition of the protein into rigid parts and flexible links. Different noncovalent interactions (salt bridges and hydrogen bonds) remain involved in maintaining the temperature-dependent structural stability of a protein. By analyzing a series of networks of such a protein and determining the involvement of those noncovalent interactions, we are able to correlate the stability retention or the rigidity loss with a thermal unfolding study. Throughout the temperature gradient MD simulation study, we have that the rigid network becomes almost flexible in mesophilic α-CA. Comparatively, we have observed that both thermophilic and mutant α-CA are prone to retain most of their interaction networks even at a high temperature. These phase transitions in all of the α-CA have been related to the thermodynamics of structural compactness provided by FEL.

Mutant residues Arg44 and Arg168 form additional salt bridge interactions with Glu226 in the newly engineered α-CA. These additional salt bridges are supposed to be responsible for the formation of a greater and rigid interaction network adjacent to those newly formed salt bridges in the mutant α-CA. This
Figure 6. continued
type of rigid interaction network is the main reason for acquiring increased thermal stability in both thermophilic and mutant α-CA.

PCA. PCA is used to determine the movement of the protein backbone during the MD simulation. The majority of the protein backbone motion is determined by the first two principal components. The protein backbone motion is plotted in two-dimensional covariance matrices as a function of principal component 1 (PC1) and principal component 2 (PC2).

Both eigenvalues (PC1 and PC2) of mesophilic Ng α-CA fluctuate between −50 and +50 Å at 300 K, whereas PC1 of Ssp α-CA ranges between −100 and +50 and PC2 from −50 to +50 Å. For the mutant, both PC1 and PC2 values range between −50 and +100 Å. The range of both PC1 and PC2 of mesophilic α-CA increases compared to thermophilic α-CA, which is (−150, +150) Å at 350 K. In the case of thermophilic α-CA, the fluctuations of the protein backbone along with the eigenvectors are less (PC1: −300, +250 Å and PC2: −200, +150 Å). In mutant α-CA, PC1 ranges between −270 and +300 Å, and PC2 ranges between −250 and +200 Å. For the mutant α-CA, PC1 ranges between −150 and +250 Å, and PC2 ranges between −200 and +75 Å. Similarly, at 500 K, larger flexibility along both eigenvectors (PC1: −35, +45 Å; PC2: −35, +30 Å) is shown by Ng α-CA. In the case of thermophilic α-CA, the fluctuations of the protein backbone along with the eigenvectors are less (PC1: −300, +250 Å and PC2: −200, +150 Å). In mutant α-CA, PC1 ranges between −270 and +300 Å, and PC2 ranges between −250 and +200 Å.

Essential Dynamics (ED). ED analyses during PCA have been performed to monitor the overall concerted motion of all of the three proteins, mesophilic, thermophilic, and mutant α-CA, at different temperatures. To examine the efficiency of the sampling of conformational space, we have calculated the cosine content of the first two principal components (PC1 and PC2) at all of the temperatures. In ED calculations, we have calculated the degree of backbone overlap of an individual protein at different temperatures. As mesophilic α-CA is comparatively unstable compared to thermophilic and mutant ones, it shows less percentage of overlap between the backbones of different conformations at various temperatures. Thermophilic and mutant α-CA show comparatively larger number of conformations with their backbone overlap. This pattern of ED reveals that thermophilic and mutant α-CA have smaller fluctuations at different local domains compared to mesophilic α-CA. In this analysis, it is also reported that the comparatively stable thermophilic and mutant

Figure 6. PCA for a. Ng α-CA, b. Ssp α-CA, and c. mutant Ng α-CA. PCA determines the movement of protein backbone during the MD simulation, which is extremely critical to understand protein unfolding. a. Ng α-CA: shows rapid increase in backbone movement with increasing temperature, b. Ssp α-CA: relatively less spreading throughout the simulations at all temperatures and c. mutant Ng α-CA: shows movement quite similar to Ssp-CA but increases significantly for the 500 K simulation.
Figure 7. continued
α-Ca proteins have their backbone motion mostly in the same direction for different conformations throughout the trajectories of temperature gradient, whereas the relatively less stable mesophilic α-CA shows residual backbone motion mostly in the opposite two directions for different conformations throughout the time scale of the temperature gradient MD simulation trajectory. The width of the ribbon in all of the three types of α-CA proteins represents the degree of backbone motion. The ribbon representation of the ED for all of these proteins shows different amplitudes of motion of a specific region of those proteins. From our study, it is clear that mesophilic α-CA has visible expansion and a larger dimension of the ribbon compared with thermophilic and mutant α-CA. The region of the largest β-sheet in the core hydrophobic β-twisted region and the surface β8 region in mesophilic α-CA shows a greater dispersion in backbone motion. Compared to that in this scenario, thermophilic and mutant α-CA show less amplitude of motion in those regions. This observation is also supported by the finding of the unfolding pathway.

From the comparative ED and PCA, we are able to correlate the role of stabilizing and destabilizing salt bridge-forming residues in maintaining thermal stability. There are some destabilizing salt bridge-forming residue pairs in mesophilic Ng α-CA: Lys102−Arg104, Arg136−Lys165, Lys165−Arg166, Arg166−Lys168, and Lys132−Arg136, which are mainly responsible for the larger domain-specific fluctuation compared to the thermophilic one. The corresponding residue pairs in the case of thermophilic α-CA are Glu99−Lys101, Glu133−Lys162, Lys162−Asp163, Asp163−Arg165, and Lys99−Glu133. Among them Glu99, Glu133, Asp163, and Arg165 are shown to form stabilizing salt bridges and assumed to have an impact on the small fluctuations of backbone atoms along the two principal eigenvectors. In the case of mutant α-CA, stabilizing salt bridge-forming residues Glu139 and Arg168 are responsible for the lower backbone atom fluctuation, as in thermophilic α-CA.

