



Article

# In Silico Investigation of Potential Applications of Gamma Carbonic Anhydrases as Catalysts of CO<sub>2</sub> Biomineralization Processes: A Visit to the Thermophilic Bacteria *Persephonella hydrogeniphila*, *Persephonella marina*, *Thermosulfidibacter takaii*, and *Thermus thermophilus*

Colleen Varaidzo Manyumwa and Özlem Tastan Bishop \*

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa; colleen.manyumwa06@gmail.com

\* Correspondence: o.tastanbishop@ru.ac.za; Tel.: +27-46-603-8072; Fax: +27-46-603-7576



**Citation:** Manyumwa, C.V.; Bishop, Ö.T. In Silico Investigation of Potential Applications of Gamma Carbonic Anhydrases as Catalysts of CO<sub>2</sub> Biomineralization Processes: A Visit to the Thermophilic Bacteria *Persephonella hydrogeniphila*, *Persephonella marina*, *Thermosulfidibacter takaii*, and *Thermus thermophilus*. *Int. J. Mol. Sci.* **2021**, *22*, 2861. <https://doi.org/10.3390/ijms22062861>

Academic Editor: Clemente Capasso

Received: 23 February 2021

Accepted: 3 March 2021

Published: 11 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Carbonic anhydrases (CAs) have been identified as ideal catalysts for CO<sub>2</sub> sequestration. Here, we report the sequence and structural analyses as well as the molecular dynamics (MD) simulations of four  $\gamma$ -CAs from thermophilic bacteria. Three of these, *Persephonella marina*, *Persephonella hydrogeniphila*, and *Thermosulfidibacter takaii* originate from hydrothermal vents and one, *Thermus thermophilus* HB8, from hot springs. Protein sequences were retrieved and aligned with previously characterized  $\gamma$ -CAs, revealing differences in the catalytic pocket residues. Further analysis of the structures following homology modeling revealed a hydrophobic patch in the catalytic pocket, presumed important for CO<sub>2</sub> binding. Monitoring of proton shuttling residue His69 (*P. marina*  $\gamma$ -CA numbering) during MD simulations of *P. hydrogeniphila* and *P. marina*'s  $\gamma$ -CAs ( $\gamma$ -PhCA and  $\gamma$ -PmCA), showed a different behavior to that observed in the  $\gamma$ -CA of *Escherichia coli*, which periodically coordinates Zn<sup>2+</sup>. This work also involved the search for hotspot residues that contribute to interface stability. Some of these residues were further identified as key in protein communication via *betweenness centrality* metric of dynamic residue network analysis. *T. takaii*'s  $\gamma$ -CA showed marginally lower thermostability compared to the other three  $\gamma$ -CA proteins with an increase in conformations visited at high temperatures being observed. Hydrogen bond analysis revealed important interactions, some unique and others common in all  $\gamma$ -CAs, which contribute to interface formation and thermostability. The seemingly thermostable  $\gamma$ -CA from *T. thermophilus* strangely showed increased unsynchronized residue motions at 423 K.  $\gamma$ -PhCA and  $\gamma$ -PmCA were, however, preliminarily considered suitable as prospective thermostable CO<sub>2</sub> sequestration agents.

**Keywords:** gamma carbonic anhydrase; homology modeling; MD simulations; *betweenness centrality*; MD-TASK; hydrothermal vents; carbon dioxide sequestration

## 1. Introduction

The increasing concentrations of CO<sub>2</sub> in the atmosphere have resulted in a search for mitigation strategies to counter global warming [1,2]. Biomineralization, a process where CO<sub>2</sub> is reacted with mineral ions, such as Ca<sup>2+</sup> or Mg<sup>2+</sup>, to produce mineral carbonates, has been identified as a viable sequestration method [3]. The carbonates produced by this way present a stable source of carbon without the possibility of CO<sub>2</sub> re-entering the atmosphere. However, a catalyst is normally required for the quick capture of CO<sub>2</sub>. Carbonic anhydrases (CAs) have been identified as preferable catalysts for this reaction, being responsible for the rapid reversible conversion of CO<sub>2</sub> to HCO<sub>3</sub><sup>-</sup> in the presence of water [4]. Eight CA classes have been identified to date, including the  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\zeta$ ,  $\eta$ ,  $\theta$ , and  $\iota$  classes, all utilizing a metal ion in the active site, which is predominantly a Zn<sup>2+</sup> [5–9]. The Zn<sup>2+</sup> is involved in catalysis by means of a Zn-bound hydroxide molecule, generated by the transfer of a

proton from a Zn-bound water. The hydroxide molecule is responsible for the nucleophilic attack on the CO<sub>2</sub> molecule, producing HCO<sub>3</sub><sup>−</sup> [10,11].

Apart from the α- and β-CAs, literature on the γ-CAs has been recently increasing. They have been found in a number of organisms, mostly bacteria, although some have been identified in fungi and plants [12–16]. Those characterized from plants, however, have displayed the inability to catalyze the reversible hydration of CO<sub>2</sub>, but instead, have been observed as essential subunits of mitochondrial complex 1, an enzyme complex in the plant respiratory chain [17–19]. All γ-CAs are structurally similar, adopting a homotrimeric architecture dominated by β-sheets and an α-helical C-terminal. These CAs are only functional in their trimeric form because each active site is located between two monomers, with two coordinating residues coming from one monomer and the third from the neighboring one. The catalytic metal ion in the active site is usually Fe<sup>2+</sup> but γ-CAs have also been observed to utilize Zn<sup>2+</sup> and Co<sup>2+</sup> [12,20,21]. Similar to the α-CAs, γ-CA's catalytic Zn<sup>2+</sup> is tetrahedrally coordinated by three His residues and the fourth coordination is fulfilled by a water molecule. Despite the functionality differences of identified γ-CAs, the Zn<sup>2+</sup> coordination residues have been completely conserved to date [12,15,17]. Variability of residues in the catalytic pocket, however, has been evident in this class. Catalytically active γ-CAs include those from *Escherichia coli* (EcoCA-γ) [22], *Methanosarcina thermophila* (Cam) [13,23], the pathogenic *Porphyromonas gingivalis* (Pg-iCA) [24], *Pseudomonas aeruginosa* PAO1 (PA5540) [14], the cyanobacteria *Nostoc* sp. PCC 7120 [25] and *Thermosynechococcus elongatus* BP-1 (CcmM) [26]. Organisms with γ-CAs whose activity could not be detected include the small plant, *Arabidopsis thaliana* [15], the cyanobacterium *Synechococcus* PCC 7942 [27] as well as *Thermus thermophilus* HB8 [28].

Thermostability is a desirable characteristic in catalysts for biotechnology purposes. This is especially applicable for carbon sequestration at industrial sites, where there is mass production of industrial flue gas at high temperatures, containing large concentrations of CO<sub>2</sub>. Accordingly, the main focus of this study was to assess selected γ-CA proteins in silico for thermostability properties. The γ-CAs were from three organisms, *Persephonella hydrogeniphila* [29], *Persephonella marina* [30] and *Thermosulfidibacter takaii* [31], which were isolated from hydrothermal vent systems. A fourth γ-CA protein, from a hot spring thermophile *Thermus thermophilus* HB8 (γ-TtCA), was included in this study, based on its alleged thermostability properties, for comparison purposes [28].

An in silico approach was motivated by a similar study of the α-CAs including those from *P. hydrogeniphila* and *P. marina*, revealing important functional characteristics of the CAs as well as identifying possible thermostable sequestration agents [32]. Computational studies of the well-studied α-CA from *Thermovibrio ammonificans* have also been performed, where possible residue mutations for enhanced thermostability were identified [33]. The variants expressed after mutation studies indeed produced a more stable CA with an increased tolerance for high temperatures [34]. The prospects for in silico studies are therefore considered monumental and worth exploring.

Here, we retrieved the γ-CA sequences and aligned with previously identified γ-CA sequences for comparison. This included the γ-CA from the hydrothermal vent originating hyperthermophilic archaeon, *Pyrococcus horikoshii* (Cap), whose activity has not yet been determined [35]. The structures for *P. hydrogeniphila*, *P. marina* and *T. takaii* CAs (γ-PhCA, γ-PmCA and γ-TtkCA respectively) were modeled as trimers substantiating the importance of the interface analysis to identify important interface residues. *T. thermophilus*' crystal structure was already available. Hotspot residues, defined as those whose mutation to alanine results in destabilization of the interface due to their significant energy contribution, were also identified [36]. Analysis of the active site revealed a hydrophobic region proposed to be functional for CO<sub>2</sub> binding. Molecular dynamics (MD) simulation analyses revealed the conformational behavior of the protein structures at high temperatures. Average betweenness centrality (BC) analysis [37] revealed important residues in protein communication linked to the function of the γ-CAs. Most of these were found in the interface, including some catalytic site residues, also located in the interface. Generally, γ-PhCA

and  $\gamma$ -PmCA showed indications of thermostability, including structure rigidity, at high temperatures. This study also unveiled the structural factors behind the lack of catalytic activity previously observed in  $\gamma$ -TtCA [28]. Altogether, this study builds on research on the  $\gamma$ -CAs and simultaneously discloses potential thermostable CO<sub>2</sub> sequestration agents.

## 2. Results and Discussion

### 2.1. Sequence and Structural Analyses

Residue numbering henceforth is in compliance with  $\gamma$ -PmCA unless explicitly stated. Given that there are three different interfaces, the chain identities of the residues involved will be stated as superscripts in cases where a bond is being discussed.

