Abstract

Novel biomimetic routes to develop ecofriendly carbon capture, utilisation and sequestration (CCUS) processes is a challenging area of research. Recent years have seen an increasing interest in the utilisation of some selected family of enzymes to carry out effective and efficient bioconversions in various dimensions of research and industrial sector, such as energy, pharamceutical, health, chemistry, agricultutre, and others. However, for several applications, these natural occuring enzymes are not contemplated to be viable owing to their poor stability at elevated temperatures. Over the years, the quest for ultramodern enzymes with high potential for biotechnological applications has evolved serveral complementary approaches, such as rational protein design engineering and mining enzymes variants from thermophiles living at high temperatures (323-363 K), to develop more stable enzymes. In this context, carbonic anhydrases (CAs) active at higher temperatures are of extreme interest. To this end, we first employed the supervised machine learning (ML) algorithms to untangle the potential patterns behind protein thermostability. Our computational analysis conclusively shows inverse gamma turns, VIII turns and propensity of CYS to be the most critical features responsible for protein thermostability. From the propensity analysis of amino acids, polar residues, GLN and SER, and charged residues, GLU and LYS, were found to favour the enhancement of protein thermostability.

The fastest member of the CA family catalysing the reversible hydration of carbon dioxide to bicarbonate ions has been recently reported to be SazCA. Due to the potential of practical implementation of this enzyme in CCUS, we examined the the structure and dynamics of thermostable SazCA, probed using molecular dynamics (MD) simulations. The widely used biomolecular force-fields (AMBER99SB*, CHARMM22, CHARMM36 and OPLS-AA) in conjunction with TIP3P water model were investigated to describe the molecular system and the comparison MD simulation results suggested AMBER99SB* to be a suitable choice to describe the structure and dynamics of SazCA. Our high temperature MD simulations identified the amino acid residues VAL98, ASN99, GLY100, LYS101, GLU145 and HIS207 as the most flexible residues. In addition to this, ion-pairs ASP113-LYS81, ASP115-LYS81, ASP115-LYS114, GLU144-LYS143, and GLU144-LYS206, were responsible for the compromised thermostability of SazCA. In this series of studies under the MD framework, next we attemped to gain insights into the thermodynamic stability of SazCA and its implications on protein folding/unfolding. The structural analysis

showed that the protein exihibited the highest structural stability at 353 K and the protein denaturation occured above 353 K. The thermodynamics analysis of protein stability revealed that the conformations that denature at higher melting temperatures tend to have greater maximum thermal stability. Our simulations revealed that SazCA has the highest melting temperature at 353 K, which was close to the experimental reported optimum temperature. The energy landscape confirmed a transition in folding/unfolding pathways of SazCA at 353 K.

Further, in this series of MD studies of SazCA, we induced a single point GLY100ALA mutation in SazCA and examined the factors governing the stability and flexibility of the mutated form, and compared it to that of the wildtype. We observed higher structural stability and lesser residual mobility in the mutated SazCA. Improved H-bonding due to GLY100ALA was observed and GLYALA100 mutation was responsible for the increment in helical contents in the mutated SazCA, while GLY100 compromized the secondary structure contents in the wildtype. A strong network of ions-pairs and high local structuring of the solvent molecules at the protein surface contributed to the enhanced stability of the mutated protein.

Apart from the thermostability of SazCA, we also attempted to explore the molecular basis for the exceptional activity of SazCA, in contrast to SspCA, under the framework of MD simulations. Our simulations, carried out at different elevated temperatures, indicated the presence of efficient proton shuttle between the active zinc centre and HIS64 residue in the two enzymes. The proton accepting HIS64 residue was identified to have "in" and "out" conformations with the in conformations being supportive to proton accpetance. Our simulations showed a large population of in conformations in SazCA making the enzyme exhibit an exceptional activity. The structural analysis confirmed the role of HIS2 and HIS207 in supporting the attainment of in conformations in SazCA resulting in exceptional activity. The molecular details provided in this comprehensive study is expected to provide a vital guide to make new thermostable variants for industrially viable CCUS processes.

Keywords: thermostability; machine learning; carbonic anhydrase; SazCA; molecular simulations