Rational Design of Thermostable Carbonic Anhydrase Mutants using Molecular Dynamics Simulations

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	jp-2018-05926u.R1
Manuscript Type:	Article
Date Submitted by the Author:	15-Aug-2018

Complete List of Authors:	Parra-Cruz, Ricardo; University of Nottingham - Malaysia Campus, Chemical and Environmental Engineering Jäger, Christof; University of Nottingham, Department of Chemical and Environmental Engineering Lau, Phei Li; The University of Nottingham Malaysia Campus, Chemical and Environmental Engineering Gomes, Rachel; University of Nottingham, Faculty of Engineering Pordea, Anca; University of Nottingham, Chemical and Environmental Engineering

Rational Design of Thermostable Carbonic Anhydrase Mutants using Molecular Dynamics Simulations

Ricardo Parra-Cruz,[†] *Christof M. Jäger,*[‡] *Phei Li Lau,*[†] *Rachel L. Gomes,*[‡] *Anca Pordea*[‡]*

[†]Department of Chemical and Environmental Engineering, University of Nottingham Malaysia Campus, Semenyih, Malaysia

[‡]Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

ABSTRACT: The stability of enzymes is critical for their application in industrial processes, which generally require different conditions from the natural enzyme environment. Both rational and random protein engineering approaches have been used to increase stability, with the latter requiring extensive experimental effort for the screening of variants. Moreover, some general rules addressing the molecular origin of protein thermostability have been established. Herein, we demonstrate the use of molecular dynamics simulations to gain molecular level understanding of protein thermostability and to engineer stabilizing mutations. Carbonic anhydrase (CA) is an enzyme with high potential for biotechnological carbon capture applications, provided it can be engineered to withstand the high temperature process environments, inevitable in most gas treatment units. In this study, we used molecular dynamics simulations at 343 K, 353 K and 363 K, to study the relationship between structure flexibility and thermostability in bacterial α-CAs, and applied this knowledge to the design of mutants with increased stability. The most thermostable α-CA known, TaCA from *Thermovibrio ammonificans*, had the most rigid structure during MD simulations, but also showed regions with high flexibility. The most flexible amino acids in these regions were identified from RMSF studies, and stabilizing point mutations were predicted based on their capacity to improve the calculated free energy of unfolding. Disulfide bonds were also designed at sites with suitable geometries, and were selected based on their location at flexible sites, assessed by B-factor calculation. Molecular dynamics simulations allowed the identification of five mutants with lower RMSF of the overall structure at 400 K, compared to wild-type TaCA. Comparison of free energy landscapes between wild-type TaCA and the most promising mutants, Pro165Cys-Gln170Cys and Asn140Gly, showed an increased conformational stability of the mutants at 400 K.

INTRODUCTION

Carbonic anhydrase (CA; EC 4.2.1.1) catalyzes the reversible hydration of carbon dioxide into hydrogen carbonate and a proton, with one of the fastest reaction rates known so far in nature. Enzymes in this class are encoded by six evolutionarily unrelated gene families (α, β, γ, δ, ξ and η), which all have a divalent metal-dependent mechanism, generally Zn-based, as a common feature.[1] One of the most studied CA families are the α-CAs, in which the Zn^{2+} ion, typically bound by three histidines, activates a hydroxyl group to perform the nucleophilic attack on CO_2. The resulting Zn-bound hydrogen carbonate is displaced by water, which is in turn deprotonated *via* a proton shuttle mechanism assisted by a neighboring histidine, to re-form the Zn-bound hydroxyl.[2]

Recently, the use of CAs as potential catalysts for the absorption of CO_2 from process gases has been investigated.[3] CO_2 absorbers employ either inorganic or organic bases in aqueous solvents, to drive the equilibrium by neutralizing the proton released during the hydrolysis, and amine-based solvents have shown particular promise.[4] In such processes, regeneration of the amine is problematic due to the high energy required to desorb CO_2. Amines with low heat of desorption have slow CO_2 capture kinetics, and CAs have been employed to accelerate the CO_2 capture step. Given the high pH and temperature conditions in the absorption – stripping processes, the stability of the CA is crucial for its application at large scale. Two approaches have been used for the development of stable CAs: isolation and characterization of enzymes from thermophilic bacteria,[5] and stability engineering by directed evolution under high temperature and pH conditions.[6]

Many research efforts are focused on the development of thermostable α-CAs, which have relatively short sequences compared to the other families, since they show the highest activities

and long-term stability. Four thermostable α-CAs were characterized so far: SspCA from *Sulfurihydrogenibium yellowstonense*, with an optimum temperature of 90 °C, a half-life of 53 days at 40 °C and k_{cat} = 9.35 x 10^5 s^{-1};[7] SazCA from *Sulfurihydrogenibium azorense*, with the highest turnover rate amongst thermostable α-CAs, k_{cat} = 4.40 x 10^6 s^{-1}, and with a similar optimum temperature to SspCA but lower stability;[8] TaCA from *Thermovibrio ammonificans*, with k_{cat} = 1.60 x 10^6 s^{-1} and the highest long-term stability of 152 days at 40 °C[9] and PmCA from *Persephonella marina*, with k_{cat} = 3.20 x 10^5 s^{-1} and a half-life of 75 days at 40 °C.[10] The former three enzymes have been structurally characterized.[11-13] Inspection of the SspCA crystal structure shows a high content of secondary structure, of charged residues and of ionic networks, all of these being consistent with features of thermostable proteins.[14] On the other hand, TaCA has fewer ion pairs, but it was suggested to form two intermolecular disulfide bonds, promoting the association of two dimers into a tetramer, which in turn triggers the formation of additional four ion pairs. These features were suggested to be responsible for the high thermostability of TaCA.[13]

Common protein stabilization principles are known, and generally rely on rigidifying features such as salt bridges, disulfide bonds and hydrophobic interactions. However, the best strategies for stability engineering differ between protein families, and thus are difficult to predict, making rational design of protein stability a non-trivial task. Flexible regions can be identified by structure-based B-factor analysis or molecular dynamics (MD) simulations.[15] Computational design of stabilizing point mutations at these hot spots, coupled with disulfide design at sites with suitable geometry and energy can increase the chances of finding stabilizing mutations with a relatively small screening effort.[16] Subsequent stability assessment by MD simulations can further reduce the number of variants to be screened for thermostability.

In the α-CA family, mesostable CAs have already been engineered for increased thermostability using the above strategies. For example, the thermostability of CA from *Neisseria gonorrhoeae* (NgCA) was 8-fold increased by engineering disulfides at the protein surface.[17] In another study, comparative MD simulations between NgCA and SspCA were used to identify stabilizing salt bridges in the thermostable scaffold, which were introduced at similar positions in the mesostable protein sequence.[18] Analysis of the stability-determining factors of the two wild-type proteins and of the newly designed Ser44Arg-Ser139Glu-Lys168Arg NgCA mutant showed an improved stability of the latter, suggested by the lower flexibility, lower solvent accessible surface area and increased propensity of stable conformations in the Free Energy Landscape (FEL) analysis during the simulations at high temperatures (500 K). Nonetheless, given that both catalytic efficiencies and stabilities of thermostable variants are higher compared to their mesophilic counterparts (for example, k_{cat}/K_M = 5.4 x 10^7 M^{-1} s^{-1} for NgCA and = 1.1 x 10^8 M^{-1} s^{-1} for TaCA), a better strategy would be to utilise thermostable enzymes as starting points for further improvement of their operation at high temperature.

Since thermostable enzymes are more rigid than their mesostable counterparts, the design of stabilizing mutations based on flexibility analysis is more challenging using these starting points, and a common approach for stability optimization relies on random mutagenesis, which involves laborious screening efforts. The only example of successful α-CA stability engineering starting from a thermostable scaffold has been the directed evolution of TaCA, which resulted in up to 3-fold improvement in half-life compared to the wild-type variant, using mutations in the *N*-terminal region of the protein.[19] Directed evolution methods have also led to impressive 4 x 10^6-fold stability improvement in the case of β-CA from *Desulfovibrio vulgaris*.[6] A better

understanding of the structural features responsible for thermostability in α-CAs is needed, in order to design mutants with improved stability, by using the minimum number of mutations.

The aim of this study is to develop a strategy for the rational design of stabilizing point mutations in α-CAs, using thermostable scaffolds as a starting point. Comparative molecular dynamics simulations at three different temperatures were used to identify rigid and flexible regions in mesostable and thermostable α-CAs, and to assess whether structure rigidity corresponds to experimentally determined stability in the case of these enzymes. Relevant thermosensitive regions from the most rigid protein structure determined from the simulations were targeted by point mutations that would stabilize the structure even further. MD simulations were then used to confirm the increased rigidity, which was associated with the stabilization of the structure.

METHODS

Protein Sequences and Structures. Protein sequences and crystal structures were retrieved from the RSCB protein data bank, as follows: hCAII (3CAJ[20]); NgCA (1KOQ[21]); TaCA (4C3T[13]); SspCA (4G7A[12]); SazCA (4X5S[11]). The amino acid numbering from the published structures was used. Sequence alignments were performed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).[22] The initial signal peptide (around 20 amino acids at the *N*-terminus) was not present in any of the sequences except for TaCA, where it was not built in the crystal structure owing to disorder, and was therefore not included in the simulations. Salt bridges between negative (aspartate, glutamate) and positive (arginine, lysine, histidine) side chain functionalities were evaluated using the ESBRI server with a cut-off at 4 Å (http://bioinformatica.isa.cnr.it/ESBRI/introduction.html).[23]

Molecular Dynamics Simulations and Analyses. All simulations were performed using GROMACS 5.1.2[24] with CHARMM36 force field[25] and the SPC/E water model.[26] The protonation states of amino acid side chains were adjusted to normal pH 7 conditions. The H++ server was used to prepare the series of protonation states.[27] The output files from the H++ server were used to determine the optimal setting of protonation and deprotonation states of Lys, Arg, Asp, Glu and His residues, which were modified using the pdb2gmx command in GROMACS. All the crystal structures were prepared before simulation, by removing all solvent atoms, such as water, oxygen and solvent molecules, which were not considered part of the protein structure or else relevant for the purpose of the study. Structural representations such as macromolecule dynamics were visualized with VMD[28] and PyMOL. Post analysis of the molecular dynamics data was graphically represented with XmGrace or Matlab packages.