The above-mentioned comparative PCA values of mesophilic and thermophilic α-CA reveal that thermophilic and mutant Ssp α-CA have less backbone motion than their mesophilic counterpart. Less backbone motion and atomic degree of motion indicate rigid and stable conformations of thermophilic and mutant α-CA at higher temperatures (Figures 6 and S7).

**FEL.** The corresponding FEL as a function of the principal components describes the energy distribution of a protein-folding pathway. It depicts the stability of protein in terms of the Gibbs free energy and analyzes different conformational states as a function of energy for each residue involved.

\[ \Delta G = \sum -K T_B \ln(P_A - P_B) \]

Here, \( \Delta G \) defines the Gibbs Free energy and is a function of \( K \) and \( T_B \), which are the equilibrium constant and gas constant, respectively, \( P_A \) and \( P_B \) are the probabilities of the occurrence of the A conformation and B conformation, respectively, of a protein in its dynamics pathway.

We have analyzed FEL from the MD simulation trajectories of the simulated systems of the wild-type mesophilic, thermophilic, and mutant α-CA. Interactive 3D plots have been generated by using the Gibbs free energy as a function of two eigenvectors (PC1 and PC2), obtained from the principal component analysis. The principal component projected covariance matrix has been highlighted with a contour map, showing different coloring patterns at the bottom of each FEL plot. Each coloring pattern defines a range of energy distribution in which the protein shows its different configurations. Blue and red define the conformational spaces of a protein with minimum energy (stable state) and maximum energy (unstable state), respec-
tively. Intermediate color patterns highlight the other transient local energy states. In the case of \( \alpha \)-CA, we get different intermediate states for different temperature simulations. By detecting the intermediate states and the energy profile of residues corresponding to each state, we can detect the residues that are important for stabilizing those intermediate states and regulating the activity. In combination, FEL and conformation path sampling methods analyzed from MD simulation trajectories help us locate different important states of \( \alpha \)-CAs during their folding pathways, which have biological significance in maintaining thermal stability. A comparative analysis of FEL of thermophilic, mesophilic, and mutant \( \alpha \)-CA shows greater stability for thermophilic and mutant \( \alpha \)-CA.

A comparative analysis of FEL of different temperatures shows that the expansion of the energy funnel opening in Ng \( \alpha \)-CA is more than that of Ssp \( \alpha \)-CA and the mutant one. Ng \( \alpha \)-CA has a larger sampling for the two principal component values than that of Ssp \( \alpha \)-CA. Significant difference in the expansion of the FEL opening can be observed at a temperature of 350 K. At high temperatures, that is, 500 K, the differences in the principal component values become much higher between the mesophilic and thermophilic \( \alpha \)-CA. The number of intermediate states in Ng \( \alpha \)-CA is more than that in Ssp \( \alpha \)-CA with multiple lower free-energy barriers in the case of Ng \( \alpha \)-CA. Because of this, the surface of FEL for Ng \( \alpha \)-CA is more rugged than that of Ssp \( \alpha \)-CA. The higher number of free-energy barriers in the Ng \( \alpha \)-CA FEL than that in Ssp \( \alpha \)-CA and mutant \( \alpha \)-CA indicates that each local minimum of the landscape contains a stable conformation. Because of the presence of a large number of intermediate conformations, Ng \( \alpha \)-CA takes longer time scales to fold into its native state. In addition, the global minimum of free energy of Ng \( \alpha \)-CA is higher than that of the Ssp \( \alpha \)-CA and mutant \( \alpha \)-CA FEL. Another cause of the slow folding of the mesophilic Ng \( \alpha \)-CA protein is that this protein has to cross more number of larger energy barriers to transform from one intermediate conformation to the other. The RMSD and \( R_g \) value differences are significantly higher between the intermediate states of Ng \( \alpha \)-CA compared to those of Ssp \( \alpha \)-CA and mutant \( \alpha \)-CA. Because of this large number of intermediate states between any two global minima in the Ng \( \alpha \)-CA protein, it has to transverse a large energy barrier for proceeding along its folding pathway to achieve the native state. Analyzing different conformations of mesophilic Ng \( \alpha \)-CA from the FEL plots of increasing temperature gradient compared to thermophilic Ssp \( \alpha \)-CA and mutant \( \alpha \)-CA, it is clear that mesophilic Ng \( \alpha \)-CA is highly flexible and more unstable than its thermophilic counterpart and mutant \( \alpha \)-CA.

All of the FELs at different temperatures show a trend that the expansion of the opening of the landscape is much larger in the mesophilic protein than its thermophilic counterpart and the mutant one (Figure 7). The sampling of landscape becomes comparatively more with increasing temperature in the case of mesophilic protein. It is further evident as the mesophilic protein has more fluctuations in backbone motion and increased degree of freedom compared to others. Because of its greater unfolding rate and flexibility, the possibility of the occurrence of intermediate conformations is much more than other variants. The wild-type mesophilic \( \alpha \)-CA sampled wider regions on the FEL than thermophilic and mutant ones, significantly at 400 and 500 K. From 350 K onward to 500 K, the FELs of both thermophilic and mutant \( \alpha \)-CA show the presence of distinct basins separated by a relatively high free-energy barrier compared to the wild-type mesophilic \( \alpha \)-CA. At 350, 400, and 500 K, thermophilic and mutant \( \alpha \)-CA show a greater coverage of conformational space along the first two PCs compared to the wild-type mesophilic \( \alpha \)-CA. In contrast, the wild-type mesophilic \( \alpha \)-CA is found to sample overlapping regions of conformational space, wherein the conformations are largely clustered in regions along PC1 and PC2 due to the formation of weak hydrogen bonds and hydrophobic collapses. Mesophilic \( \alpha \)-CA shows large concerted motions in different parts of the trajectories, represented by the movements along PC1 and PC2 eigenvectors. In contrast, thermophilic and mutant \( \alpha \)-CA show restricted motions, significantly in the N-terminal region (\( \alpha \)-helix, twisted \( \beta \)-sheets in the protein hydrophobic core, and loops) and in the surface loops of the protein. Mesophilic \( \alpha \)-CA possesses a higher free energy of unfolding than thermophilic and mutant \( \alpha \)-CA. For this reason, thermophilic and mutant \( \alpha \)-CA show more thermolability than mesophilic \( \alpha \)-CA.