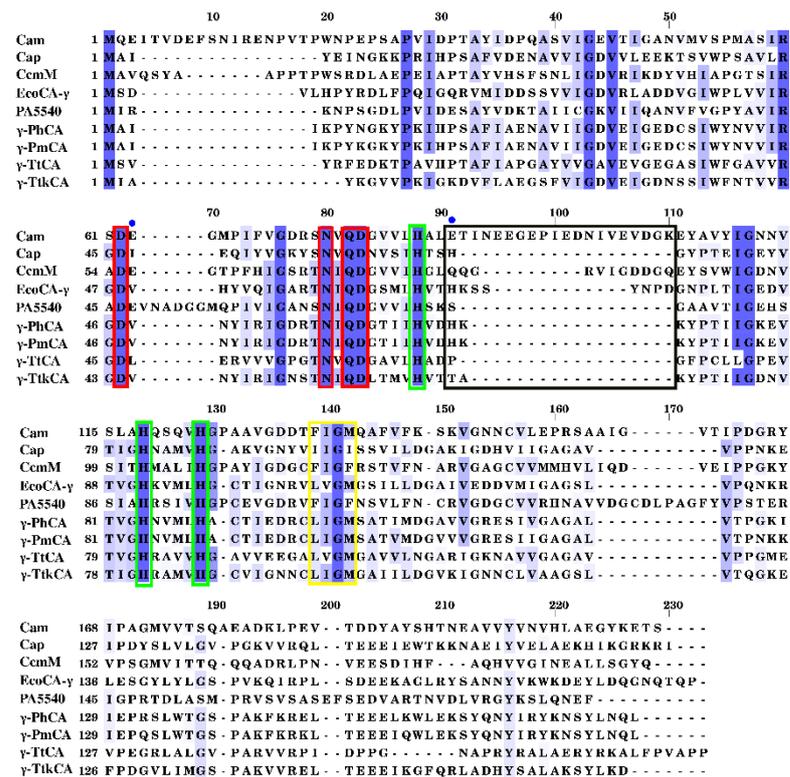
#### 2.1.1. Multiple Sequence Alignment (MSA)

MSA of the  $\gamma$ -CA sequences was performed by PROMALS3D [38] alignment program (Figure 1). *Methanosarcina thermophila* (Cam), *Pyrococcus horikoshii* (Cap), *Thermosynechococcus elongatus* BP-1 (CcmM), *Esherischia coli* (EcoCA- $\gamma$ ), *Pseudomonas aeruginosa* (PA5540), *Persephonella hydrogeniphila* ( $\gamma$ -PhCA), *Persephonella marina* ( $\gamma$ -PmCA), *Thermus thermophilus* ( $\gamma$ -TtCA) and *Thermosulfidibacter takaii* ( $\gamma$ -TtkCA) sequences were included in the MSA. High variability was observed amongst the sequences. Residues that coordinate the catalytic Zn<sup>2+</sup> metal, His66, His84, His89, were completely conserved across sequences as expected. Additionally, conserved were Arg45, Asp47, Gln60 and Asp61 which play a role in catalysis [28,35,39]. Arg45<sup>ABC</sup> and Asp61<sup>BCA</sup> have been reported to form an ion pair and are marked by red boxes in the  $\gamma$ -CA MSA, along with Asp47 and Gln60 [39,40]. The most distinct variation was the insert found in Cam between Leu83 and Gly102 (Cam numbering). Loops for CcmM and EcoCA- $\gamma$  were also slightly longer than the four CAs being investigated. Two proton transfer residues crucial in the catalytic mechanism for Cam, Glu62 and Glu84 with the former found on the elongated loop, were absent in  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA, and  $\gamma$ -TtkCA. In the position for Glu84 (Cam numbering),  $\gamma$ -PhCA and  $\gamma$ -PmCA, similar to Cap and EcoCA- $\gamma$ , contained a His residue, which has been proposed to be involved in proton transfer [41]. This residue has, however, been observed occasionally coordinating Zn<sup>2+</sup> in EcoCA- $\gamma$ , resulting in a “closed” conformation. This in turn, resulted in a mechanism similar to the  $\beta$ -CAs in which they switch between “open” and “closed” states [42,43]. This phenomenon was, thus, searched for  $\gamma$ -PhCA and  $\gamma$ -PmCA following MD simulations.  $\gamma$ -TtCA and  $\gamma$ -TtkCA were different from these CAs, containing a Pro and Thr residue respectively, in this position.

#### 2.1.2. Homology Modeling and Structural Analyses

3D structures of  $\gamma$ -PhCA,  $\gamma$ -PmCA, and  $\gamma$ -TtkCA proteins were modeled as generic  $\gamma$ -CAs, which are known to exist as trimers and all three passed the verification process (Table 1). Verify3D [44] scores were above the threshold of 80% and PROCHECK [45] confirmed no residues to be in the disallowed region thus the model qualities were satisfactory. The crystal structure for  $\gamma$ -TtCA (PDB ID: 6IVE) was used for all further analyses.

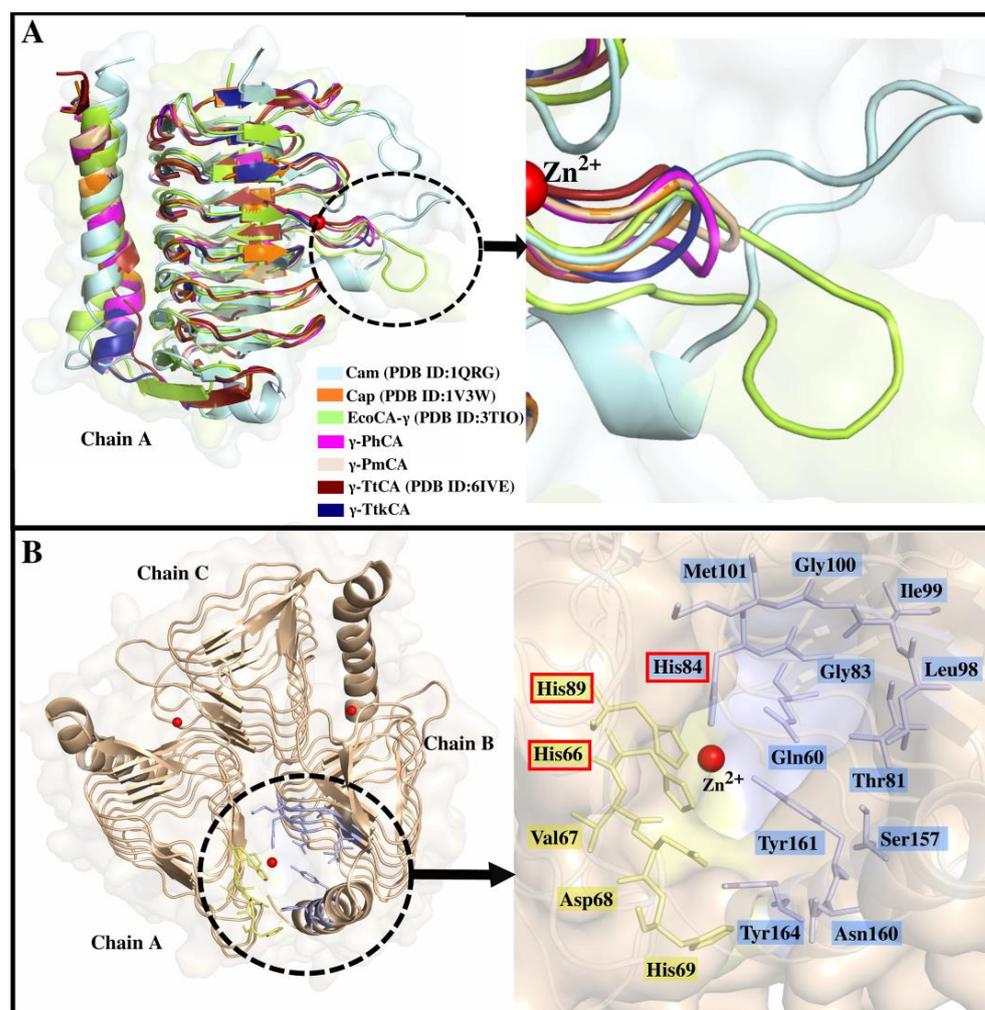
Each monomer contained the typical left-handed  $\beta$ -helix resembling a prism-like shape, with loops between  $\beta$ -sheets. The C-terminal was made of a long  $\alpha$ -helix. The loop between  $\beta$ -sheet 5 and 6, shown in Figure 2A displayed variability amongst  $\gamma$ -CAs upon comparison with Cam, Cap and EcoCA- $\gamma$ .  $\gamma$ -PhCA,  $\gamma$ -PmCA, and  $\gamma$ -TtkCA revealed a shorter loop compared to Cam and EcoCA- $\gamma$  but similar to Cap and  $\gamma$ -TtCA. This region corresponds to the inserts displayed in the MSA (Figure 1).



**Figure 1.** Multiple sequence alignment of the  $\gamma$ -CA sequences performed by PROMALS3D alignment program. Abbreviations: *Methanosarcina thermophila*—Cam; *Pyrococcus horikoshii*—Cap; *Thermosynechococcus elongatus* BP-1—CcmM; *Escherichia coli*—EcoCA- $\gamma$ ; *Pseudomonas aeruginosa*—PA5540; *Persephonella hydrogeniphila*— $\gamma$ -PhCA; *Persephonella marina*— $\gamma$ -PmCA; *Thermus thermophilus*— $\gamma$ -TtCA; and *Thermosulfidibacter takaii*— $\gamma$ -TtkCA. Residues are colored by the extent of conservation. Residues in the red boxes are important for catalysis while Zn<sup>2+</sup> coordinating His residues are shown in green boxes. The black box depicts the insert present for Cam, CcmM, and EcoCA- $\gamma$ . The blue dots show proton shuttling Glu residues identified in Cam. Yellow box depicts residues around the hydrophobic region in the catalytic site of  $\gamma$ -PmCA.