The protein structures were inserted in a triclinic geometry box of water molecules with a minimum distance of 1.2 nm between the protein and the box boundaries. The systems were neutralized by adding Na^+ or Cl^- ions. Long range electrostatic interactions were modelled using the particle-mesh Ewald method,[29] with a cutoff of 0.8 nm, a Fourier spacing of 0.12 nm and cubic spline interpolation. Non-bonding van der Waals interactions were applied with a twin range potential of 0.8 nm and 1.4 nm.[30] All bonds involving hydrogen atoms were constrained with the LINCS algorithm.[31] Energy minimization (with 2000 steps, by using the steepest descent algorithm), temperature, and pressure equilibration were applied with position restraints on the protein. The overall system was integrated to a temperature isotherm using Nose-Hoover thermostat with a speed of 0.002 ps.[32] Initial velocities applied for the different temperatures (343 K, 353 K, 363 K, 400 K) were represented by Maxwell Boltzmann distribution. The

pressure of the system was isotropically coupled to a barostat of 1 bar by Parrinello – Rahman calculations with a temperature coupling constant of 2 ps.[33] All simulations were run for 100 ns.

The analysis of simulation trajectories, including principal component analysis and free energy landscapes, is described in detail in the Electronic Supplementary Information. The root mean square deviation (RMSD) values were calculated using GROMACS embedded tools.[34] These were based on standard RMSD calculations following rigid body translation and rotation for structural superposition in Cartesian space, minimizing the resulting RMSD value. Root mean square fluctuation (RMSF), radius of gyration (*R*g), solvent accessible surface area (SASA), hydrogen bonds and salt bridges (< 0.4 nm) were calculated using tools within the GROMACS simulation package.

Highly flexible regions were identified from the RMSF analysis, as follows (see also Table S2). First, for each isotherm simulation (343 K, 353 K, and 363 K), residues were selected, with an RMSF higher than the true average limit μ, where $\mu = \bar{x} + s$, with $\bar{x}$ = mean of RMSF per isotherm and s = standard deviation of RMSF per isotherm. For every sequence of consecutive amino acids, residues with the highest RMSF were selected as representatives, and this was performed separately at 343 K, 353 K and 363 K. Residues that were selected as representatives for at least two isotherm simulations were considered as highly flexible amino acids. Amino acids at the *N*-terminus are naturally flexible, and therefore were excluded from this selection.

Design of Stabilizing Mutations. Point mutations in TaCA were constructed using Yasara.[35] The relative changes in the folding free energies ($\Delta\Delta G^{Fold}$) due to point mutations were calculated by FoldX (foldx.crg.es)[36] using the wild-type structure of TaCA as a reference ($\Delta\Delta G^{Fold} = \Delta G^{Fold}_{mutation} - \Delta G^{Fold}_{wild\text{-}type}$). The standard settings of the software were used (T = 343 K, 353 K, and 363 K; pH = 8.0, ionic strength = 0.05 M). Mutations were evaluated as

stabilizing if $\Delta\Delta G^{Fold} < -2$ kJ mol^{-1}, neutral if -2 kJ mol^{-1} $< \Delta\Delta G^{Fold} <$ 2 kJ mol^{-1} and destabilizing if $\Delta\Delta G^{Fold} > 2$ kJ mol^{-1}.[36-37]

Disulfide bonds were predicted with the Disulfide by Design 2.0 web-based platform,[38] which includes residue proximity and geometry, the highest sum of B-factors and the location of the residue as criteria for pair selection. Visual inspection allowed further refinement of this selection, and exclusion of candidates disrupting high-consensus residues, catalytic site and existing salt bridges. Out of 36 possible pairs identified with DbD 2.0, three disulfide bonds were selected for further study, due to their location in high flexibility regions and/or to their favorable predicted bonding energy and dihedral angle (see Table S5 for details of the selection process).

RESULTS AND DISCUSSION

The sequences and structures of mesophilic and thermophilic α-CAs were compared to identify consensus and non-consensus regions, as well as common stabilizing features such as salt bridges and disulfide bonds. Molecular dynamics simulations were then performed at three different temperatures, to understand which regions of the protein are the most flexible. Stabilizing mutations were subsequently designed in these regions and their ability to reduce flexibility was analyzed by molecular dynamics.

Sequence and Structural Analysis of α-CAs. Six α-carbonic anhydrases from different organisms were initially compared by sequence alignment and by inspection of their crystal structures (Figure S1): four from thermophiles (SspCA, SazCA, TaCA and PmCA, the latter has no published crystal structure), one from a mesophile (NgCA) and one eukaryotic (hCAII). They showed similar sequence identities of 45-65%, whilst a lower identity was observed with the

eukaryotic hCAII (33-42%), which contains three additional large insertions at positions 98-103, 126-136 and 230-239 (hCAII numbering)[12]. The majority of secondary structures is conserved throughout all CAs. Non-consensus regions were identified at positions 76-81, 93-99 and 165-182 (TaCA numbering). Evaluation of the ionic networks (4 Å cut-off) by inspection of the crystal structures showed an increased number of intra-monomer ionic interactions for the α-CAs from thermophiles (10-13 contacts for the thermostable enzymes, compared to 8-9 contacts for NgCA and hCAII). The three thermostable α-CA structures showed a glutamate-lysine interaction between Glu156 and Lys159 (TaCA numbering; corresponding to Glu133-Lys136 in SspCA and SazCA), which was absent from NgCA and hCAII, where the glutamate was replaced by arginine and glycine, respectively. In TaCA, these two residues are part of a further ionic network with Asp179 and Lys182, connecting helices η4/α2 and η6 (Figure 1; secondary structures are numbered according to hCAII, see Figure S1). Another common ionic interaction was between His119 and Glu129 (TaCA numbering), which is present in all thermostable CAs and hCAII, but is absent from NgCA, where the histidine is replaced by asparagine. Whilst in NgCA and hCAII most ionic interactions occur between pairs of amino acids, all three thermostable CAs exhibit clusters of charged amino acids, involving ionic networks: Asp102-Arg192-Arg207 in TaCA, Lys39-Lys41-Arg165-Glu223 in SspCA (Lys39-His41-Glu223 in SazCA), and Asp73-Lys75-Glu82-His84 in SazCA (Figure 1). These ionic networks have previously been suggested to increase thermostability.[12]

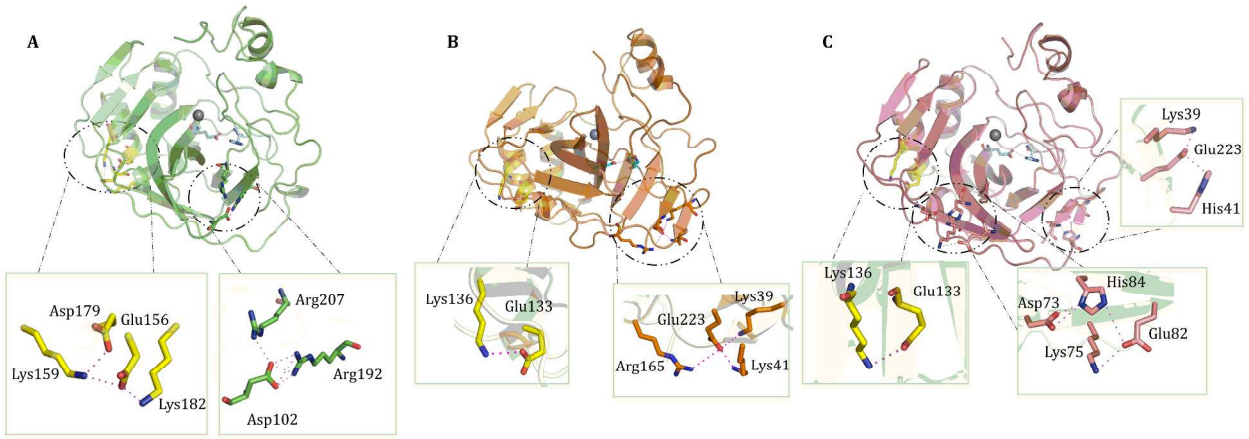


Figure 1. Relevant salt bridges and ionic networks in thermostable α-carbonic anhydrases: A) TaCA; B) SspCA; C) SazCA.

Furthermore, all bacterial CAs have an intra-monomer disulfide bridge between Cys47-Cys202 (TaCA numbering). In contrast to the other bacterial enzymes, the highly stable TaCA adopts a tetrameric structure, which is held together by a small core at the center of the tetramer. This core is formed by an additional inter-monomer disulfide link between cysteines at position 67, and by an inter-monomer ion pair between Lys65-Lys247.[13]

Molecular Dynamics Simulations and Analyses. Molecular dynamics simulations at 343 K, 353 K and 363 K were performed starting from the TaCA, SspCA and SazCA structures devoid of signal peptides, to determine the regions with high conformational flexibility, and that are hypothesized to affect the thermostability. For comparison, the mesostable NgCA was also included in this analysis. To simplify the simulations, monomeric structures were used for all enzymes, which did not take into account any stabilization provided by formation of the quaternary structures.