We have also characterized different states and their free-energy values. Different states of protein unfolding pathway, such as folded, partially unfolded, hydrophobic collapsed, and unfolded states, are labeled with 1, 2, 3, and 4 (major basins), respectively, in the energy contour map at the bottom of the FEL plot (Figure 7). Thermophilic and mutant \( \alpha \)-CA show the least conformational fluctuation and movement in native (folded states) or near-to-native basins, suggesting the existence of large free energy barriers between native and non-native states (partially unfolded states). In contrast, mesophilic \( \alpha \)-CA shows rapid transitions from native to non-native FEL basins, suggesting its increased tendency to sample conformationally distant non-native states at high temperatures. Notably, the wild-type mesophilic \( \alpha \)-CA shows a higher fluctuation than thermophilic and mutant \( \alpha \)-CA and follows a constricted transition pathway from native to non-native FEL basins (Figure 7).

Comparing the average RMSD values and analyzing the variation in hydrogen bonds and salt bridges between major FEL basins, it was evident that thermophilic and mutant \( \alpha \)-CA resist the breaking of hydrogen bonds and salt bridges and hence, remain close to the native conformations at high temperatures (Tables S2–S4).

**pK\(_a\) Calculation.** From pK\(_a\) calculations, we have found that Ssp \( \alpha \)-CA has a pK\(_a\) value (6.9) near 7 at a pH of 8, whereas mesophilic Ng \( \alpha \)-CA has a comparatively lower pK\(_a\) value of 6.7 at the same pH. This result implicates that the histidine (His64 in Ssp \( \alpha \)-CA) remains in a more deprotonated state relative to mesophilic Ng \( \alpha \)-CA His66. Thus, His64 from Ssp \( \alpha \)-CA is more capable of abstracting the proton from the water attached to the zinc ion. Therefore, the rate of nucleophilic attack on the carbon of CO\(_2\) by the deprotonated hydroxyl group (Zn—OH) is more in Ssp \( \alpha \)-CA. More interestingly, in the case of the newly designed mutant Ng \( \alpha \)-CA, we have calculated the pK\(_a\) value of His66 to be 6.8, higher than that of the wild-type histidine. From this, it can be assumed that mutant Ng \( \alpha \)-CA comes up with more enzymatic activity than the wild-type Ng \( \alpha \)-CA in addition to its thermal stability, as in thermophilic Ssp \( \alpha \)-CA.

**Dynamics Study of Zn—HisND1 Involved in the Proton Shuttle.** From the MD simulation trajectories, we have compared the conformations of the catalytic histidine residue (involved in the proton shuttle) and the Zn—HisND1 distance for all of the three systems: Ng \( \alpha \)-CA, Ssp \( \alpha \)-CA, and mutant Ng \( \alpha \)-CA. Considering the HisND1—Zn distance of up to 8 Å for the optimum catalytic activity of \( \alpha \)-CA, we have identified those “in” conformers that facilitate the proton shuttle. At 300 K, we have reported two “in” conformers of the catalytic histidine each.
The distances of HisND1−Zn for Ng α-CA for these two conformers are 7.4 and 7.5 Å, whereas in the case of Ssp α-CA, these distances are 7.2 and 7.4 Å, respectively. At 350 K, His66 of Ng α-CA shows a comparatively larger number of "in" conformations for time scales 10−50, 70, and 90 ns. The HisND1−Zn distances in Ng α-CA for those time scales at 350 K are 7.6, 7.7, 8.0, 7.2, 7.3, 7.9, and 6.7, respectively. At the same temperature, Ssp α-CA shows an "in" conformer of His64 only at the 90 ns time scale, which has a HisND1−Zn distance of 7.4.