**Table 1.** Template coverage, sequence identity and model validation for  $\gamma$ -CA sequences.

CA	Organism	Template PDB ID	Template—Sequence Identity	Template—Sequence Coverage	z-DOPE Score	Verify 3D (%)	Procheck	
							Most Favored Region (%)	Disallowed Region (%)
$\gamma$ -PhCA	<i>Persephonella hydrogeniphila</i>		46%	98%	−1.42	89.6	86.6	0
$\gamma$ -PmCA	<i>Persephonella marina</i>	1V3W	46%	98%	−1.62	88.6	88.8	0
$\gamma$ -TtkCA	<i>Thermosulfidibacter takaii</i>		44%	99%	−1.80	89.9	89.9	0



**Figure 2.** (A): Aligned monomeric  $\gamma$ -CA structures of Cam (cyan), Cap (orange), EcoCA- $\gamma$  (green),  $\gamma$ -PhCA (magenta),  $\gamma$ -PmCA (wheat),  $\gamma$ -TtCA (maroon), and  $\gamma$ -TtkCA (dark blue). The loop region indicated in the MSA is shown by the black dotted circle and zoomed into in the image pointed to by the black arrow. (B): Trimeric structure of  $\gamma$ -PmCA with the active site which is located between Chains A and B delimited by the black dotted circle. The black arrow points to the enlarged active site showing the residues that contribute to its formation.  $\text{Zn}^{2+}$  is shown as a red sphere and the three His residues coordinating it are boxed in red.

The open active site of the structures expectedly resembled that of previously crystallized  $\gamma$ -CAs, containing three His residues and a free fourth coordination position to be fulfilled by a water molecule (Figure 2B). This coordination was shared between two monomers. The catalytic pocket was further scrutinized in this study, and other residues were identified as contributing to its surface. Hydrophobic residues Leu98, Ile99, Gly100, and Met101 form a hydrophobic patch at one end of the pocket. This is surmised to play a role in  $\text{CO}_2$  binding similar to the hydrophobic pocket in  $\alpha$ -CAs [11,32,46,47]. Met135 (Cam numbering) has been previously noted as part of a possible hydrophobic  $\text{CO}_2$  binding site in Cam, but with a different set of non-conserved residues, including some in the extended loop shown in Figure 2A [23]. It was interesting to observe that the hydrophobicity of this region in the alignment was conserved amongst  $\gamma$ -PhCA,  $\gamma$ -PmCA, and  $\gamma$ -TtkCA as well as previously crystallized Cam, Cap, EcoCA- $\gamma$ , and  $\gamma$ -TtCA (see yellow box in Figure 1). In the absence of the two Glu proton transfer residues from Cam, a different channel is evidently followed by these CAs, given the high catalytic activity displayed by recently characterized EcoCA- $\gamma$ . Orientation of Tyr161 towards  $\text{Zn}^{2+}$  in the catalytic site evinces its role in proton transfer as previously suggested for the  $\gamma$ -CA from *Arabidopsis thaliana* as

well as Cap [15,35]. It occupies the structural position of Cam's Asn202, whose involvement in catalysis has been confirmed. Tyr164, which is also contributing to the catalytic pocket surface (Figure 2B), is also suggested to participate in proton transfer but was absent in both  $\gamma$ -TtCA and  $\gamma$ -TtkCA. It should be pointed out that  $\gamma$ -TtCA did not contain any effective substitutions for proton transfer residues His69 and Tyr164 ( $\gamma$ -PmCA numbering). Both the residues were hydrophobic, disrupting the proton transfer pathway, thus explaining the absence of activity previously observed in vitro [28]. However, the substitute residues for Cam's Glu62 and  $\gamma$ -PmCA's Tyr164 in  $\gamma$ -TtCA (Leu47<sup>ABC</sup> and Leu156<sup>BCA</sup> respectively) were observed to form hydrophobic interactions in the interfaces.

Besides this interaction between monomers, other interactions have been observed following submission of  $\gamma$ -CA structures to the five web servers used to identify interface residues in this study and four of these servers for hotspot residue identification. Importance of the hotspot residues in interface stability is significantly more compared to other interface residues. Results from each server are listed in Supplementary material, Table S1. A consensus of the residues contributing to the interfaces of the structure, consequently supporting the trimeric form of the CAs, is displayed in Table 2. The previously discussed salt bridge formed by Arg45<sup>ABC</sup> and Asp61<sup>BCA</sup> was recognized between monomers in all structures. Both these residues were recognized as hotspot residues in all four  $\gamma$ -CAs. Arg45 was also involved in an intra-subunit ionic interaction with Asp47. This chain of ionic bonds is hereby referred to as the Asp-Arg-Asp\* ionic triad, with the asterisk indicating a residue from a different monomer. Met101 which is part of the hydrophobic patch in the catalytic pocket was also recognized as a hotspot residue in all  $\gamma$ -CAs. It was interesting to note that  $\gamma$ -TtCA was the only CA amongst the four containing hotspot residues in the N-terminal, specifically Tyr4 and Phe6. Tyr4<sup>ABC</sup> was observed to form hydrophobic interactions with Leu164<sup>BCA</sup> and Pro166<sup>BCA</sup> while Phe6<sup>ABC</sup> forms hydrophobic interactions with Leu156<sup>BCA</sup>, Tyr160<sup>BCA</sup>, and Leu164<sup>BCA</sup> in the C-terminal.

Interface characteristics from PDBePISA [48] are detailed in Table S2 (Supplementary material). These revealed the total interface areas, also regarded as the buried surface areas (BSAs) for  $\gamma$ -PhCA and  $\gamma$ -PmCA to be 5155 Å<sup>2</sup> and 5353 Å<sup>2</sup> respectively while a lower BSA of 4749 Å<sup>2</sup> was seen for  $\gamma$ -TtkCA.  $\gamma$ -TtCA showed an extensive BSA of 6846 Å<sup>2</sup> which, when correlated to the total surface area of the multimer, the BSA accounted for 28% of the protein. For the other three CAs, the BSAs constituted a smaller percentage, approximately 20–21% of the proteins. This correlated with the higher number of hydrogen bonds and salt bridges observed for  $\gamma$ -TtCA using both PDBePISA and Protein Interactions Calculator (PIC) [49] servers compared to the other structures. Occupancy of the hydrogen bonds was further analyzed across MD simulations in a later section (Section 2.2.3). Structures of previously crystallized Cap (PDB ID: 1V3W) and Cam (PDB ID: 1QRM) were also submitted to the PDBePISA web server and were observed to have BSAs of totals 7842 Å<sup>2</sup> and 6180 Å<sup>2</sup> respectively, which is irreconcilable to values previously reported (18,775 Å<sup>2</sup> and 23,352 Å<sup>2</sup> respectively) [35]. These covered approximately 26.5% and 25% of the total surface area of the structures. Models were thus less closely packed compared to their template as well as Cam.

**Table 2.** Residues common to three programs identified as participating in interface formation. Hotspot residues are in bold.