The complete analysis of the molecular dynamics simulation results is presented in the ESI. Simulations showed that TaCA and SspCA were stable at the temperatures investigated,

indicated by the low RMSD observed between the simulation frames and the initial crystal topology (average RMSD was 0.14 ± 0.07 nm for TaCA and 0.15 ± 0.08 for SspCA). On the other hand, the most active of the three enzymes, SazCA displayed a drastic increase in RMSD after 20 ns at 363 K (to approx. 0.50 nm), suggesting denaturation of this enzyme at high temperatures. NgCA also showed an unstable behavior at the higher temperature, although on a longer timeframe (71 ns, RMSD = 0.40 nm) compared to SazCA. The secondary structures were calculated with the DSSP algorithm,[39] for each protein and all of the MD trajectories, and indicated that no major conformational changes occurred during the simulations. This observation confirmed the overall rigidity of the four bacterial CAs at the temperatures investigated. A further analysis of the radius of gyration (R_g), used as an indicator of protein compactness, indicated a loosening of the structural network at higher temperatures in the case of NgCA and SazCA (R_g > 1.80 nm at 80 ns and 25 ns, respectively), while the other two proteins had stable R_g values over the given temperature range, with TaCA showing the least fluctuations (average R_g = 1.68 ± 0.01 nm). RMSF analysis of the MD trajectories revealed regions with increased flexibility, in all proteins studied (Figure 2), and led to the following observations:

a) The lowest average RMSF values were observed with TaCA (0.09 ± 0.01 nm), followed by SspCA (0.10 ± 0.01 nm), suggesting these two proteins to be more rigid. NgCA and SazCA were most affected by changes in temperature, showing the highest standard deviation of RMSF values at the different temperatures, calculated as a sum for all amino acids (Figure S5);

b) All structures showed enhanced flexibility at both termini, but particularly at the *N*-terminus. Since the signal peptide was not included in any of the crystal structures, and thus in the simulations, the flexibility was attributed to the *N*-terminus structure in α-CAs, perhaps due to its proximity to the active site. TaCA showed least flexibility in this region.

c) Most regions with enhanced flexibility were common to all enzymes. In particular, amino acids 165-182, identified as a non-consensus region from the sequence alignment, showed enhanced flexibility in all enzymes. Another common region with enhanced flexibility was at position Asn138 (TaCA numbering);

d) More flexibility was observed in catalytically relevant regions of SspCA and SazCA than in the other enzymes. The proton shuttle His64, and the gate-keeper residues Glu95 and Thr175 (SspCA numbering) were all situated in regions with enhanced flexibility in these two α-CAs.

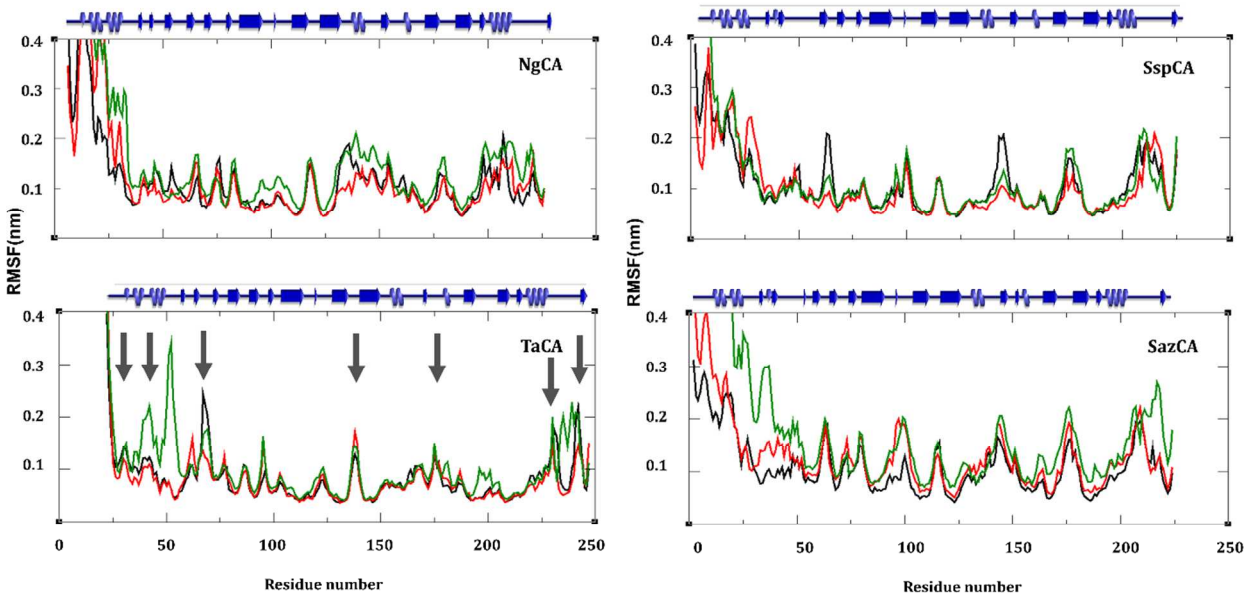


Figure 2. Backbone (Cα) RMSF of α-CA amino acids, at different simulation temperatures: 343 K (black line), 353 K (red line), 363 K (green line). Arrows show the enhanced flexibility regions in TaCA.

The RMSF analysis identified residues with high flexibility, starting from the most rigid TaCA structure. They were hypothesized to be suitable starting points for the rational design of an ultra-thermostable α-CA. For every sequence of consecutive amino acids showing enhanced flexibility during an isothermal simulation, one of the residues with the highest RMSF was

selected as representative. Representative residues that were selected during at least two isothermal simulations, as far away as possible from the active site and not part of secondary structures, were considered as highly flexible amino acids (see Methods and Table S2). Seven amino acids were identified as highly flexible using this selection procedure: Gly30, Pro42, Cys67, Asn138, Thr175, Asp232 and Ala242. From these, Cys67 was previously shown to be involved in inter-monomer disulfide bond formation promoting association into a tetramer. Therefore, this residue was excluded from further analysis, as the flexibility determined by our simulations might not be relevant in solution state, where the enzyme is present in the tetrameric form.[13] Analysis of the crystallographic B-factors obtained from the tetrameric structure indicated values above average for regions around Gly30, Pro42 and Ala 242, but not for the other amino acids identified as flexible from the RMSF analysis at different temperatures. Hence, we reason that temperature-dependent MD studies are a more reliable way to assess flexible regions within protein structures.

Taken together, the MD simulation observations described above suggest that the rigidity of the protein structure can be correlated with the thermostability of α-CAs. Previous experimental reports suggested TaCA to be the most thermostable enzyme, followed by SspCA; this is in line with our simulation results, which showed that TaCA had the least flexible structure. Simulation data also suggests that NgCA is more rigid than SazCA, which is surprising given that the latter was isolated from a thermophilic bacterium. This rigidity correlates with the unusual thermostability of NgCA at higher temperatures, which was experimentally demonstrated by Jo *et al.*, who determined a half-life of 4 h at 70 °C, higher than the SazCA half-life of 3 h at 70 °C.[8,17] On the other hand, the flexibility of SazCA seems to explain its highest activity amongst thermostable α-CAs. A clear link between flexibility and activity could not be established,

however, since the rigidity of TaCA appears not to be detrimental to its activity, which is higher than the activity of SspCA and NgCA.

Common regions with enhanced flexibility were identified by MD simulations in all the bacterial α-CAs, in particular at the protein termini. Most residues at the *C*-terminus are not in secondary structures, and therefore some natural flexibility is expected. On the other hand, the high flexibility observed for amino acids located on the α-helices at the *N*-terminus suggests that this region might play a role in defining thermostability in α-CAs. Previous directed evolution of TaCA resulted in the most stabilizing mutations being identified at the *N*-terminus of the protein devoid of the signal sequence (*i.e.* amino acids 26-43 using the numbering from this paper), supporting our finding that this region is susceptible to thermosensitivity.[19]

The analysis of standard parameters of structural stability showed that TaCA had the highest number of intra-protein hydrogen bonds at different temperatures, whilst SspCA had the highest number of protein-water interactions, as well as the highest percentage of salt bridge solvent accessible surface area (Table 1). Again, this result was correlated with the thermostability of these two enzymes. This analysis also suggested that there was space for improving TaCA thermostability, by increasing the ionic interactions between the protein and the solvent, analogous to SspCA.[14]

Table 1. **Comparison of Structural Features of Thermostability in Bacterial α-CAs**

		Hydrogen bonds		SAS (Å)	
Protein	Structure[a]	Protein-protein	Protein-water	Salt bridges[b]	Total protein
NgCA	initial	159	398	1052 (9.6)	10922
	363 K	154.2	420.2	1088 (9.7)	11235
TaCA	initial	175	429	2219 (20.5)	10800
	363 K	159.7	424.3	2320 (20.4)	11386
SspCA	initial	165	470	3124 (27.4)	11399
	363 K	151.8	462.4	3262 (27.3)	11968
SazCA	initial	171	435	2124 (19.9)	10635
	363 K	153.7	438.6	2177 (18.2)	11989

[a] The data refers to either the initial crystal structure, or the calculated averages for the MD simulation at 363 K; [b] The percentage of SAS of salt bridge residues, compared to the total SAS of the protein is given in parenthesis. Salt bridge residues refer to all amino acids involved in a salt bridge (< 4 Å) with either another amino acid, or with a charged solute molecule in the solvent. The data for all temperatures including standard deviations is represented in the ESI.