### Table 2. Comparative HisND1−Zn Distances of Mesophilic Ng α-CA, Thermophilic Ssp α-CA and Mutant Ng α-CA

|        | 300 K      | 350 K      | 400 K      | 500 K      |
|--------|------------|------------|------------|------------|
|        | meso       | thermo     | mutant     | meso       | thermo     | mutant     | meso       | thermo     | mutant     | meso       | thermo     | mutant     |
| 10 ns  | 9.9        | 9.3        | 11.7       | 7.6        | 9.7        | 14.3       | 8.3        | 9.8        | 13.0       | 12.6       | 6.8        | 6.3        |
| 20 ns  | 10.2       | 7.2        | 12.3       | 7.7        | 8.5        | 13.9       | 6.4        | 8.2        | 11.0       | 9.8        | 6.7        | 8.8        |
| 30 ns  | 11.1       | 11.2       | 11.3       | 8.0        | 12.3       | 13.4       | 10.2       | 10.2       | 10.6       | 10.6       | 14.4       | 18.5       |
| 40 ns  | 12.0       | 13.0       | 10.5       | 7.2        | 11.3       | 13.7       | 10.6       | 8.7        | 7.5        | 16.9       | 12.9       | 7.5        |
| 50 ns  | 7.4        | 11.4       | 10.5       | 7.3        | 8.4        | 13.2       | 7.9        | 4.5        | 8.1        | 6.9        | 10.7       | 10.4       |
| 60 ns  | 8.4        | 7.4        | 11.6       | 10.1       | 8.1        | 13.9       | 7.6        | 8.5        | 7.7        | 15.1       | 16.4       | 13.2       |
| 70 ns  | 8.2        | 9.2        | 11.3       | 7.9        | 12.4       | 12.9       | 7.8        | 4.2        | 8.4        | 7.8        | 17.4       | 13.9       |
| 80 ns  | 7.5        | 11.9       | 10.7       | 11.5       | 8.1        | 14.0       | 7.4        | 4.0        | 9.0        | 10.1       | 13.6       | 12.6       |
| 90 ns  | 8.5        | 10.5       | 11.2       | 6.7        | 7.4        | 13.0       | 9.4        | 11.2       | 9.2        | 14.7       | 20.9       | 10.6       |
| 100 ns | 8.2        | 10.8       | 11.0       | 8.3        | 8.7        | 13.9       | 9.5        | 8.7        | 7.2        | 17.1       | 16.1       | 10.6       |

*All distances are in Angstrom (Å). HisND1−Zn distances ≤8 Å are in bold.*

Figure 8. Comparative denaturation temperature ($T_m$) of mesophilic Ng α-CA, thermophilic Ssp α-CA, and mutant Ng α-CA. Vertical red bar indicates $T_m$ for each of the system.
At 400 K, mesophilic Ng α-CA shows five “in” conformers of His66 within a HisND1–Zn distance of 8 Å at the time scale of 20 ns and from 50 to 80 ns. For thermophilic Ssp α-CA, we have reported three “in” conformers of His64 at the 50, 70, and 80 ns time scale. In these three conformers, the distances between Zn and HisND1 are much less, which are 4.5, 4.2, and 4.0 Å, respectively. The comparative HisND1–Zn distance between those “in” conformers of catalytic histidine of Ng α-CA and Ssp α-CA implies that less HisND1–Zn distance is more compatible for a fast proton transfer in the case of Ssp α-CA to attain its maximum enzymatic activity even at temperatures as high as 400 K. In the case of mutant Ng α-CA, three “in” conformers have been observed at 400 K, which have HisND1–Zn distances of 7.5, 7.7, and 7.2 at 40, 60, and 100 ns, respectively. At 500 K, mesophilic Ng α-CA has two “in” conformers of His66 with HisND1–Zn distances of 6.9 and 7.8 Å at 50 and 70 ns, respectively. Even for thermophilic Ssp α-CA, there are two “in” conformers of His64 with HisND1–Zn distances of 6.8 and 6.7 Å at 10 and 20 ns, respectively. At 500 K, the mutant Ng α-CA shows two “in” conformers of its His66 with HisND1–Zn distances of 6.3 and 7.5 Å at 20 and 40 ns, respectively. More interestingly, if we successively compare the HisND1–Zn distance of mutant Ng α-CA at 10 ns with the wild-type Ng α-CA HisND1–Zn distance at 50 ns and the same at 40 ns for mutant Ng α-CA and at 70 ns for wild-type Ng α-CA, then we can find that the distances are comparatively less for mutant Ng α-CA (Table 2). In addition, for thermophilic Ssp α-CA, the HisND1–Zn distances for the two “in” conformers of His64 are also less compared to those for mesophilic Ng α-CA. This quantitative analysis of HisND1–Zn distance and the conformation of catalytic histidine residues among Ng α-CA, Ssp α-CA, and mutant Ng α-CA imply that the rate of proton shuttle is faster in both Ssp α-CA and mutant Ng α-CA. Faster proton shuffling results higher enzymatic activity of thermophilic Ssp α-CA and mutant Ng α-CA compared to the wild-type mesophilic Ng α-CA at a high temperature of 500 K. The higher enzymatic activity of Ssp α-CA and mutant Ng α-CA compared to that of wild-type Ng α-CA is also supported by the relative pKₐ values that we have estimated later and is in good agreement with the analysis of the HisND1–Zn distance from the MD simulation study.

**Determination of the Melting Temperature (Tₘ)**. In this study, we have tried to complement the experimental evidence with one of the most efficient theoretical predictions. Here, we have theoretically calculated the denaturation temperature or Tₘ of mesophilic Ng α-CA, thermophilic Ssp α-CA, and mutant Ng α-CA, using the CNA. The method of Tₘ prediction can well mimic the experimental method of Circular Dichroism (CD), which determines the effects of mutations and ligands on protein and polypeptide stability with respect to the change in CD as a function of temperature. This also determines the enthalpy (ΔH) and entropy (ΔS) of unfolding, the midpoint of the unfolding transition (TM), and the free energy (ΔG) of unfolding (1). Likewise, in the Tₘ prediction method, a thermal unfolding simulation has been performed on each of the system. During the calculation of Tₘ, two other order parameters have also been considered, which are protein’s backbone rigidity and deformation energy for intraprotein hydrogen bonds. The unfolding transition calculated in CD is calibrated as a phase transition in the Tₘ prediction method (2). A comparative stability graph of the calibrated melting temperatures of the three proteins reveals that thermophilic Ssp α-CA has the highest Tₘ among all, and it is 348 K (75 °C), whereas its mesophilic counterpart Ng α-CA has the lowest Tₘ of 312 K (39 °C) compared to the other two protein systems. Mutant Ng α-CA has a Tₘ of 333 K (60 °C), close to that of thermophilic Ssp α-CA (Figure 8). The other two factors, rigidity and H-bond energy, calculated along with the melting temperature are also in good agreement with the comparative stability of the above three proteins. The melting temperature actually defines the moment when the rigidity of the protein backbone and the H-bond energy suddenly drop to a significantly low level. In the figure, the red bar indicates the melting temperature for individual proteins. It can be commented from the figure that mesophilic Ng α-CA starts to denature quickly than thermophilic Ssp α-CA and mutant Ng α-CA. Thermophilic Ssp α-CA is shown to retain its backbone rigidity the most, even at a high temperature of 348 K (75 °C). Mutant Ng α-CA is also shown to contain more number of high-energy H-bonds and a higher rigidity index up to a temperature of 333 K (60 °C), reflecting its greater stability than wild-type Ng α-CA.