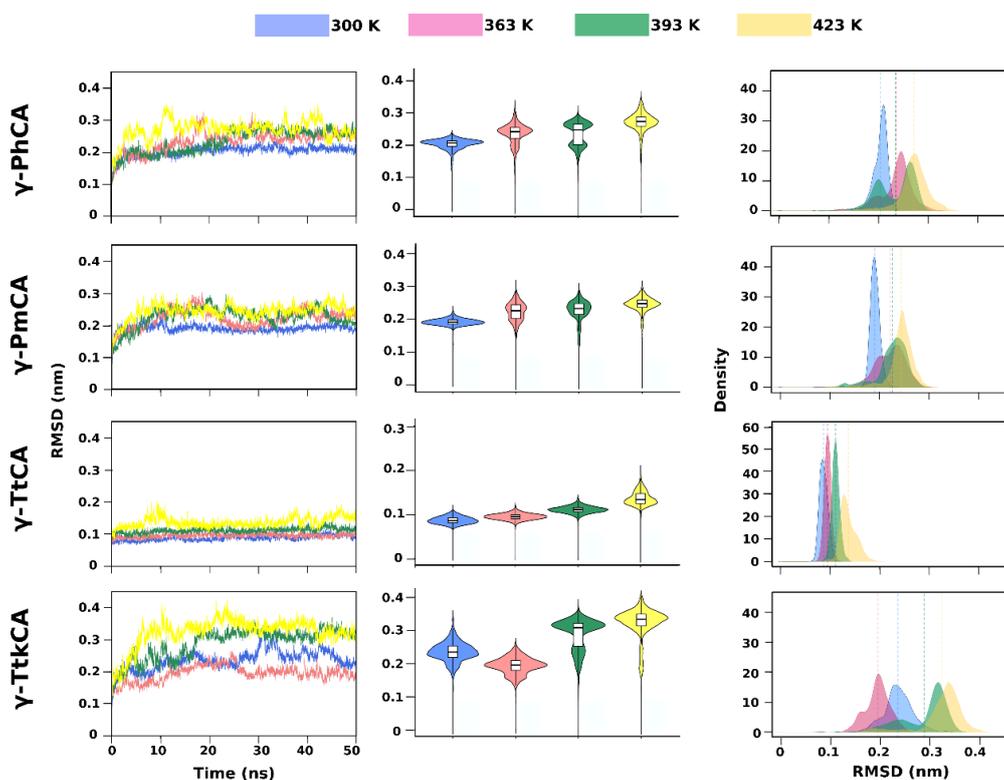
CA	Residues
$\gamma$ -PhCA	<b>Interface AB</b> A: M1, A2, I4, K5, P6, Y7, N8, V25, I27, N41, V43, <b>R45</b> , D47, T63, <b>I64</b> , H66, H69, K70, N85, V86, M87, H89, T104, I105, M106, L122, S138 B: E22, N23, W39, <b>Y40</b> , N41, Q60, <b>D61</b> , G62, <b>H84</b> , <b>N85</b> , <b>M101</b> , S102, A119, T136, W153, N160, Y161, Y164, S167, <b>Y168</b> , Q171
	<b>Interface BC</b> B: I4, K5, P6, Y7, N8, V25, I27, <b>N41</b> , V43, <b>R45</b> , <b>D47</b> , V48, G62, T63, <b>I64</b> , H66, D68, H69, K70, N85, V86, M87, H89, S102, T104, I105, M106, L122, S138 C: N23, W39, <b>Y40</b> , N41, Q60, <b>D61</b> , G62, <b>H84</b> , <b>N85</b> , <b>M101</b> , S102, A119, T136, W153, K156, N160, Y161, <b>Y164</b> , S167, Y168, Q171
	<b>Interface AC:</b> A: N23, W39, <b>Y40</b> , N41, Q60, <b>D61</b> , H84, N85, M101, S102, A119, T136, N160, Y161, <b>Y164</b> , S167, Y168, N170, Q171 C: A2, I4, K5, Y7, N8, V25, I27, N41, V43, <b>R45</b> , D47, G62, T63, <b>I64</b> , H66, H69, V86, M87, H89, S102, T104, M106, D107, L122, S138
$\gamma$ -PmCA	<b>Interface AB</b> A: I3, I4, K5, P6, Y7, K8, Y11, N23, V25, I27, N41, V42, V43, <b>R45</b> , D47, V48, I64, <b>H66</b> , V67, D68, H69, N85, V86, M87, H89, A102, T104, M106, L122, S138 B: N23, W39, <b>Y40</b> , N41, Q60, <b>D61</b> , <b>H84</b> , N85, <b>M101</b> , S102, A119, T136, W153, N160, Y161, <b>Y164</b> , S167, <b>Y168</b> , N170, Q171, L172
	<b>Interface BC</b> B: I3, I4, K5, P6, Y7, K8, V25, <b>I27</b> , N41, V43, <b>R45</b> , <b>D47</b> , V48, G62, T63, <b>I64</b> , H66, H69, N85, V86, M87, H89, S102, T104, V105, M106, L122, S138 C: N23, W39, <b>Y40</b> , N41, Q60, <b>D61</b> , <b>H84</b> , <b>N85</b> , <b>M101</b> , S102, A119, T136, N160, Y161, Y164, S167, <b>Y168</b> , N170, Q171, L172
	<b>Interface AC:</b> A: E22, N23, W39, <b>Y40</b> , N41, <b>D61</b> , H84, N85, <b>M101</b> , S102, A119, G120, W153, N160, Y161, <b>Y164</b> , S167, <b>Y168</b> , Q171, L172 C: M1, A2, I3, I4, K5, P6, Y7, K8, V25, I27, N41, V43, <b>R45</b> , D47, G62, <b>I64</b> , H66, D68, H69, <b>N85</b> , M87, H89, S102, T104, V105, M106, L122
$\gamma$ -TtCA	<b>Interface AB</b> A: S2, V3, <b>Y4</b> , R5, <b>F6</b> , E7, K9, T10, Y24, V26, V42, <b>R44</b> , D46, L47, <b>V63</b> , <b>H65</b> , <b>R83</b> , V85, H87, G100, A101, V102, L104, V120, V136 B: W38, <b>F39</b> , Q59, <b>D60</b> , <b>H82</b> , <b>R83</b> , <b>M99</b> , A117, G118, V136, R152, <b>Y153</b> , L156, R159, <b>Y160</b> , A163, <b>L164</b> , F165, V167
	<b>Interface BC</b> B: V3, <b>Y4</b> , R5, <b>F6</b> , E7, Y24, <b>V26</b> , V42, <b>R44</b> , D46, L47, <b>V63</b> , <b>H65</b> , <b>R83</b> , V85, H87, V102, L104, V120, V136 C: P21, G22, Y24, W38, <b>F39</b> , Q59, <b>D60</b> , <b>H82</b> , <b>R83</b> , M99, A117, G118, V136, R152, <b>Y153</b> , <b>L156</b> , R159, <b>Y160</b> , A163, <b>L164</b> , F165, P166, V167
	<b>Interface AC:</b> A: P21, G22, W38, <b>F39</b> , Q59, <b>D60</b> , <b>H82</b> , <b>R83</b> , <b>M99</b> , A117, G118, L134, V136, R152, <b>Y153</b> , <b>L156</b> , R159, <b>Y160</b> , A163, L164, F165, P166, V167, A168, T169 C: M1, S2, V3, <b>Y4</b> , R5, <b>F6</b> , E7, T10, Y24, V26, V42, <b>R44</b> , <b>D46</b> , L47, <b>V63</b> , <b>H65</b> , D67, P68, R83, A84, V85, H87, G100, A101, V102, L104, V120, V136
$\gamma$ -TtkCA	<b>Interface AB</b> A: I2, Y4, K5, <b>F22</b> , I24, <b>N38</b> , T39, <b>V40</b> , <b>R42</b> , D44, V45, <b>L59</b> , T60, M61, <b>H63</b> , R82, A83, M84, H86, <b>I101</b> , L103, L119, S135, P136 B: E19, G20, W36, <b>F37</b> , N38, Q57, <b>D58</b> , <b>L59</b> , <b>H81</b> , <b>R82</b> , M98, G99, A116, G117, M133, H157, Y158, L161, Y165, D168
	<b>Interface BC</b> B: Y4, K5, V7, <b>F22</b> , I24, N38, V40, <b>R42</b> , D44, V45, L59, T60, <b>M61</b> , H63, R82, M84, H86, I101, L103, L119, S135 C: E19, W36, <b>F37</b> , <b>N38</b> , Q57, <b>D58</b> , <b>L59</b> , <b>H81</b> , <b>R82</b> , <b>M98</b> , A116, G117, M133, H157, Y158, L161, Y165, D168
	<b>Interface AC:</b> A: E19, G20, W36, <b>F37</b> , N38, Q57, <b>D58</b> , L59, H81, R82, <b>M98</b> , G99, A116, G117, H157, Y158, L161, Y165, D168 C: Y4, K5, G6, <b>F22</b> , I24, N38, V40, <b>R42</b> , D44, V45, L59, T60, M61, H63, R82, M84, H86, G99, I101, L103, L119, S135

## 2.2. Molecular Dynamics Simulations

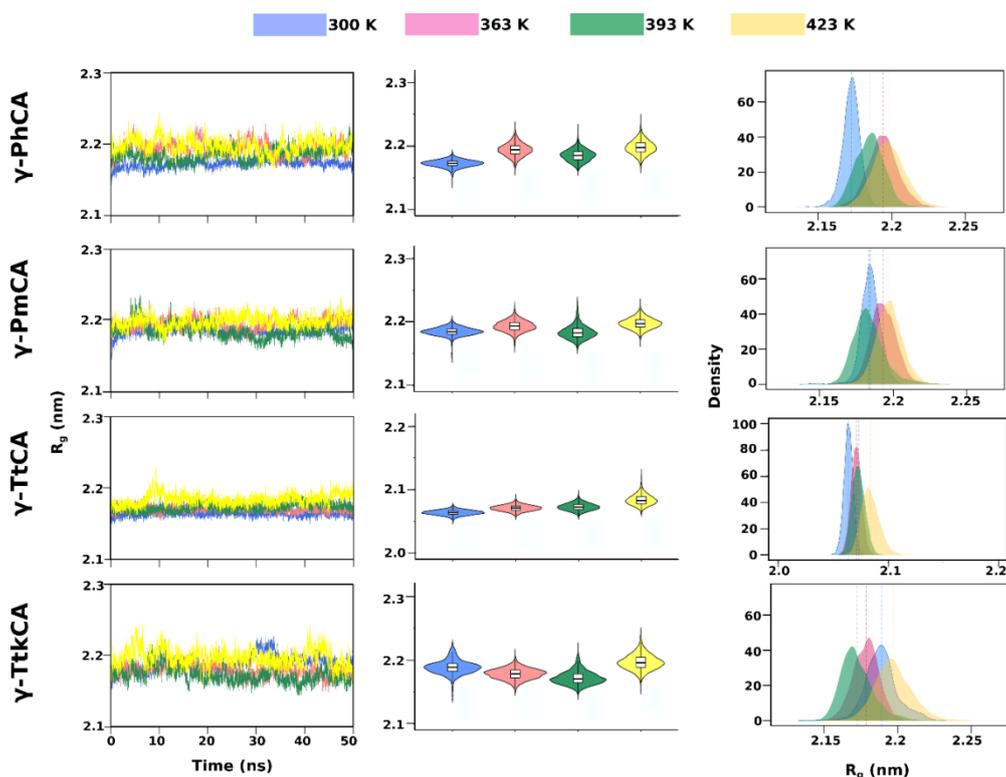
Simulations at increasing temperatures were mainly to identify possible thermostability in the  $\gamma$ -CAs from *P. hydrogeniphila*, *P. marina*, *T. takaii*, and *T. thermophilus*. This was done mostly through conformational dynamics analyses. However, probing protein functionality at residue level gave rise to insights on residues important for stability and function.

### 2.2.1. Conformational Analysis

Investigation of the conformations visited by the  $\gamma$ -CAs during simulations at 300 K, 363 K, 393 K, and 423 K was achieved by root mean square deviation (RMSD) and radius of gyration ( $R_g$ ) calculations. These were visualized as line graphs, violin plots and kernel density estimation (KDE) plots in Figure 3 (RMSD) and Figure 4 ( $R_g$ ). The line graphs convey the evolution of conformations as a function of time, facilitating a perception of particular occurrences such as equilibration and increases or decreases in RMSD/ $R_g$ . The violin plots and the KDE plots illustrate the probability densities of the conformations sampled, enabling the observation of RMSD/ $R_g$  distribution of conformations across the simulations. Although the violin plots are basically mirrored KDE plots, they also contain a box plots showing the median and interquartile range of the conformational densities.



**Figure 3.** Root mean square deviation (RMSD) line graphs (**left**), violin plots (**middle**), and kernel density estimation (KDE) plots (**right**) of  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA and  $\gamma$ -TtkCA at 300 K, 363 K, 393 K and 423 K. Median values in the violin plots are shown by the black line in the white box plots and those in the KDE plots are shown as a line with a color corresponding to the respective plot color.



**Figure 4.** Radius of gyration ( $R_g$ ) line graphs (left), violin plots (middle), and kernel density estimation (KDE) plots (right) of  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA and  $\gamma$ -TtkCA at 300 K, 363 K, 393 K and 423 K. Median values in the violin plots are shown by the black line in the white box plots and those in the KDE plots are shown as a line with a color corresponding to the respective plot color.