The variation of the distance between atoms involved in salt bridges was also analyzed for 100 ns of molecular dynamics simulations at different temperatures, to gain insight on the contribution of the salt bridges to thermostability (Table S4). Previous studies which compared NgCA and SspCA suggested that more rigid salt bridges were responsible for the increased thermostability of the latter.[18] This was confirmed by our simulation results, by increased lengths for the NgCA salt bridges, and relatively constant lengths for the SspCA salt bridges. For both SspCA and SazCA, we found the charged cluster Lys39-Lys41-Glu223 to maintain the interatomic distances during the simulations. However the interaction between Arg165-Glu223, which is part of the same cluster in SspCA, exhibited lower stability. In SazCA, the Asp73-Lys75-Glu82-His84 ionic network showed increased bond lengths after the simulations, suggesting the instability of this salt bridge network. In TaCA, the ion pair latch formed by Asp102, Arg192 and Arg207 was very stable at different temperatures. In contrast, the extended ionic network between Lys158-Asp179-Lys182 seemed to be disrupted when temperatures were increased during simulation, indicating the lability of these salt bridges in TaCA. The conserved His119-Glu129 and Glu156-Lys159 salt bridges (TaCA numbering) remained stable over the conditions studied, showing almost no fluctuations.

From the MD simulations, the unfolding pathways of the three proteins were analyzed, by inspecting the protein conformations at different timescales and temperatures (Figures S7-S10). For NgCA, and SazCA, some loss of secondary structure can be observed over time, in particular at the surface of the protein. Interestingly, a disruption of the parallel β-sheet structure containing

Lys39 (in β2), His41 and Glu223 (in β15) could be observed during the simulation in SazCA, particularly at 353 K and 363 K. This was not suggested by the analysis of the salt bridge distances presented above, but it might mean that this salt bridge network, present in both SspCA and SazCA, is not involved in maintaining the secondary structure. The antiparallel β-sheet β12-β13 also shortens during the simulation. The *N*-terminal region, including the helices, appear to be very flexible in both NgCA and SazCA, whilst TaCA and SspCA show more rigidity overall.

Another way to analyze the flexibility and tendency of unfolding is to inspect the free energy landscapes (FELs) at different temperatures, which represent the conformational space occupied by the enzymes during the simulations. A more expanded landscape with a higher number of separated minima indicates higher flexibility, whilst a lower barrier between separated minima indicates that a transition into non-native or non-active enzyme conformations is more likely to take place. In the case of α-CAs, the differences most prominently unravel in simulations at higher temperature (Figure 3). The less stable NgCA and SazCA showed a relatively high number of intermediate unfolding states and a broader conformational space overall. On the other hand, the more stable TaCA and SspCA displayed a smaller conformational subspace during all simulations. The stable CAs showed up to two minima on the FEL, separated by relatively low energy barriers. Examining the structures of these minima showed that they were all very similar to the crystal structure (as depicted by low RMSD values), while the minima structures of the least stable CAs depicted significant differences to the crystal structure. TaCA showed the lowest structural variance, which once again strongly supports the better thermostability of TaCA, and also gives confidence for the quality of the simulation procedures applied.

In conclusion, the molecular dynamics simulations presented in this section proposed TaCA to be the most rigid of the thermostable α-CAs characterized so far, and this is correlated with the increased stability of this enzyme at high temperatures. Nevertheless, flexible sites were identified within the TaCA structure, and they were similar to the flexible sites identified in the other thermostable CAs. The salt bridge networks identified in SspCA and SazCA, by inspection of their crystal structures, were shown to be relatively flexible during the simulations. This was in contrast to the ion pair latch of TaCA, which was relatively rigid.

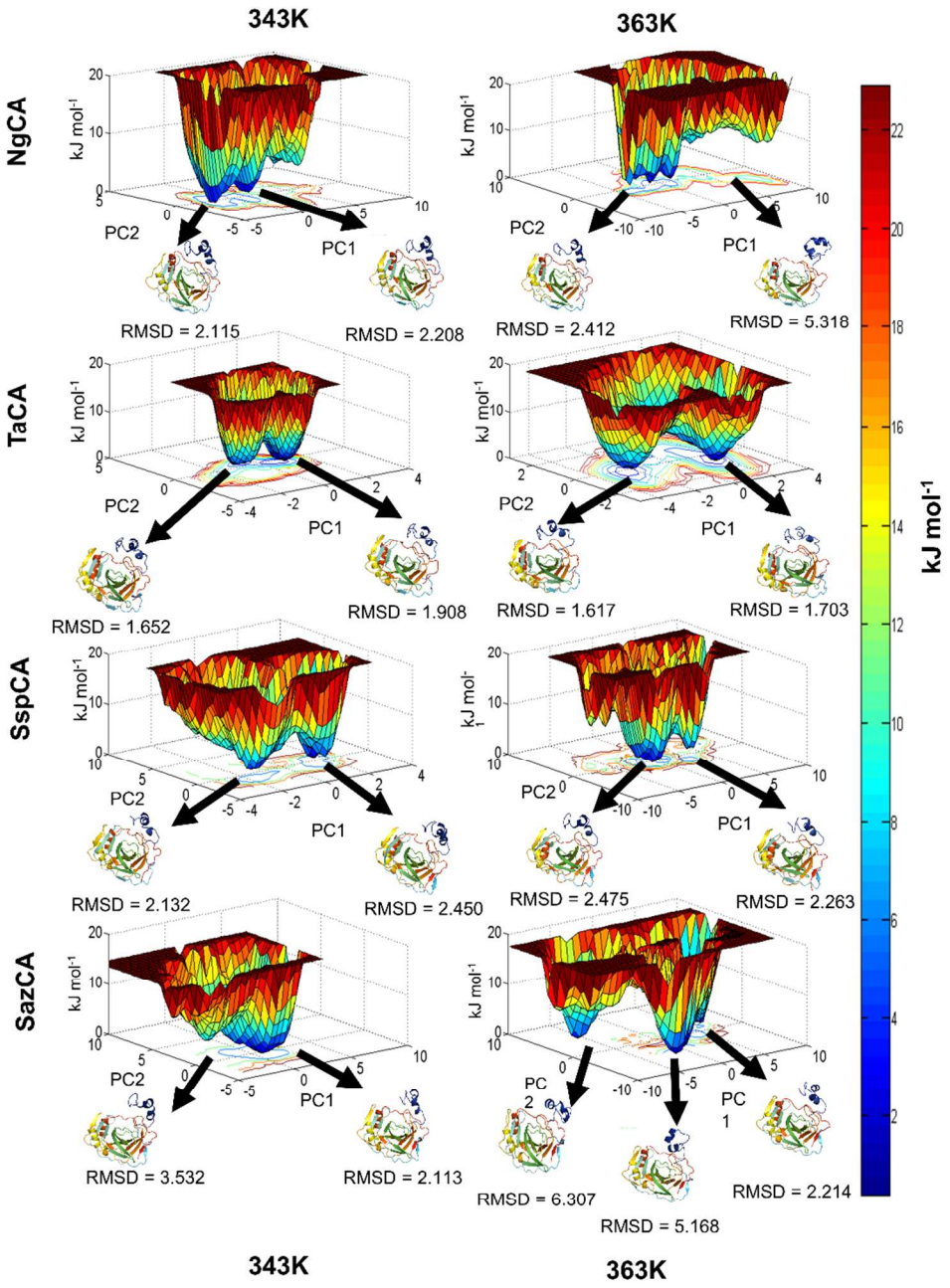


Figure 3. Free energy landscapes (FEL) for NgCA, TaCA, SspCA and SazCA from simulations at 343 K and 363 K. The x- and y-axis represent the first two eigenvectors, PC1 and PC2 of the C-alpha atomic fluctuation. The z-axis represents the free energy in kJ mol $^{-1}$. The color conventions are depicted as red (energy maxima) and blue (energy minima). A complete description of FEL construction is given in the ESI. RMSD values given in Å.

Design of Mutations to Increase Thermostability. Previous engineering of protein stability has successfully been achieved by decreasing the flexibility of the polypeptide backbone *via* rigidifying interactions, such as disulfide bonds, ion pairs, hydrogen bonds and hydrophobic interactions. Disulfide bonds have been suggested to decrease the conformational entropy of the denatured state, thus increasing the half-life of the folded protein.[15,40] The interaction of charged surface residues with solvent molecules at high temperatures was also suggested to increase thermostability.[15,41-42] Additionally, small volume non-polar residues such as glycine, and conformationally restrained proline have been shown to prevail in thermostable proteins, because they minimize hindrance and favor entropic stabilization.[41,43] Following successful identification of flexible regions within the TaCA backbone, the next focus was to assess whether the introduction of stabilizing mutations at these hot spots would yield an improved stability at high temperature, assessed by a decreased flexibility during MD simulations. These studies were performed using the monomeric structure of TaCA, and therefore could only evaluate the impact of intra-monomer rigidifying interactions. Since the flexible residues selected by MD simulations were not involved in inter-subunit interactions, we hypothesized that any stabilizing effect on the monomer would be similar for the tetramer in solution state. Additionally, the residues were not in proximity of the active site, in order to maintain flexibility and thus not to influence enzyme turnover.

In a first approach, single mutations at the flexible sites were designed, based on their ability to form ionic protein-solvent interactions, and to provide compactness, thus yielding a total of seven possible mutations at each site: Asp, Glu, Lys, Arg, His, Pro and Gly. *In silico* mutagenesis was performed using the Yasara software, and the corresponding mutations were evaluated based on the difference in folding energy between the mutant and the wild-type protein

($\Delta\Delta G^{Fold}$), calculated as an average of three temperatures using the FoldX algorithm.[36] The folding free energy calculation was performed by following the original publication by Schymkowitz *et al.,* and are presented in the ESI. Most single point mutations designed using this method were either neutral or destabilizing, in particular at the *N*-terminus, where the removal of Gly30 or Pro42 had a negative impact on stability (Figure S11). On the other hand, introduction of prolines at positions 175, 232 and 242 had an overall stabilizing effect. All designed mutations at position 138 were destabilizing, and this was also the case for position 139 (data not shown), whilst in the same flexible region, Asn140Gly was found to be a stabilizing mutation. Following this analysis, four potential stabilizing mutations were identified as candidates for MD simulation: Asn140Gly, Thr175Pro, Asp232Pro and Ala242Pro.