## Conclusions

In the current study, comparative MD simulation has been used for understanding the molecular basis of different thermostabilities in mesophilic and thermophilic α-CA. Different stability comparing factors, such as RMSD, RMSF, Rs, SASA, hydrogen bonds, salt bridges, secondary structure content, unfolding pathway, PCA, CNA, ED and FEL, pKₐ, melting temperature, have been detected for both the mesophilic and thermophilic α-CA. Comparative analyses of those stability-determining factors of mesophilic and thermophilic α-CA guided us to develop a set of mutations in mesophilic α-CA in search for increased thermostability. Finally, we have designed to induce three mutations (S44R, S139E, and K168R) in mesophilic α-CA through a detailed analysis from sequence comparison, RMSF values, and detection of stabilizing and destabilizing salt bridges. All of the above-mentioned stability-determining factors have also been analyzed for mutant α-CA, and the results clearly show that the mutant becomes more thermostable than the wild type and becomes comparable to thermophilic α-CA with regard to stability. Salt bridge analysis shows that mutant α-CA evolves to form five new salt bridges that impart extra thermostability. In addition, the RMSD, RMSF, Rs, and SASA values for mutant α-CA become reduced compared to the mesophilic one at higher temperatures. PCA shows that mutant α-CA traverses through comparatively fewer conformational spaces than wild-type mesophilic α-CA, representing greater structural rigidity of mutant α-CA. The Comparative ED analysis of these three proteins explains the difference in thermal stability in accordance with PCA results. CNA also proves that the newly designed mutant α-CA has a more compact interaction network compared to the wild-type protein. The FEL analysis shows that wild-type mesophilic α-CA follows a constricted pathway from a native to a non-native state through less populated intermediate states. This actually supports the experimental observation regarding the irreversible folding-to-unfolding transition of wild-type mesophilic α-CA at higher temperatures. The comparative FEL analysis reveals that mutant α-CA has more number of native-like intermediate states with lower aggregation propensities compared to wild-type mesophilic α-CA. This finding explains the variation in stabilities between mesophilic and mutant α-CA, which is also in good agreement with other different analyses from the MD simulation trajectory. The newly designed mutant revealed that the improved stability has resulted from a reduced rate of protein.
unfolding and a decreased rate of precipitation of the unfolding intermediates. In contrast to mesophilic α-CA, mutant α-CA, like the thermophilic homologue, shows a deep and rugged FEL surface near the native state at 500 K. Throughout the analysis of different states of FEL, we have shown that mutant α-CA stays mostly in those conformations that retain stable salt bridges and hydrogen bonds. Such conformations help the mutant variant to remain stable even at extremely high temperatures. On the basis of the discussed results, we propose that mutant α-CA possesses both structural rigidity required to adapt in high temperatures (like the thermophilic homologue) and flexibility in catalytically important regions (like its originator, mesophilic α-CA) needed for optimum enzymatic activity at those high temperatures.

The in silico approach followed in this study provides new insights into protein engineering. Mutational study to understand or alter protein functions has been a very powerful tool for decades, but there is less correlation between the structural and functional roles of an amino acid. This frequently leads to the design of unstable proteins, resulting in poor or no expression of the mutant variants when treated experimentally. This clearly results in huge loss of manpower and money. The most interesting analyses of this study are the pK_a calculation and the determination of melting temperature (T_m), both of which actually compensate for the experimental study. The pK_a calculation shows that the designed mutant mesophilic α-CA is a more active enzyme than the wild-type one, and the T_m calculation confirms its thermal stability as being close to that of thermophilic α-CA. In future, this study would work as a platform not only to help further design thermally stable and industrially important proteins, which would be applicable in diverse biotechnological, pharmaceutical, and biomedical fields at relatively high temperatures, but also to provide a basis for designing mutants with better chance to be produced successfully under experimental conditions.

### MATERIALS AND METHODS

#### Basis of Mutant Construction of α-CA

To construct a thermostable mutant α-CA, we have gone through successive steps of analyzing the sequence, structure, and then detecting its stability using MD simulations of the mutant protein.36,37 We have first analyzed a number of comparative sequence and structural factors, such as number of proline and charged residues, total SASA, number of hydrogen bonds and salt bridges, for mesophilic and thermophilic α-CA.38 From their corresponding sequence alignments, we have found that the number of proline residues is greater in the mesophilic α-CA protein than that in the thermophilic one. Proline residues take part in secondary structure deformation and hence are mostly found in the loop region. Proline is also known for making the loops rigid. Structural analysis shows that thermophilic α-CA has all its proline residues in the loop region, whereas in its mesophilic counterpart, two proline residues are excluded from the loop, and thus, we have hypothesized that they might have some role in destabilizing the secondary structure contents. Thermophilic α-CA has more number of charged residues (Arg, Lys, Asp, and Glu) than its mesophilic homologue. In particular, there are more surface-oriented charged residues in the case of thermophilic α-CA compared to the mesophilic one.39 The number of intrachain hydrogen bonds and salt bridges are also greater in thermophilic α-CA. The total SASA is greater for mesophilic α-CA than thermophilic, which implies rapid unpacking of the core hydrophobic region of the mesophilic protein than its thermophilic counterpart. All of the above results support the enhanced stability of thermophilic α-CA over mesophilic α-CA (Table 3).