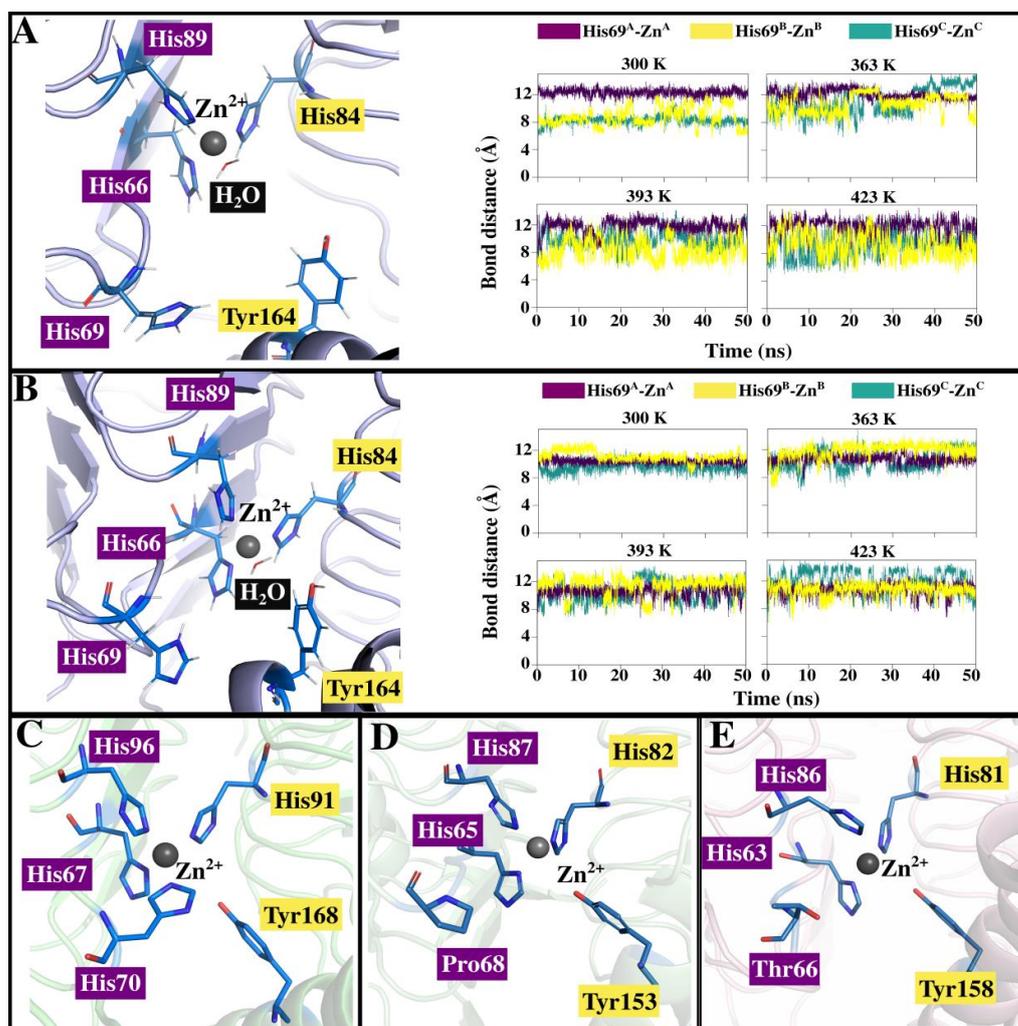
The RMSD line graphs showed that most structures had attained equilibration by 20 ns. This excluded  $\gamma$ -TtkCA at 300 K and 363 K, as well as  $\gamma$ -PhCA at 393 K. The latter sustained an RMSD between 0.17 nm and 0.2 nm at 393 K, from approximately 3 ns until around 22 ns where a gradual increase was seen before equilibration which was attained around 25 ns. This was conveyed by the KDE and violin plots as two RMSD peaks at 393 K, representing two distinct conformations visited during the simulation.  $\gamma$ -PhCA and  $\gamma$ -PmCA showed an increase in conformations visited at 363 K, indicated by the lower RMSD peaks in the KDE plots. All three different RMSD plots for  $\gamma$ -TtCA show that its structure deviations were lower than those for the other three CAs, with values below 0.15 nm being observed at 300–393 K and a slight increase at 423 K.

In the  $R_g$  line graphs (Figure 4), all four structures appear to maintain their compactness showing obscure differences across temperatures. The KDE and violin plots displayed a clearer illustration of the conformational occurrences. Some structures at higher temperatures proved to be less compact than those of lower temperature simulations. This observation was seen in the KDE and violin plots of (i)  $\gamma$ -PhCA at 393 K compared to 363 K, (ii)  $\gamma$ -PmCA at 393 K compared to 300 K and 363 K, (iii)  $\gamma$ -TtkCA at 363 K compared to 300 K and (iv)  $\gamma$ -TtkCA at 393 K compared to both 300 K and 363 K. Minimal differences in  $R_g$  were seen for  $\gamma$ -TtCA, with a notable increase at 423 K, clearly shown by a higher positioned violin plot (yellow) and KDE plot slightly shifted to the right.  $\gamma$ -TtkCA behaved slightly differently from the other three  $\gamma$ -CAs, with a higher number of conformations being visited at 300 K. The density peaks in the KDE plots for  $\gamma$ -TtCA remained more or less the same height with increase in temperature as did those for  $\gamma$ -PhCA and  $\gamma$ -PmCA for the last three temperatures.

It was interesting to note that the increase in RMSD noted in  $\gamma$ -TtkCA at 393 K (Figure 3) corresponded with the structure becoming more compact (decrease in  $R_g$ )

whereas the increase in its RMSD at 423 K corresponded with the structure becoming less compact (increase in  $R_g$ ).

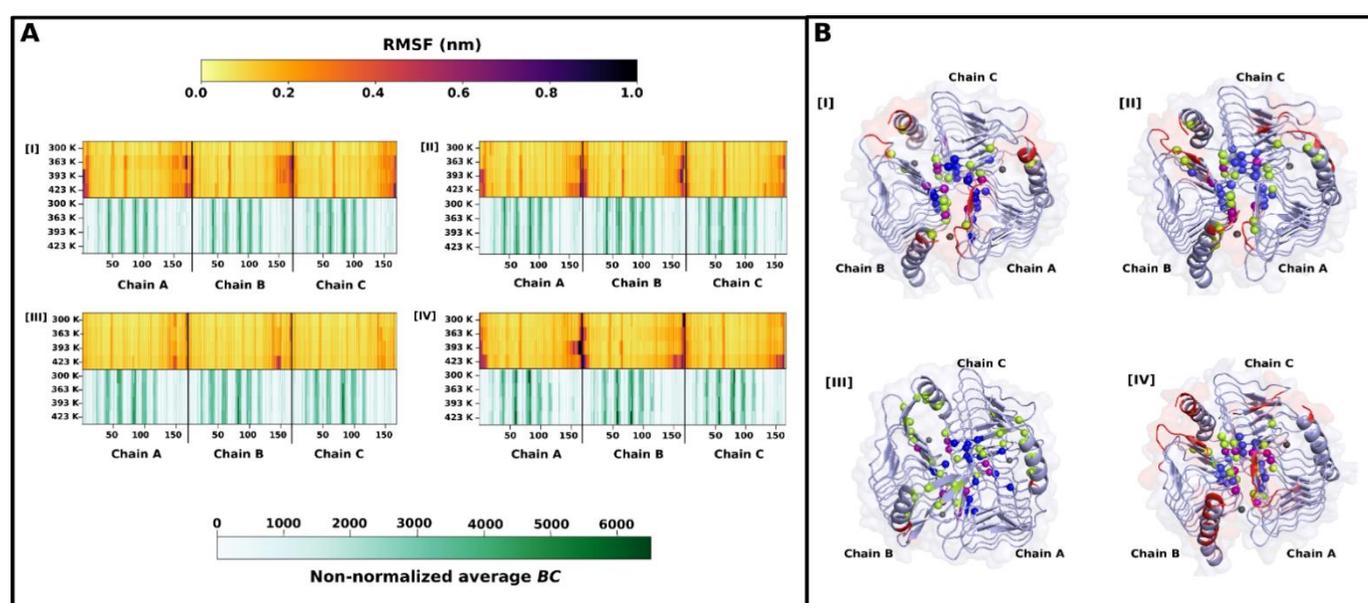
Further, the distance between  $\delta$ -nitrogen atom (ND1) of His69 and  $Zn^{2+}$  in catalytic sites for  $\gamma$ -PhCA and  $\gamma$ -PmCA (see structures in Figure 5A,B) was monitored throughout all trajectories to see if a possible “closed” conformation as the one observed in EcoCA- $\gamma$  (Figure 5C) was being exhibited. The Zn-His coordination bond distance is  $2.09 \pm 0.14 \text{ \AA}$  [50,51]. In all cases, the active site remained open with His69 staying more than 5  $\text{\AA}$  away from  $Zn^{2+}$  (see plots in Figure 5A,B). Coordination of water molecules to  $Zn^{2+}$  was also observed during simulations and is shown for  $\gamma$ -PhCA and  $\gamma$ -PmCA structures. Proximity of His69 to the active site as well as a similar spatial orientation as the one in EcoCA- $\gamma$  corroborates its role in proton transfer [41]. The loop on which EcoCA- $\gamma$ 's His70 is however longer as observed in Figure 2A compared to  $\gamma$ -PhCA and  $\gamma$ -PmCA, possibly allowing flexibility for  $Zn^{2+}$  coordination. This investigation was irrelevant for  $\gamma$ -TtCA and  $\gamma$ -TtkCA on account of the presence of Pro68 and Thr65 respectively in place of the His residue (Figure 5D,E). Unlike the non-polar Pro residue, Thr has been identified in other proton transfer pathways [52], due to its capability to donate and accept hydrogen atoms, thus this substitution in  $\gamma$ -TtkCA might have little to no effect on proton shuttling in this CA.