The second strategy for introducing rigidifying interactions was the computational design of intra-monomer disulfide bridges. The geometries, distances and energy constraints of all residue pairs were evaluated with Disulfide by Design 2.0[38] within both chains of the protein structure, in order to predict possible sites which might form disulfide bonds when mutated into cysteines. Given that the stabilizing impact of disulfides is higher in regions with greater flexibility, the ranking of potential disulfides was performed using the sum of the B-factors of the residue pair. Existing intra-monomer disulfides (Cys47 and Cys 202) were ranked highest, thus validating the design method used. A total of 36 additional potential residue pairs were predicted within TaCA, in either chain A, B or both (Table S5). From these, after exclusion of unusual torsion angles and of potential interference with TaCA salt bridges and catalytic activity, only three pairs were selected for MD studies: Gly27Cys-Ile32Cys, Lys159Cys-Ala176Cys and Pro165Cys-Gln170Cys, all of which are located in high flexibility regions.

To assess whether the designed mutations had a stabilizing effect on the TaCA structure, MD simulations were performed and the flexibility of the structures was compared to the wild-type enzyme. Molecular dynamics simulations at 400 K and over 100 ns identified five mutants with an increased rigidity in the protein backbone, indicated by the relatively stable RMSD trace during the simulation (Figure 4). Most of these mutations, with the exception of Ala242Pro, decreased the flexibility at the *N*-terminus, although this region was far away from the mutation points. Only the disulfide mutant Cys165-Cys170 showed a decreased RMSF in the region corresponding to the mutations. Interestingly, although at considerable distance, both the disulfide mutant and Thr175Pro introduced flexibility in the loop region 119-125, whilst Asn140Gly reduced the flexibility of this loop. Since we have shown that rigidity and thermostability are correlated in α-CAs, we propose that the mutants designed here are likely to exhibit higher thermostability than the wild-type enzyme.

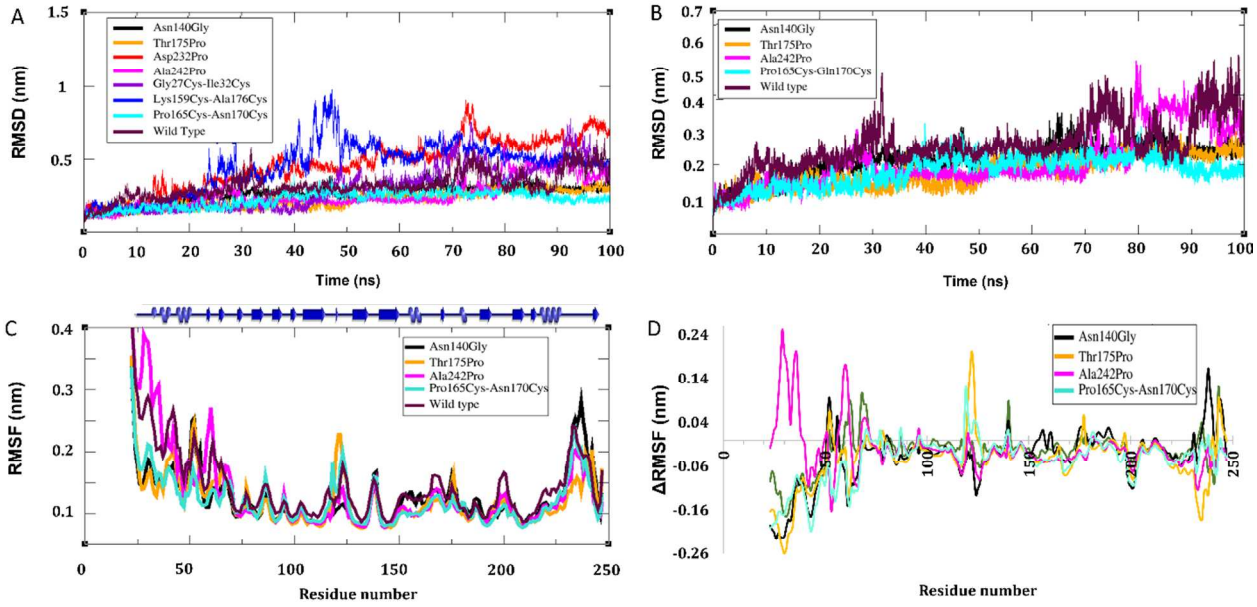


Figure 4. Molecular dynamics simulation results for the designed TaCA mutants, over 100 ns at 400 K. A) RMSD of the eight mutants; B) RMSD of the five mutants with improved average

RMSD compared to wild-type TaCA; C) RMSF of the five mutants with improved average RMSD compared to wild-type TaCA; d) ΔRMSF calculated between the five mutants and wild type, with negative values representing an increased rigidity of the protein.

The best two mutants with overall smaller RMSF values compared to the wild-type protein were Pro165Cys-Asn170Cys and Asn140Gly. Although the overall rigidity of the protein was increased, the RMSF of the active site residues (Val133, Val143, Leu197, Val206 and Trp209 in the CO_2-binding pocket) were relatively similar. This suggests that these mutations did not interfere with the flexibility of the active site, required for enzyme turnover. Interestingly, the hydrophilic proton-shuttle residues (Tyr28, Asn85, Thr198 and Thr199) showed lower flexibility in the mutants compared to the wild-type enzyme, which might be a result of a more compact protein structure, favoring stronger hydrogen-bonding interactions within the protein, and between residues and water molecules.

Inspection of the FEL landscapes for the most promising mutants, Pro165Cys-Asn170Cys and Asn140Gly, showed a clear improvement compared to wild-type TaCA (Figure 5). At this higher temperature, the wild-type protein showed a high flexibility, the FEL resembled the pattern of the less stable proteins (NgCA and SazCA) at 363 K, and conformations with a disordered *N*-terminus were identified in the minima structures. On the other hand, the FEL of the disulfide mutant was very close to that of wild-type TaCA at the lower temperature (Figure 3 *vs* Figure 5), thus demonstrating the higher stability of this mutant. Introduction of a rigidifying bond in a loop region is likely to be responsible for this stabilization of the protein. Asn140Gly showed an intermediate stability, sampling a larger conformational space than the disulfide mutant, but with minima structures relatively close to the wild-type TaCA.

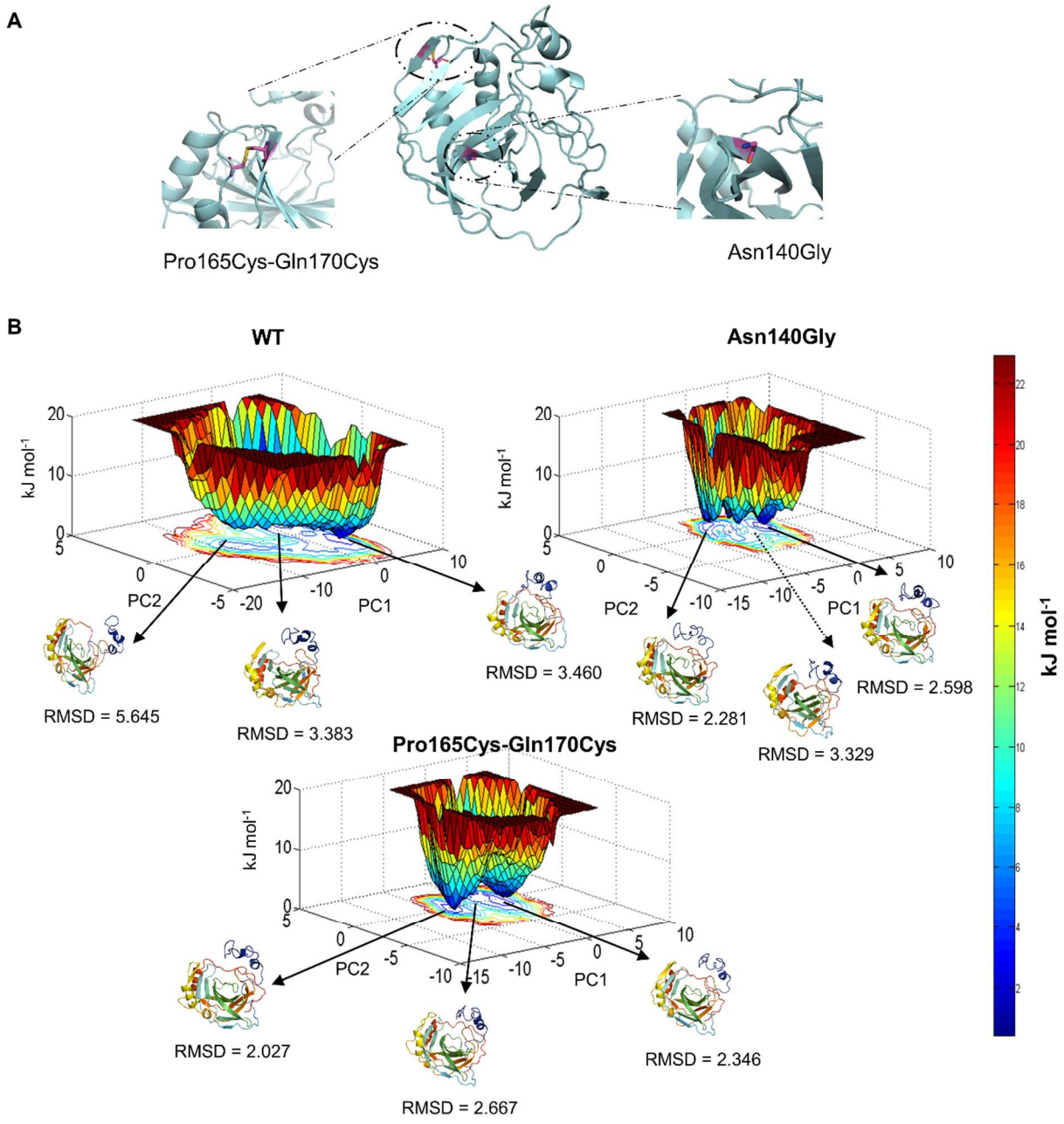


Figure 5. A) Representation of rigidifying mutation sites within TaCA; B) Free energy landscapes for wild-type TaCA, Asn140Gly and Pro165Cys-Gln170Cys from simulations at 400 K. The x- and y-axis represent the first two eigenvectors, PC1 and PC2 of the C-alpha atomic fluctuation. The z-axis represents the free energy in kJ mol $^{-1}$. The color conventions are

depicted as red (energy maxima) and blue (energy minima). A complete description of FEL construction is given in the ESI. RMSD values given in Å.