#### Table 3. Different Structural Factors of Protein Stability

| factors                           | mesophilic α-CA | thermophilic α-CA |
|-----------------------------------|-----------------|-------------------|
| total surface area                | 12 420 Å²       | 11 620 Å²         |
| number of hydrogen bond           | 487             | 526               |
| number of salt bridges            | 4 (all confined in a local region) | 5 (dispersed in different places) |
| number of proline residues        | 15 (13 in loop)  | 11 (all in loop)  |
| charged residues                  | 44              | 62                |

From the above sequence and structure-based analysis, we accumulate some ideas about the significant flexible region in the mesophilic protein compared to its thermophilic homologue, where we can introduce mutation to enhance its stability. We have first used the RMSF values obtained from the MD simulation to identify these types of flexible regions in mesophilic proteins.40 From the sequence alignment between mesophilic and thermophilic α-CA, we have then further identified the residues in the mesophilic protein, which correspond to the salt bridge-forming residues in the thermophilic one. Then, we have selected the surrounding residues (8 Å radius) of those individual mesophilic residues and have further compared the RMSF values of all of those residues with their thermophilic counterpart. We have found some residues of mesophilic α-CA showing high RMSF values at all of the temperatures as compared to the thermophilic one. Residues of mesophilic α-CA that show approximately two to four times greater RMSF than the corresponding thermophilic α-CA residues are Ser44, Asn134, Arg136, Ser139, Thr141, Pro164, Arg166, Lys168, Tyr169, Arg171, Leu188, and Tyr193. In addition to that, in thermophilic α-CA there is a compact network between the salt bridge-forming residues, such as Lys36, Lys39, Lys41, Glu99, Arg165, and Glu223. Structural analysis also revealed that the main differences between wild-type mesophilic α-CA and its thermophilic homologue are the organization of salt bridges involved in the above-mentioned surface residues. The corresponding mesophilic α-CA residues in this region are Glu41, Ser44, Lys102, Lys168, and Glu226, which show comparatively increased RMSF value at all temperatures.81,42 Increased RMSF values might also be a reason for not forming any salt bridges in this region in the case of mesophilic α-CA. We have also found some destabilizing salt bridges in the mesophilic protein, which are Lys102–Arg104, Arg136–Lys165, Lys165–Arg166, Arg166–Lys168, and Lys132–Arg136. The corresponding residues in the thermophilic protein for these positions are interestingly a pair of two oppositely charged residues. In thermophilic α-CA, these pairs of residues are Glu99–Lys101, Glu133–Lys162, Lys162–Asp163, Asp163–Arg165, and Lys129–Glu133 (Table 4).43 All of these residues of thermophilic α-CA show significantly low RMSF values than the corresponding mesophilic α-CA residues (Table 5). Glu99, Glu133, Asp163, and Arg165 from those pairs of oppositely charged residues are shown to form salt bridges in thermophilic α-CA.

According to the significant changes in RMSF values and destabilizing salt bridges in the mesophilic protein, we have introduced some mutations. Our first choice of mutation is S44R because we have found that the residue Serine in this position of mesophilic α-CA does not form any salt bridge,
whereas at the same position, Lysine form a salt bridge with Glu223 in thermophilic α-CA. Moreover, the RMSF value of Ser44 is about double in mesophilic α-CA compared to Lys41 of thermophilic α-CA because of the absence of the salt bridge. By mutating Ser44 to arginine, we want to develop maximum fluctuation of thermophilic residues corresponding to its mesophilic residues at the respective temperatures.

### System-Setup

The crystal structures of two homologous α-CA from mesophilic N. gonorrhoeae (PDB ID: 1KOQ) and thermophilic Sulfurohydrogenibium sp. (PDB ID: 4G7A) were used as starting models for the MD simulations. Crystal structures of mesophilic and thermophilic α-CA have resolutions of 1.80 and 1.90 Å and the B-factors of 22.94 and 17 Å² with residue numbers 2-225 and S-226, respectively. We have performed MD simulations of both wild-type and mutant α-CAs. We have introduced three mutations into mesophilic Ng α-CA (S44R, S139E, and K168R) by using the COOT software. We have then performed MD simulations of mutant protein and have shown comparative thermal stabilities for both wild-type α-CAs.

### MD Simulation

We have used GROMACS 4.5 and OPLS-AA all atom force field, for MD simulations, on an Intel Xeon Quad Core W3530 2.8 GHz Processor with LINUX environment. Mesophilic and thermophilic α-CA were solvated in a cubic box (dimension with 74.8 × 74.8 × 74.8 and 74.6 × 74.6 × 74.6, respectively) filled with SPC216 water molecules (12 655 and 12 604, respectively). To simulate the solvated system in a neutral pH, we have replaced solvent molecules with counter ions (5 Cl⁻ for mesophilic α-CA and 11 Cl⁻ for thermophilic α-CA).

The steepest descent method of energy minimization was applied for all of the systems for an iteration cycle of maximum 50 000 steps. The minimized systems were then equilibrated at four different temperatures (300, 350, 400, and 500 K) to relax the system and maintain the constant temperature and pressure of the system. A production run of 100 ns was performed. Periodic boundary conditions were applied under isothermal and isobaric conditions, using Berendsen coupling algorithm with relaxation times of 0.1 and 0.2 ps, respectively. The LINCS algorithm was used to constrain bond lengths using a time step of 2 fs for both systems.