**Figure 5.** (A,B) show position of His69 in the active sites of  $\gamma$ -PhCA and  $\gamma$ -PmCA respectively as well as plots illustrating its bond distance from  $Zn^{2+}$  during simulations at 300 K, 363 K, 393 K and 423 K. (C) shows His70 taking up the fourth coordination position in EcoCA- $\gamma$ . (D,E) show the active site of  $\gamma$ -TtCA and  $\gamma$ -TtkCA, with Pro68 and Thr66 respectively, in place of  $\gamma$ -PhCA's His69.

### 2.2.2. RMSF and Average Betweenness Centrality (BC) Analyses

Generally, the  $\gamma$ -CA simulations displayed remarkable rigidity with increase in simulation temperatures, shown by root mean square fluctuation (RMSF) heat maps in Figure 6A. This is expected given their intricate folding with numerous beta sheets across the structures. Fluctuation was mainly seen in the beta-sheet and successive loop making up the N-terminal as well as number of residues in the C-terminal helix. This was not the case, however, for  $\gamma$ -TtCA whose N-terminals barely displayed any fluctuation compared to the other three CAs. The N-terminal for each chain is generally oriented close to the C-terminal of the neighboring monomer, and in the case for  $\gamma$ -TtCA, the C-terminal has a unique beta-sheet, extending over the N-terminal (Figure 6B). These regions are close enough for interactions, which are seemingly responsible for the reduced RMSF values. These interactions are further probed in Section 2.2.3. Beta-sheets are known to be rigid and exhibit low fluctuations. It is worth recalling that  $\gamma$ -TtCA had hotspot residues in the N-terminal (Table 2) unlike the rest of the CAs, also illustrated in Figure 6B (III). This observation accounts for  $\gamma$ -TtCA's larger BSA noted in Section 2.1.2 during interface analysis. *Betweenness centrality* (BC) quantifies how involved a residue is in the relay of information within a protein's residue interaction network. Average BC is accordingly the mean of BC values calculated for specified number of frames across a trajectory and is used as an indication of residue usage [37]. Regions that were fluctuating considerably coincided with residues exhibiting low average BC values, which are regarded as low communication residues. This is in alignment with the inverse relationship between RMSF and BC originally observed by Penkler et al. [53]. The top 5% highest communication residues with high average BC values, were picked for further analysis. These are outlined in Table 3 and are mapped onto structures of  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA and  $\gamma$ -TtkCA in Figure 6B. Location of these residues, along with other identified hotspot residues (Section 2.1.2), was observed towards the center of the trimer. A number of high communication residues coincided with interface and hotspot residues.  $\gamma$ -PhCA and  $\gamma$ -PmCA had hotspot residues toward the end of the C-terminal, specifically, Tyr164 and Tyr168. Tyr164 was previously mentioned as important in the formation of catalytic pocket and possibly in proton transfer.  $Zn^{2+}$  coordinating residues were also found appearing in all four proteins. His66 was observed in  $\gamma$ -PmCA,  $\gamma$ -TtCA, and  $\gamma$ -TtkCA, and formed ionic interaction with Asp61 of the Asp-Arg-Asp\* ionic triad.  $CO_2$  binding pocket residue Met101 appeared in one or more chains for all proteins. Met98 for  $\gamma$ -TtkCA formed hydrophobic interactions with interface residue Met84 as well as hotspot residue Ile101, both which were identified as high average BC residues in one or more chains. The residue equivalents of  $\gamma$ -TtkCA's Met84 in  $\gamma$ -PmCA and  $\gamma$ -TtCA are Met87 and Val85 respectively in the MSA, and this was also a high usage residue in both CAs, forming hydrophobic interactions with M99 ( $\gamma$ -TtCA numbering) in the  $CO_2$  binding pocket. This was not observed in the top 5% residues for  $\gamma$ -PhCA. Asn41 which was an interface residue in  $\gamma$ -PmCA and a hotspot residue in  $\gamma$ -PhCA and  $\gamma$ -TtkCA appeared with high average BC values in all three proteins.  $\gamma$ -TtCA had Gly40 in this position, which is also listed in Table 3 as a high communication residue.



**Figure 6.** (A): Heat maps of root mean square fluctuation (RMSF) (top) and average BC (bottom) of residues from I— $\gamma$ -PhCA, II— $\gamma$ -PmCA, III— $\gamma$ -TtCA and IV— $\gamma$ -TtkCA. (B): Structures of I— $\gamma$ -PhCA, II— $\gamma$ -PmCA, III— $\gamma$ -TtCA and IV— $\gamma$ -TtkCA. Hotspot residues (green), top 5% average BC residues (blue) and an intersection of the two (purple) are mapped as spheres. Regions fluctuating above 0.4 nm at 423 K were colored red. Zn<sup>2+</sup> metal ions are depicted as grey spheres.

**Table 3.** Top 5% average *betweenness centrality* (BC) residues. Interface residues are in bold text and hotspot residues are bold and italicized.

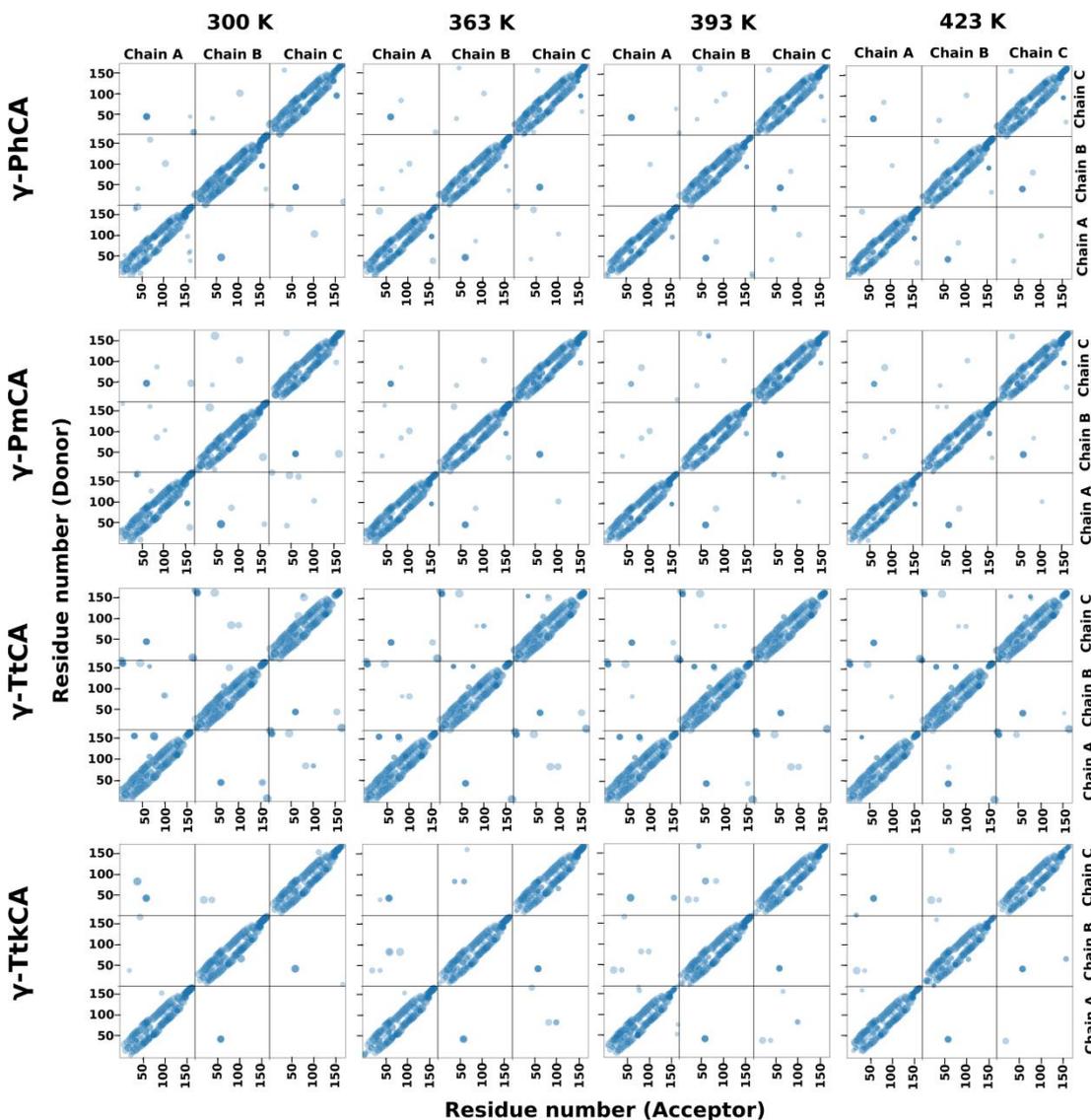
CA	Residue
$\gamma$ -PhCA	Chain A: N41, V42, <b>D61</b> , T63, <b>I64</b> , V86, H89 <sup>h</sup> , M101 <sup>c</sup> , A103, T104
	Chain B: N41, V42, T63, <b>I64</b> , <b>H84</b> <sup>h</sup> , V86, A103
	Chain C: N41, V42, <b>D61</b> , T63, <b>H84</b> <sup>h</sup> , V86, H89 <sup>h</sup> , <b>M101</b> <sup>c</sup> , A103, T104
$\gamma$ -PmCA	Chain A: N41, V42, T63, <b>I64</b> , N85, V86, A103, A119
	Chain B: N41, V42, T63, <b>I64</b> , <b>H66</b> <sup>h</sup> , <b>H84</b> <sup>h</sup> , V86, <b>M101</b> <sup>c</sup> , A103
	Chain C: N41, V42, V43, T63, I64, V82, <b>H84</b> <sup>h</sup> , V86, <b>M87</b> , T104
$\gamma$ -TtCA	Chain A: G40, N57, Q59, <b>D60</b> , A62, V63, <b>H65</b> <sup>h</sup> , <b>H82</b> <sup>h</sup> , V85, V86
	Chain B: G40, <b>D60</b> , <b>H65</b> <sup>h</sup> , <b>H82</b> <sup>h</sup> , A84, V85, <b>M99</b> <sup>c</sup>
	Chain C: G40, V42, <b>D60</b> , A62, <b>H82</b> <sup>h</sup> , A84, V85, <b>H87</b> <sup>h</sup> , V97 <sup>c</sup>
$\gamma$ -TtkCA	Chain A: N38, T39, <b>D58</b> , T60, M61, <b>H63</b> <sup>h</sup> , <b>H81</b> <sup>h</sup> , R82, A83, M84
	Chain B: T39, <b>D58</b> , L59, T60, <b>H81</b> <sup>h</sup> , A83, M84, I101
	Chain C: T39, L59, T60, M61, <b>H81</b> <sup>h</sup> , A83, M84, <b>M98</b> <sup>c</sup>

<sup>h</sup>—Active site His residue; <sup>c</sup>—CO<sub>2</sub> binding pocket residue.