CONCLUSION

The results presented here provide an increased insight into the thermostability of bacterial α-CAs, and suggest a rational approach for the design of mutants of TaCA with increased stability, for application in carbon capture biotechnologies. The comparison of bacterial α-carbonic anhydrases by molecular dynamics simulations showed that the origin of their stability at high temperatures most likely resides in the rigidity of the protein structure. The most thermostable TaCA and SspCA enzymes showed the least fluctuation in the protein backbone, and the least flexibility in the ion pair networks. FEL analysis confirmed TaCA as the enzyme with the highest rigidity, and thus the most stable. The *N*- and *C*- termini regions of all enzymes were determined by RMSF analysis to be highly flexible. At the difference from previous studies, where rigidifying features from thermostable scaffolds were re-designed into mesostable ones, we used the most thermostable α-CA as a starting point to systematically investigate all flexible sites and their potential for further stabilization. Specific amino acids with RMSF above average during dynamics at 343 K, 353 K and 363 K were identified and mutated *in silico* to a selection of charged or small non-polar residues, in an effort to stabilize the protein structure of TaCA, and create an ultra-thermostable enzyme. Initial refinement performed with FoldX allowed the identification of five stabilizing mutations. In an alternative stabilization approach, three disulfide bonds were designed at regions with high flexibility. Out of the eight mutants analyzed by MD simulations at 400 K, five showed a more rigid structure than wild-type TaCA, with lower RMSD and RMSF values. In particular, the Cys165-Cys170 disulfide bond showed

decreased flexibility at the rigidification site, whilst Asn140Gly showed an overall decreased flexibility. Comparison of free energy landscapes between the wild-type and mutant TaCA demonstrated a higher stability of the mutants at 400 K, in particular for the mutant containing a disulfide bond. Future work will include an analysis of combined mutations, to assess their effect on TaCA stability. Efforts are currently underway to experimentally characterize the mutants suggested by this computational study.

Associated Content

Supporting information. The following Supporting Information is available free of charge.

Sequence analysis and salt bridge analysis of thermostable alpha-CAs; complete analysis of MD simulation results; FoldX and DbD 2.0 mutation analysis (PDF).

Author Information

Corresponding author

E-mail: anca.pordea@nottingham.ac.uk

Telephone: +44 (0) 115 951 4087

ORCID Anca Pordea: 0000-0001-8453-0743

Acknowledgements

The research was funded by the Malaysian Ministry of Higher Education through the Fundamental Research Grant Scheme, project no. FRGS/1/2014/TK05/UNIM/02/2. RPC was funded by a PhD studentship from the University of Nottingham. CMJ acknowledges funding

through the Nottingham Advanced Research Fellowship and EU FP7 Marie Curie Actions - People, Co-funding of Regional, National and International Programmes (COFUND) under Grant Agreement no PCOFUND-GA-2012-600181. We also gratefully acknowledge support and access to the University of Nottingham High Performance Computing Facility.

REFERENCES

1. Supuran, C. T. Structure and Function of Carbonic Anhydrases. *Biochem. J.* **2016,** *473*, 2023-2032.

2. Zheng, Y. J.; Merz, K. M. Mechanism of the Human Carbonic Anhydrase II-Catalyzed Hydration of Carbon Dioxide. *J. Am. Chem. Soc.* **1992,** *114*, 10498-10507.

3. Bose, H.; Satyanarayana, T. Microbial Carbonic Anhydrases in Biomimetic Carbon Sequestration for Mitigating Global Warming: Prospects and Perspectives. *Front. Microbiol.* **2017,** *8*, 1615.

4. Savile, C. K.; Lalonde, J. J. Biotechnology for the Acceleration of Carbon Dioxide Capture and Sequestration. *Curr. Opin. Biotechnol.* **2011,** *22*, 818-823.

5. Di Fiore, A.; Alterio, V.; Monti, S.; De Simone, G.; Ambrosio, K. Thermostable Carbonic Anhydrases in Biotechnological Applications. *Int. J. Mol. Sci.* **2015,** *16*, 15456-15480.

6. Alvizo, O.; Nguyen, L. J.; Savile, C. K.; Bresson, J. A.; Lakhapatri, S. L.; Solis, E. O. P.; Fox, R. J.; Broering, J. M.; Benoit, M. R.; Zimmerman, S. A.; et al. Directed Evolution of an Ultrastable Carbonic Anhydrase for Highly Efficient Carbon Capture from Flue Gas. *Proc. Natl. Acad. Sci.* **2014,** *111*, 16436-16441.

7. Capasso, C.; De Luca, V.; Carginale, V.; Cannio, R.; Rossi, M. Biochemical Properties of a Novel and Highly Thermostable Bacterial α-Carbonic Anhydrase from *Sulfurihydrogenibium yellowstonense* Yo3aop1. *J. Enzyme Inhib. Med. Chem.* **2012,** *27*, 892-897.

8. Luca, V. D.; Vullo, D.; Scozzafava, A.; Carginale, V.; Rossi, M.; Supuran, C. T.; Capasso, C. An α-Carbonic Anhydrase from the Thermophilic Bacterium *Sulphurihydrogenibium azorense* Is the Fastest Enzyme Known for the CO_2 Hydration Reaction. *Bioorg. Med. Chem.* **2013,** *21*, 1465-1469.

9. Jo, B. H.; Seo, J. H.; Cha, H. J. Bacterial Extremo- α-Carbonic Anhydrases from Deep-Sea Hydrothermal Vents as Potential Biocatalysts for CO_2 Sequestration. *J. Mol. Catal. B: Enzym.* **2014,** *109*, 31-39.

10. Kanth, B. K.; Jun, S.-Y.; Kumari, S.; Pack, S. P. Highly Thermostable Carbonic Anhydrase from *Persephonella marina* Ex-H1: Its Expression and Characterization for CO_2-Sequestration Applications. *Process Biochem.* **2014,** *49*, 2114-2121.

11. De Simone, G.; Monti, S. M.; Alterio, V.; Buonanno, M.; De Luca, V.; Rossi, M.; Carginale, V.; Supuran, C. T.; Capasso, C.; Di Fiore, A. Crystal Structure of the Most Catalytically Effective Carbonic Anhydrase Enzyme Known, SazCA from the Thermophilic Bacterium *Sulfurihydrogenibium azorense*. *Bioorg. Med. Chem. Lett.* **2015,** *25*, 2002-2006.

12. Di Fiore, A.; Capasso, C.; De Luca, V.; Monti, S. M.; Carginale, V.; Supuran, C. T.; Scozzafava, A.; Pedone, C.; Rossi, M.; De Simone, G. X-Ray Structure of the First `Extremo-[Alpha]-Carbonic Anhydrase', a Dimeric Enzyme from the Thermophilic Bacterium *Sulfurihydrogenibium yellowstonense* Yo3aop1. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013,** *69*, 1150-1159.

13. James, P.; Isupov, M. N.; Sayer, C.; Saneei, V.; Berg, S.; Lioliou, M.; Kotlar, H. K.; Littlechild, J. A. The Structure of a Tetrameric [Alpha]-Carbonic Anhydrase from *Thermovibrio ammonificans* Reveals a Core Formed around Intermolecular Disulfides That Contribute to Its Thermostability. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014,** *70*, 2607-2618.

14. Vogt, G.; Woell, S.; Argos, P. Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs. *J. Mol. Biol.* **1997,** *269*, 631-643.

15. Modarres, H. P.; Mofrad, M. R.; Sanati-Nezhad, A. Protein Thermostability Engineering. *RSC Adv.* **2016,** *6*, 115252-115270.

16. Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally Designed Libraries for Rapid Enzyme Stabilization. *Protein Eng., Des. Sel.* **2014,** *27*, 49-58.

17. Jo, B. H.; Park, T. Y.; Park, H. J.; Yeon, Y. J.; Yoo, Y. J.; Cha, H. J. Engineering De Novo Disulfide Bond in Bacterial [Alpha]-Type Carbonic Anhydrase for Thermostable Carbon Sequestration. *Sci. Rep.* **2016,** *6*, 29322.

18. Bharatiy, S. K.; Hazra, M.; Paul, M.; Mohapatra, S.; Samantaray, D.; Dubey, R. C.; Sanyal, S.; Datta, S.; Hazra, S. In Silico Designing of an Industrially Sustainable Carbonic Anhydrase Using Molecular Dynamics Simulation. *ACS Omega* **2016,** *1*, 1081-1103.