### Table 4. Destabilizing Salt Bridge

| Destabilizing salt bridge pair in mesophilic α-CA | Corresponding residues pair in thermophilic α-CA |
|--------------------------------------------------|--------------------------------------------------|
| Lys102–Arg104 | Glu99–Lys101 |
| Arg136–Lys165 | Glu133–Lys162 |
| Lys165–Arg166 | Lys162–Asp163 |
| Arg166–Lys168 | Asp163–Arg165 |
| Lys132–Arg136 | Lys129–Glu133 |

“Residues in bold form a salt bridge in both proteins.”

### Table 5. Comparative RMSF of Residues Forming Destabilizing Salt Bridge and the Corresponding Thermophilic Residues

| Mesophilic α-CA (M) | Thermophilic α-CA (T) | 300 K | 350 K | 400 K | 500 K |
|---------------------|-----------------------|-------|-------|-------|-------|
| Lys102              | Glu99                 | 0.1293| 0.1388| 0.1632| 0.1948| 0.2436| 0.4467| 0.3085|
| Arg104              | Lys101                | 0.0778| 0.1777| 0.1173| 0.2029| 0.1729| 0.2314| 0.3876| 0.2895|
| Lys132              | Lys129                | 0.2391| 0.1812| 0.2182| 0.1896| 0.2709| 0.198 | 0.5366| 0.329 |
| Arg136              | Glu133                | 0.2914| 0.086 | 0.2896| 0.0933| 0.3214| 0.1142| 0.7206| 0.3868|
| Lys165              | Lys162                | 0.2519| 0.1952| 0.2491| 0.2053| 0.2403| 0.2226| 0.7082| 0.3132|
| Arg166              | Asp163                | 0.3193| 0.0943| 0.2513| 0.1149| 0.2589| 0.1266| 0.6994| 0.2229|
| Lys168              | Arg165                | 0.1765| 0.0753| 0.1638| 0.166 | 0.1747| 0.2322| 0.5154| 0.2847|

“RMSF values in bold represent comparatively less fluctuation of thermophilic residues corresponding to its mesophilic residues at the respective temperature.”
were calculated using the Particle Mesh Ewald method, and van der Waals and Coulombic interactions were calculated with a cutoff at 1.0 nm. The tools provided by the GROMACS package were utilized to analyze different MD trajectories. Secondary structure analyses were performed using the program DSSP. We used a web-based server 2Struc that provides detailed information regarding secondary structures, including α helices, 3α helices, β strands, β bridges, β turns, and bends. Web server ESBRi was used for salt bridge calculation and PyMOL. VMD, COOT was used as the visualizing software. PyMOL, Xmgrace, and MATLAB programs were used to analyze and prepare the figures.

We have analyzed the key order parameters that determine the molecular basis of the thermal stability for each of the wild-type mesophilic α-CA, thermophilic α-CA, and mutant mesophilic α-CA. RMSD, RMSF, \( R_g \) and SASA have been calculated and analyzed for the determination of thermal stability and their differences.

The salt bridges in both homologues of α-CA protein were calculated using a web-based tool ESBRi using a cutoff value of 4.0 Å for salt bridge lengths. By comparing the salt bridges in both mesophilic and thermophilic α-CA, we identified some extra salt bridges, which maintain greater stability in thermophilic α-CA, and replaced the corresponding residues in mesophilic α-CA without perturbing the salt bridges already present in wild-type mesophilic α-CA. After introducing those residues, we have found that the initial structure of the mutant mesophilic α-CA gained some extra salt bridges. In this way, we have designed a mutant mesophilic α-CA containing three mutations (S44R, S139E, and K168R) (Figure 6). Then, we have found that the initial structure of the mutant mesophilic α-CA gained some extra salt bridges. In this way, we have designed a mutant mesophilic α-CA containing three mutations (S44R, S139E, and K168R) (Figure 6). Then, we have investigated the increasing thermostability due to the newly originated salt bridges in mutant α-CA. We have also analyzed the variation of salt bridge length in all of the α-CA systems as a function of different temperatures. The length variation of salt bridges was analyzed with respect to the average PDB at a given temperature. Improvement of the side chain conformation of each residue from all of the average PDBs was performed using Scwrmr4.

The covariance matrices of the positional fluctuations of \( C_{\alpha} \) atoms were analyzed with PCA or ED. The first two principal components (PC1, PC2) contribute to the major protein backbone motion and have been used for the PCA. The covariance matrix was diagonalized to obtain the eigenvectors and eigenvalues that provide information about correlated motions throughout the protein. The FEL prepared using MATLAB describes the energy distribution and biological interrelation between thermodynamics and protein stability.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.6b00041.

Additional discussion, supplementary figures and tables (PDF)

- Dynamics of mesophilic NgCA (AVI)
- Dynamics of thermophilic Ssp-CA (AVI)
- Dynamics of mutant NgCA (AVI)

**AUTHOR INFORMATION**

Corresponding Authors

*E-mail: sdattfbt@iitr.ac.in, saugata.iitk@gmail.com (S.H.).

**Author Contributions**

S.K.B., M.H., and M.P. contributed equally.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

A part of the work was conducted in the context of the SFI/HEA Irish Centre for High-End Computing (ICHEC), IIT, IIT-Roorkee, Bioinformatics Facility, Department of Biotechnology, IIT-Roorkee, Bioinformatics Resources and Applications Facility (BRAF), C-DAC, Pune for the provision of computational facilities and support.

**REFERENCES**

1. Retooling of Coal-Fired Power Plants for CO₂ Emissions Reductions; MIT Energy Initiative Massachusetts Institute of Technology, 23rd March, 2009. [http://news.mit.edu/2009/whitewater.pdf](http://news.mit.edu/2009/whitewater.pdf).