### 2.2.3. Hydrogen Bond Analysis

Hydrogen bonds are important interactions in protein structures and their occurrence across the trimers and their evolution with increase in temperatures was studied using hydrogen bond analysis. These are represented as scatterplots in Figure 7. An intricate network of hydrogen bonds was observed within the individual chains for all CAs. Those in the interfaces were significantly less, a phenomenon predicted during interface analysis (Table S1, Supplementary material), and a slight decrease in bonds was observed at higher temperatures. An increase in temperature did not appear to cause a perceivable change in the intra-subunit hydrogen bonds.  $\gamma$ -TtCA had noticeably more hydrogen bonds in the interface compared to the other three CAs and  $\gamma$ -TtkCA had the least. The most prominent hydrogen bonds in all interfaces for all four CAs were those formed between Arg45<sup>ABC</sup> and Asp61<sup>BCA</sup> which were previously mentioned to form an ionic bond. Multiple bonds

between these two residues were observed across all four temperatures. In  $\gamma$ -TtCA, some of the hydrogen bonds were persistent in all three interfaces at all four temperatures. The terminals were observed to have a number of hydrogen bonds between Val3<sup>ABC</sup>-Val167<sup>BCA</sup>, Arg5<sup>ABC</sup>-Phe165<sup>BCA</sup>, Glu7<sup>ABC</sup>-Arg159<sup>BCA</sup> and Glu7<sup>ABC</sup>-Ala163<sup>BCA</sup>, which supports the RMSF results showing low fluctuations in the terminals. Another such bond was seen between hotspot residue Tyr160<sup>ACB</sup> and Asp46<sup>CBA</sup>, of which the latter was only observed as a hotspot residue in chain C. In light of the sparsity of hydrogen bonds in the interface regions, the trimers are considered to be stabilized mostly by hydrophobic interactions, most of which have been identified during interface analysis as well as average BC analysis.



**Figure 7.** Hydrogen bond plots for  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA, and  $\gamma$ -TtkCA at 300 K, 363 K, 393 K, and 423 K. Hydrogen bonds present for  $\geq 25\%$  of the simulation are illustrated as translucent blue spheres and the color intensity increase with an increase in bonds formed by those particular residues. Sphere size is directly proportional to the fraction of the simulation a hydrogen bond was present, i.e., the larger the spheres show the longer the bond was present and vice-versa.

#### 2.2.4. Dynamic Cross Correlation Analysis

Dynamic cross correlation (DCC) was utilized to observe the extent of concerted movements in the same direction across the simulations and how these are affected by temperature. Residue behavior differed from protein to protein as shown in Figure 8.  $\gamma$ -PhCA and  $\gamma$ -TtkCA displayed a decrease in correlated motions at 363 K, mostly in the

terminal regions and maintained nearly the same correlations at 393 K and 423 K.  $\gamma$ -PmCA showed the high correlated residue motions at all four temperatures, with the highest being observed at 363 K and 423 K. The behavior for  $\gamma$ -TtCA residues was quite peculiar, manifesting some anti-correlated motions between chains at 393 K which became even more pronounced at 423 K, especially for residue pairs involving C-terminal residues. A distinct increase in RMSF of residues around this region was correspondingly observed in Figure 6A (III) at 423 K. Given that the catalytic sites are located between chains, it was important that movements of the constituent residues were correlated. Further investigation revealed positive correlations between each active site's residues at both 393 K and 423 K.

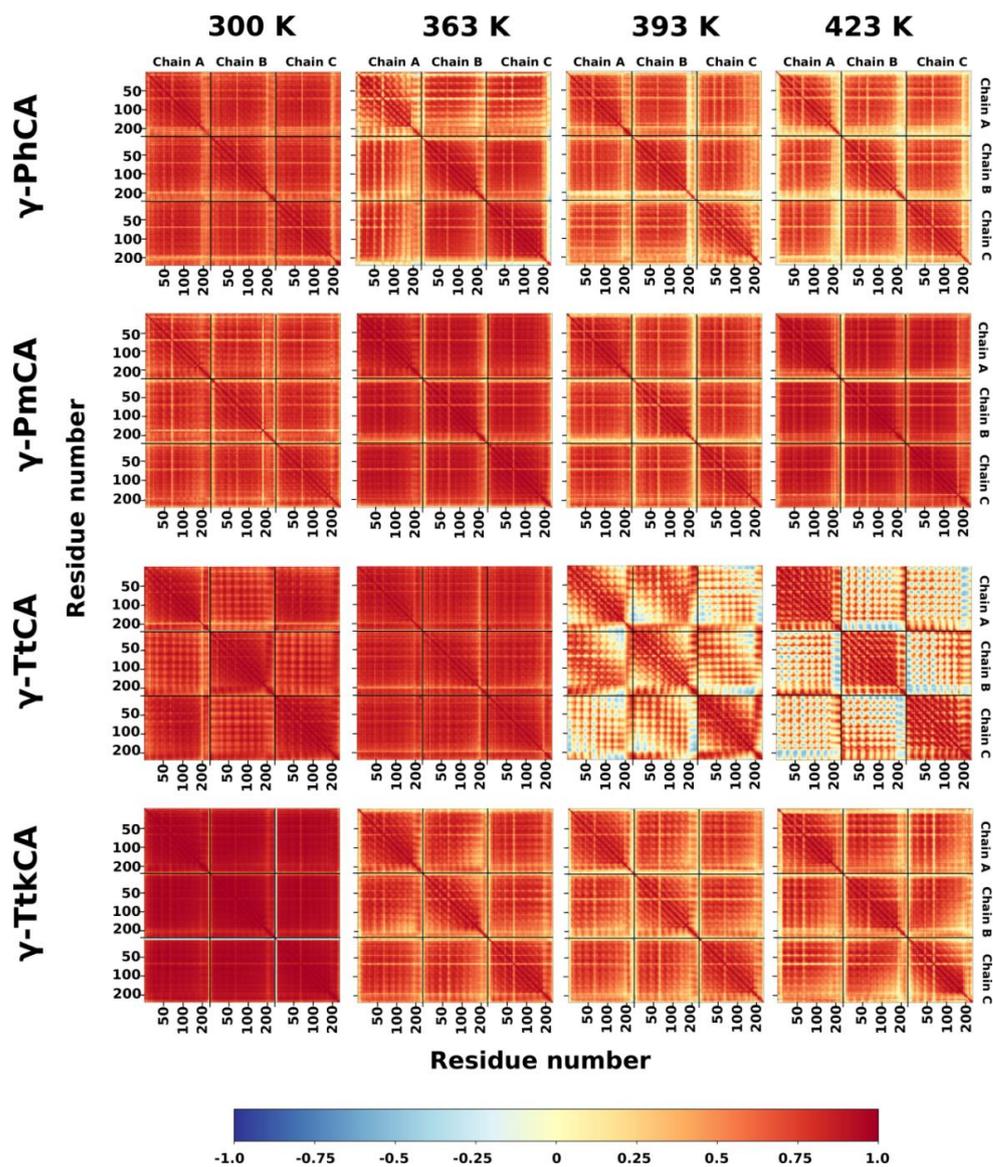
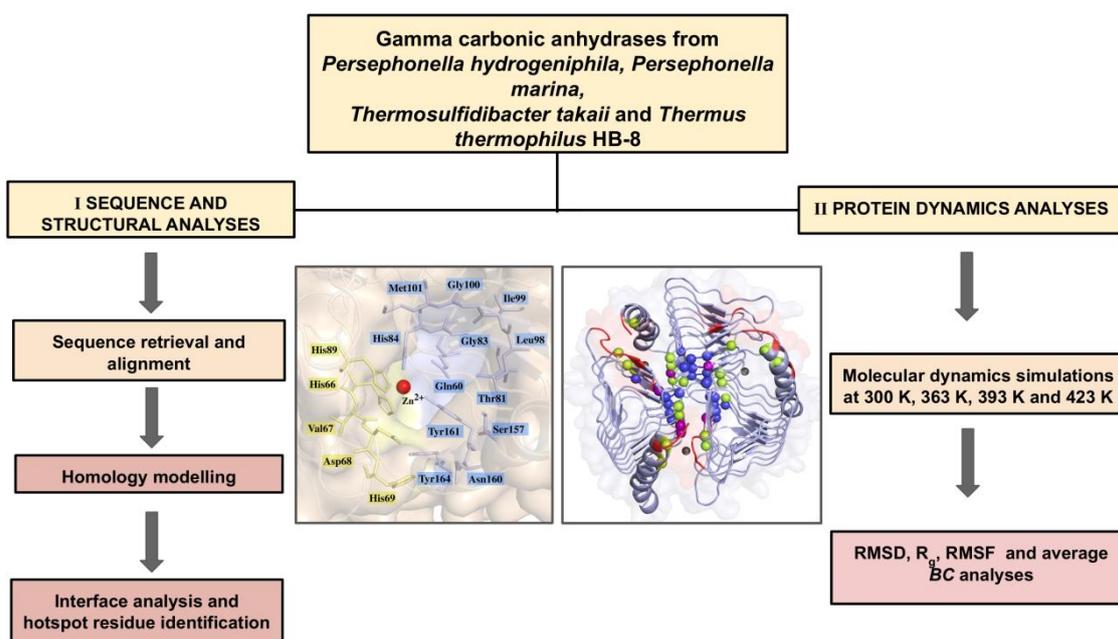


Figure 8. Dynamic cross correlation (DCC) heat maps for  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA, and  $\gamma$ -TtkCA at 300 K, 363 K, 393 K, and 423 K.