19. Voyer, N.; Daigle, R.; Madore, É.; Fradette, S. Variants of *Thermovibrio ammonificans* Carbonic Anhydrase and CO_2 Capture Methods Using *Thermovibrio ammonificans* Carbonic Anhydrase Variants. Int. Pat. Appl. WO2017035667 A1, Mar 09, 2017.

20. Di Fiore, A.; Pedone, C.; Antel, J.; Waldeck, H.; Witte, A.; Wurl, M.; Scozzafava, A.; Supuran, C. T.; De Simone, G. Carbonic Anhydrase Inhibitors: The X-Ray Crystal Structure of Ethoxzolamide Complexed to Human Isoform Ii Reveals the Importance of Thr200 and Gln92 for Obtaining Tight-Binding Inhibitors. *Bioorg. Med. Chem. Lett.* **2008,** *18*, 2669-2674.

21. Huang, S.; Xue, Y.; Sauer-Eriksson, E.; Chirica, L.; Lindskog, S.; Jonsson, B. H. Crystal Structure of Carbonic Anhydrase from *Neisseria gonorrhoeae* and Its Complex with the Inhibitor Acetazolamide. *J. Mol. Biol.* **1998,** *283*, 301-310.

22. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High□Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7*, 539.

23. Costantini, S.; Colonna, G.; Facchiano, A. M. Esbri: A Web Server for Evaluating Salt Bridges in Proteins. *Bioinformation* **2008,** *3*, 137-138.

24. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

25. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, Ψ and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8*, 3257-3273.

26. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987,** *91*, 6269-6271.

27. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating pK(a)s and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005,** *33*, W368-W371.

28. Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual Molecular Dynamics. *J. Mol. Graphics* **1996,** *14*, 33-38.

29. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

30. Reißer, S.; Poger, D.; Stroet, M.; Mark, A. E. Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017,** *13*, 2367-2372.

31. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. Lincs: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997,** *18*, 1463-1472.

32. Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985,** *83*, 4069-4074.

33. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981,** *52*, 7182-7190.

34. Maiorov, V. N.; Crippen, G. M. Size-Independent Comparison of Protein Three-Dimensional Structures. *Proteins: Struct., Funct., Bioinf.* **1995,** *22*, 273-283.

35. Van Durme, J.; Delgado, J.; Stricher, F.; Serrano, L.; Schymkowitz, J.; Rousseau, F. A Graphical Interface for the FoldX Forcefield. *Bioinformatics* **2011,** *27*, 1711-1712.

36. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005,** *33*, W382-W388.

37. Studer, R. A.; Christin, P.-A.; Williams, M. A.; Orengo, C. A. Stability-Activity Tradeoffs Constrain the Adaptive Evolution of Rubisco. *Proc. Natl. Acad. Sci.* **2014,** *111*, 2223-2228.

38. Craig, D. B.; Dombkowski, A. A. Disulfide by Design 2.0: A Web-Based Tool for Disulfide Engineering in Proteins. *BMC Bioinformatics* **2013,** *14*, 346.

39. Carter, P.; Andersen, C. A. F.; Rost, B. Dsspcont: Continuous Secondary Structure Assignments for Proteins. *Nucleic Acids Res.* **2003,** *31*, 3293-3295.

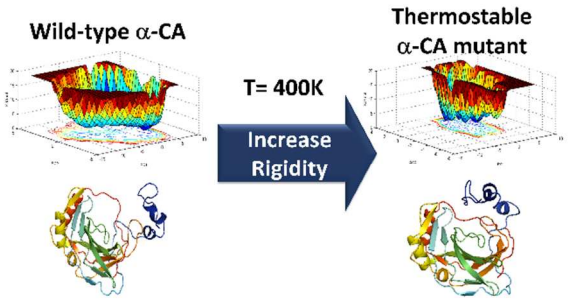
40. Yu, H.; Huang, H. Engineering Proteins for Thermostability through Rigidifying Flexible Sites. *Biotechnol. Adv.* **2014,** *32*, 308-315.

41. Panja, A. S.; Bandopadhyay, B.; Maiti, S. Protein Thermostability Is Owing to Their Preferences to Non-Polar Smaller Volume Amino Acids, Variations in Residual Physico-Chemical Properties and More Salt-Bridges. *PLoS ONE* **2015,** *10*, e0131495.

42. Szilágyi, A.; Závodszky, P. Structural Differences between Mesophilic, Moderately Thermophilic and Extremely Thermophilic Protein Subunits: Results of a Comprehensive Survey. *Structure* **2000,** *8*, 493-504.

43. Yi, Z.-L.; Pei, X.-Q.; Wu, Z.-L. Introduction of Glycine and Proline Residues onto Protein Surface Increases the Thermostability of Endoglucanase Cela from Clostridium Thermocellum. *Bioresour. Technol.* **2011,** *102*, 3636-3638.

TOC graphic


Wild-type α-CA
Thermostable
α-CA mutant
T= 400K
Increase
Rigidity

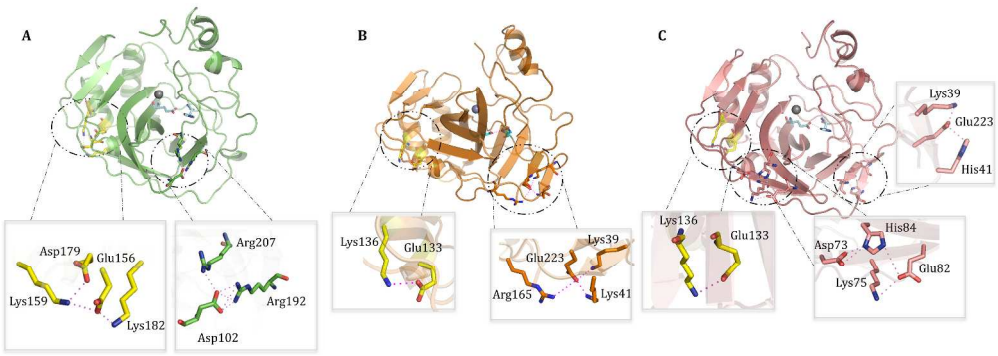


Figure 1. Relevant salt bridges and ionic networks in thermostable α-carbonic anhydrases: A) TaCA; B) SspCA; C) SazCA.

2106x738mm (96 x 96 DPI)

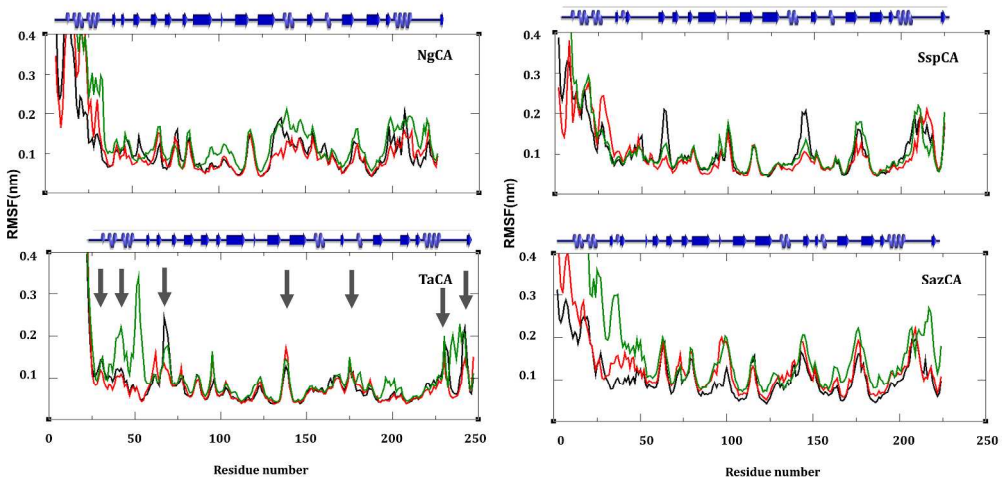


Figure 2. Backbone (Cα) RMSF of α-CA amino acids, at different simulation temperatures: 343 K (black line), 353 K (red line), 363 K (green line). Arrows show the enhanced flexibility regions in TaCA.

2110x1054mm (96 x 96 DPI)