2. Yu, C. A. Review of CO₂ Capture by Absorption and Adsorption; Aerosol and Air Quality Research, 2012; Vol. 12, pp 745–769.

3. Advanced Post-Combustion CO₂ Capture, 1st ed., April 2009. [https://sequestration.mit.edu/pdf/Advanced_Post_Combustion_CO2_Capture.pdf](https://sequestration.mit.edu/pdf/Advanced_Post_Combustion_CO2_Capture.pdf).

4. Kanbar, B. Enzymatic CO₂ Sequestration by Carbonic Anhydrase, 1st ed., 2008. [http://openaccess.itye.edu/bitstream/handle/11147/3957/T000767.pdf?sequence=1](http://openaccess.itye.edu/bitstream/handle/11147/3957/T000767.pdf).

5. Supuran, C. Carbonic Anhydrases An Overview. Curr. Pharm. Des. 2008, 14, 603–614.

6. Domnic, J.; McKenna, R. Sequestration Of Carbon Dioxide By The Hydrophobic Pocket Of The Carbonic Anhydrases. Biochim. Biophys. Acta, Proteins Proteolectomics 2010, 1804, 326–331.

7. Tripp, B.; Smith, K.; Ferry, J. Carbonic Anhydrase: New Insights For An Ancient Enzyme. J. Biol. Chem. 2001, 276, 48615–48618.

8. Karali, T.; Silverman, D. Kinetics Of The Hydration Of CO₂ Catalyzed By Carbonic Anhydrase III From Skeletal Muscle Of The Cat. Ann. N.Y. Acad. Sci. 1984, 429, 129–136.

9. Wells, J.; Kandel, S.; Koenig, S. The pH Dependence Of Solvent Proton Relaxation In Carbonic Anhydrase Solutions: Paramagnetic And Diamagnetic Effects. Biochemistry 1979, 18, 1989–1995.

10. Villauer, F.; Swietach, P.; Patar, S.; Harris, A.; Vaughan-Jones, R. Comparison Of Ph-Dependence Of Carbonic Anhydrase Activity In Vitro And In Living Cells. Biophys. J. 2009, 96, 625a.

11. Peirce, S.; Russo, M.; De Luca, V.; Capasso, C.; Rossi, M.; Olivieri, G.; Salatino, P.; Marzocchella, A. Immobilization Of Carbonic Anhydrase For Biomimetic CO2 Capture In Slurry Absorber. New Biotechnol. 2014, 31, S20–S21.

12. Research and Development Goals for CO₂ Capture Technology; U.S. Department of Energy: Washington, DC, 2011. [https://www.netl.doe.gov/energy-analyses/temp/ResearchAndDevelopmentGoalsforCO2CaptureTechnology_120111.pdf](https://www.netl.doe.gov/energy-analyses/temp/ResearchAndDevelopmentGoalsforCO2CaptureTechnology_120111.pdf).

13. Krishnamurthy, V.; Kaufman, G.; Urbach, A.; Gittin, L.; Gadiksen, K.; Weibel, D.; Whitesides, G. Carbonic Anhydrase As A Model For Biophysical And Physical-Orgnic Studies Of Proteins And Protein-Ligand Binding. Chem. Rev. 2008, 108, 946–1051.

14. Smith, K. Prokaryotic Carbonic Anhydrases. FEMS Microbiol. Rev. 2000, 24, 335–366.

15. Smith, K.; Jakubzick, C.; Whittam, T.; Ferry, J. Carbonic Anhydrase Is An Ancient Enzyme Widespread In Prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 15184–15189.

16. Capasso, C.; Supuran, C. An Overview Of The Alpha, Beta- And Gamma-Carbonic Anhydrases From Bacteria: Can Bacterial Carbonic Anhydrases Shed New Light On Evolution Of Bacteria? J. Enzyme Inhib. Med. Chem. 2015, 30, 325–332.

17. Liljas, A.; Laurberg, M. A Wheel Invented Three Times. EMBO Rep. 2000, 1, 16–17.
(56) Costantini, S.; Colonna, G.; Facchiano, A. ESBRI: A Web Server For Evaluating Salt Bridges In Proteins. *Bioinformation* 2008, 3, 137–138.

(57) DeLano, W. *The Pymol Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002. http://www.pymol.org.

(58) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera: A Visualization System For Exploratory Research And Analysis. *J. Comput. Chem.* 2004, 25, 1605–1612.

(59) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 1996, 14, 33–38.

(60) Vaught, A. Graphing with Gnuplot and Xmgr: Two graphing packages available under Linux. *Linux J.* 1996, 1996, 7. http://dl.acm.org/citation.cfm?Id=326334.

(61) MATLAB R2014b; The MathWorks Inc.: Natick, MA, 2014.

(62) Paul, M.; Hazra, M.; Barman, A.; Hazra, S. Comparative Molecular Dynamics Simulation Studies For Determining Factors Contributing To The Thermostability Of Chemotaxis Protein “CheY”. *J. Biomol. Struct. Dyn.* 2014, 32, 928–949.

(63) Krivov, G.; Shapovalov, M.; Dunbrack, R. Improved Prediction Of Protein Side-Chain Conformations With SCWRL4. *Proteins* 2009, 77, 778–795.

(64) Amadei, A.; Ceruso, M.; Di Nola, A. On The Convergence Of The Conformational Coordinates Basis Set Obtained By The Essential Dynamics Analysis Of Proteins’ Molecular Dynamics Simulations. *Proteins: Struct., Funct., Bioinf.* 1999, 36, 419–424.

(65) Amadei, A.; Linssen, A.; Berendsen, H. Essential Dynamics Of Proteins. *Proteins: Struct., Funct., Bioinf.* 1993, 17, 412–425.