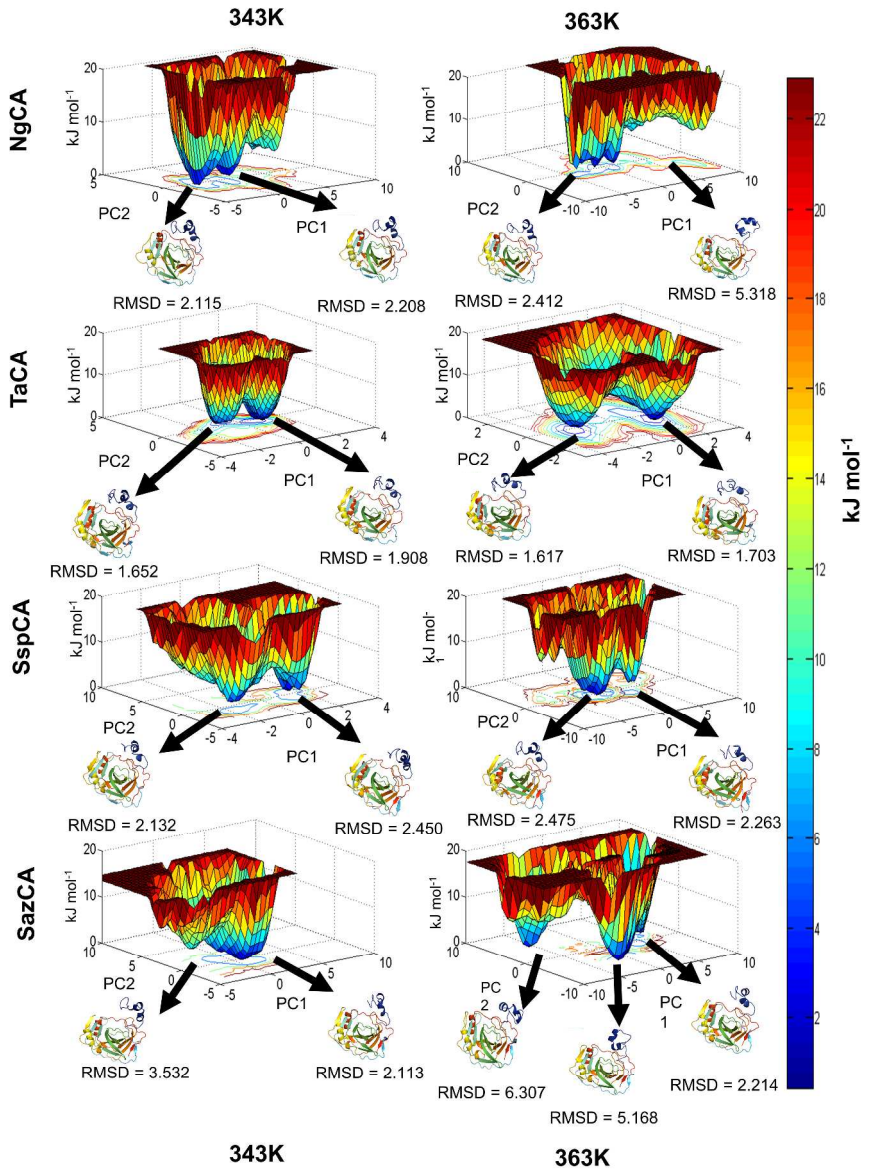


Figure 3. Free energy landscapes (FEL) for NgCA, TaCA, SspCA and SazCA from simulations at 343 K and 363 K. The x- and y-axis represent the first two eigenvectors, PC1 and PC2 of the C-alpha atomic fluctuation. The z-axis represents the free energy in kJ mol $^{-1}$. The color conventions are depicted as red (energy maxima) and blue (energy minima). A complete description of FEL construction is given in the ESI. RMSD values given in Å.

1164x1602mm (96 x 96 DPI)

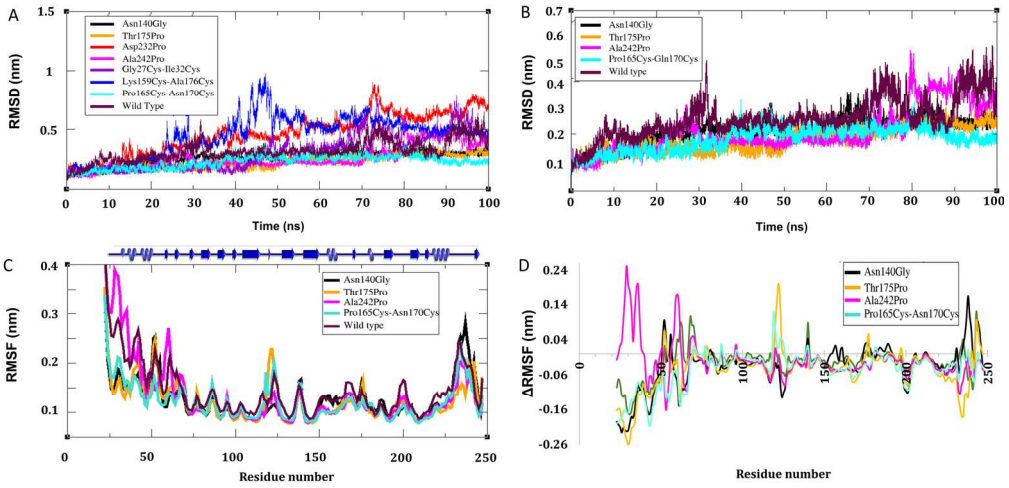


Figure 4. Molecular dynamics simulation results for the designed TaCA mutants, over 100 ns at 400 K. A) RMSD of the eight mutants; B) RMSD of the five mutants with improved average RMSD compared to wild-type TaCA; C) RMSF of the five mutants with improved average RMSD compared to wild-type TaCA; d) ΔRMSF calculated between the five mutants and wild type, with negative values representing an increased rigidity of the protein.

169x84mm (300 x 300 DPI)

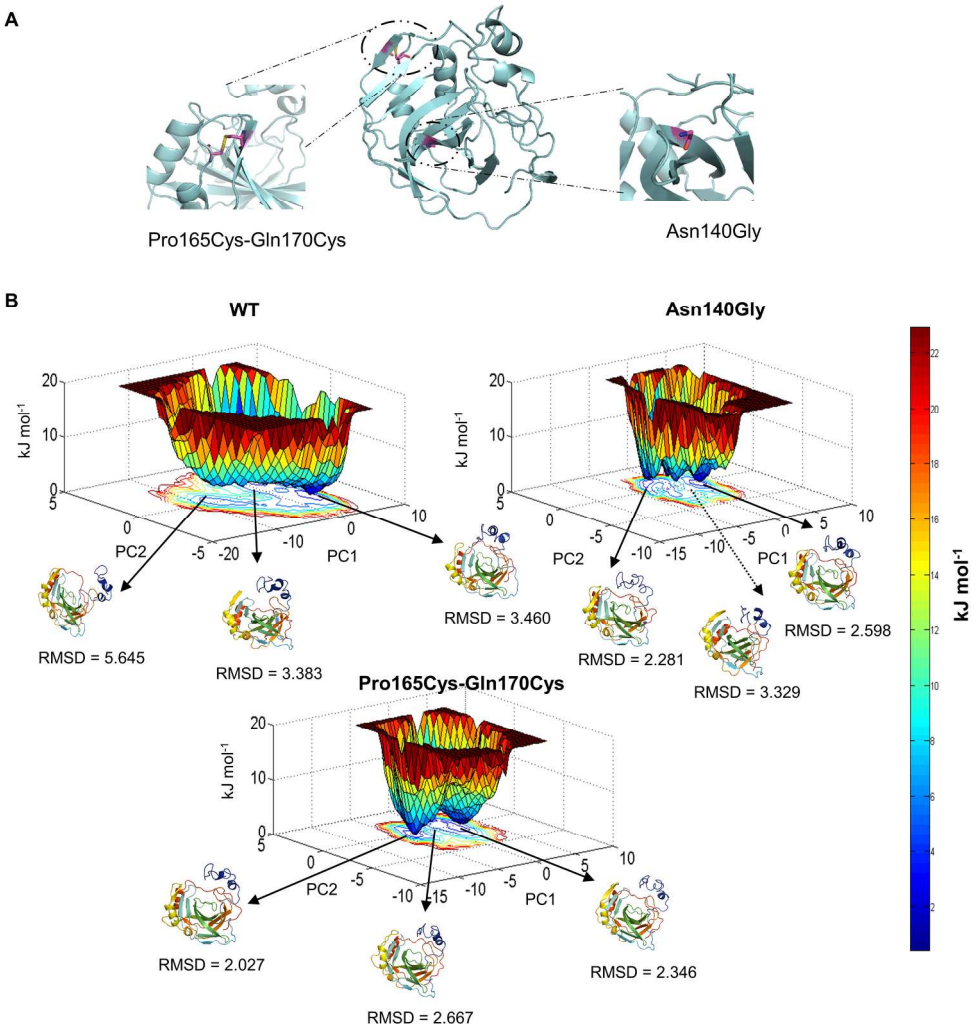
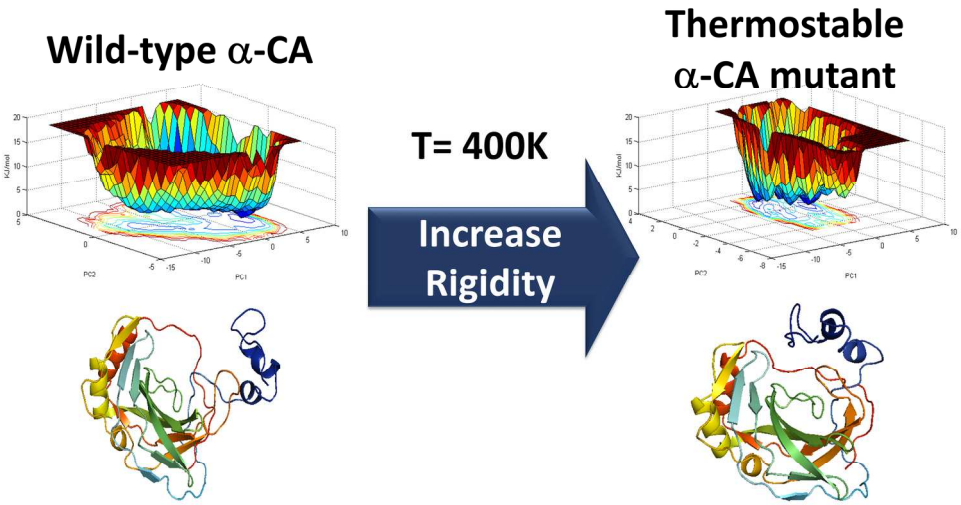


Figure 5. A) Representation of rigidifying mutation sites within TaCA; B) Free energy landscapes for wild-type TaCA, Asn140Gly and Pro165Cys-Gln170Cys from simulations at 400 K. The x- and y-axis represent the first two eigenvectors, PC1 and PC2 of the C-alpha atomic fluctuation. The z-axis represents the free energy in kJ mol $^{-1}$. The color conventions are depicted as red (energy maxima) and blue (energy minima). A complete description of FEL construction is given in the ESI. RMSD values given in Å.

203x214mm (300 x 300 DPI)



190x107mm (300 x 300 DPI)