### 3. Materials and Methods

The present work comprised of sequence retrieval of four  $\gamma$ -CAs followed by static and dynamic structural analyses of these proteins. The workflow is illustrated in Figure 9.



**Figure 9.** General methodology followed in this work. RMSD,  $R_g$ , RMSF, and BC refer to root mean square deviation, radius of gyration, root mean square fluctuation and *betweenness centrality* respectively.

### 3.1. Sequence Retrieval and Alignment

The thermophiles *Persephonella hydrogeniphila* [29], *Persephonella marina* [30], *Thermosulfidibacter takaii* [31], and *Thermus thermophilus* HB8 [28] were identified through literature search. The  $\alpha$ -CA proteins from *P. hydrogeniphila* and *P. marina* have been previously investigated [32,54,55].  $\gamma$ -CA sequences for these organisms were queried and acquired from the National Center for Biotechnology Information (NCBI) database [56] and have accession numbers WP\_097000475 (*P. hydrogeniphila*,  $\gamma$ -PhCA), WP\_012675364 (*P. marina*,  $\gamma$ -PmCA) and WP\_083498668 (*T. takaii*,  $\gamma$ -TtkCA). The sequence for *T. thermophilus*' CA ( $\gamma$ -TtCA) was obtained from the Protein Data Bank (PDB) [57] with the PDB ID 6IVE. Sequences of previously characterized  $\gamma$ -CAs were also queried from the NCBI as well as the PDB and included for the multiple sequence alignment (MSA). These include  $\gamma$ -CAs from *Escherichia coli* (UniProtKB Accession number: P0A9W9.1, EcoCA- $\gamma$ ) [22], *Methanosarcina thermophila* (PDB ID:1QRL, Cam) [13,23], *Pseudomonas aeruginosa* (GenBank Reference number: QLJ92275.1, PA5540) [14], *Pyrococcus horikoshii* (PDB ID: 1V3W, Cap) [35], *Thermosynechococcus elongatus* BP-1 (PDB ID:3KWD, CcmM) [26]. PROfile Multiple Alignment with predicted Local Structures and 3D constraints (PROMALS3D) [38] was utilized to execute the MSA and to include structural information for the crystallized CAs, their respective PDB IDs and chain identities were provided. Visualization was performed using Jalview v2 [58].

### 3.2. Homology Modeling

Homology models for  $\gamma$ -PhCA,  $\gamma$ -PmCA, and  $\gamma$ -TtkCA were calculated using the  $\gamma$ -CA structure from the hyperthermophilic archaeon, *P. horikoshii* (Cap, PDB ID: 1V3W) [35]. It was identified as the most suitable template for all three  $\gamma$ -CA sequences, with template-sequence identities and coverages above 43% and 97% respectively. For each of the sequences, the slow refinement option and the automodel class in MODELLER version 9.20 [59] were used to generate 100 trimeric models, including the  $Zn^{2+}$  HETATM in the active site. Initial validation of the models was achieved using z-DOPE score calculations, with the top five showing the lowest scores proceeding for evaluation using PROCHECK [45] and Verify3D [44] web servers.  $\gamma$ -TtCA's structure (PDB ID: 6IVE) was obtained from the PDB database and used going forward.

### 3.3. Interface Analysis and Hot Spot Identification

The Protein Interactions Calculator (PIC) web server [49] was used to survey hydrophobic and ionic interactions across the trimers. Hotregion [60], HSPred [61], PPCheck [62], Protein Interfaces, Surfaces and Assemblies (PDBePISA) [48] and Robetta [36] web servers were queried for interface residues. These web servers except PDBePISA were also searched for hotspot residues. In both instances, interface and hotspot residues concurring in at least three servers were considered in this study. PyMOL v1.7.2.1 [63] was used to map these residues onto the models.

### 3.4. Molecular Dynamics Simulations

Carbonation of metals during sequestration is favored by alkaline pHs. Consequently,  $\gamma$ -PhCA,  $\gamma$ -PmCA,  $\gamma$ -TtCA, and  $\gamma$ -TtkCA were protonated at pH 8 using H++ web server [64]. Given the  $\text{Zn}^{2+}$  coordinating residues for  $\gamma$ -CAs are the same as in  $\alpha$ -CAs, previously generated and validated parameters for  $\alpha$ -CAs were inferred onto these structures prior to simulation [32,65]. Correctness of protonation states of the His residues was verified manually by checking the PDB files. His66 ( $\gamma$ -PmCA numbering) (HIE) coordinates the  $\text{Zn}^{2+}$  by the  $\delta$ -nitrogen with the  $\epsilon$ -nitrogen protonated, and the other two His (HID) residues are the reverse. *tleap* [66] from the AmberTools20 [67] package was used alongside ACPYPE [68] to generate coordinate and topology files utilizing the AMBER ff14SB force field [69]. A TIP3P cubic water box with a clearance space of 10 Å was employed in the solvation of the systems. Steepest descent minimization of the systems was performed using GROMACS v2016.1 [70], and this process was complete upon attainment of a maximum force  $< 1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . Canonical ensemble equilibration followed by isothermal-isobaric equilibration were performed for 100 ps each at four separate temperatures of 300 K, 363 K, 393 K and 423 K. MD simulations of 50 ns proceeded at these same temperatures for each  $\gamma$ -CA. These calculations as well as trajectory analyses for root mean square deviation (RMSD), radius of gyration ( $R_g$ ) and root mean square fluctuation (RMSF) were also performed using GROMACS v2016.1. Simulations were run using a total of 12,806 CPU hours on the Center for High Performance Computing (CHPC) Cape Town, South Africa. Distance between His69 ( $\gamma$ -PmCA numbering) and  $\text{Zn}^{2+}$  in each active site was monitored using Visual Molecular Dynamics (VMD) [71] and plotted in Gnuplot 5.2 [72]. A maximum bond distance of 3.5 Å was applied for hydrogen bond analysis of the trimers which was carried out using AmberTools20's *cpptraj* [73].

### 3.5. Average Betweenness Centrality Analysis

The frequency of residue usage during MD simulations was calculated using *betweenness centrality* (BC) analysis in MD-TASK (Research Unit in Bioinformatics, Makhanda, South Africa) [37]. Dynamic residue networks are constructed using  $\text{C}_\alpha$  atoms ( $\text{C}_\beta$  for Gly), regarded as nodes and an edge is created when two nodes are within a certain distance from each other. In this study the maximum distance between two nodes was specified as 6.7 Å. Using the *calc\_network.py* script, the shortest paths between two nodes in the residue interaction network were calculated for each 100th frame over the last 25 ns of each trajectory. In every frame used, BC was calculated for each residue as the total number of shortest paths going through it using the calc-BC option in the *calc\_network.py* script. The *avg\_network.py* script was then utilized to calculate the average BC for each residue from those calculated for each used frame by the *calc\_network.py* script.

### 3.6. Dynamic Cross Correlation Analysis

In order to analyze the pairwise residue correlated motions in each trimer, the displacement of  $\text{C}_\alpha$  atoms ( $\text{C}_\beta$  for Gly), from their initial position was calculated using the *calc\_correlation.py* script in MD-TASK [33]. A time step of 100 ps was specified and dynamic cross correlation (DCC) was calculated for the full 50 ns trajectories. Python scripts were used to plot heat maps from the correlation matrices produced.

#### 4. Conclusions

In order to investigate the potential use of  $\gamma$ -CAs as CO<sub>2</sub> sequestration agents as well as to achieve an understanding of the functional properties of these proteins, we applied a number of computational techniques in the analysis of four  $\gamma$ -CAs from bacteria found in high temperature environments. These bacteria included *P. hydrogeniphila*, *P. marina*, *T. takaii* and *T. thermophilus*. Alignment of these sequences with those of other known  $\gamma$ -CAs revealed low levels of conservation within the class. Variability was the most noticeable for proton transfer residues, with their absence in  $\gamma$ -TtCA evincing its lack of activity. Active site similarities between EcoCA- $\gamma$  and the two *Persephonella*  $\gamma$ -CA proteins, particularly the His residue which is involved in proton transfer, suggest that the latter two might be as active as the former  $\gamma$ -CA which has been recently characterized. This His residue was also present for the  $\gamma$ -CA from *P. horikoshii* which was also isolated from a hydrothermal vent. Dissimilar to EcoCA- $\gamma$ , however, this residue was not perceived coordinating the catalytic Zn<sup>2+</sup> during simulations in  $\gamma$ -PhCA and  $\gamma$ -PmCA. The occurrence of a “closed” active site in  $\gamma$ -CAs is thus proposed to be hinged on two features, the first being the presence of a residue capable of coordinating Zn<sup>2+</sup> residing on the loop after the first Zn<sup>2+</sup> coordinating His (His66  $\gamma$ -PmCA numbering) and the second feature being that the loop is elongated to allow flexibility of this particular residue to bond with Zn<sup>2+</sup>. This projection thus raises the question if Cam also periodically exhibits a “closed” active site, since the proton transfer residue, Glu is capable of coordinating Zn<sup>2+</sup>. Future studies could possibly utilize MD simulations to address this, given that the structure was crystallized with an open active site.

Following the modelling of trimeric structures for  $\gamma$ -PhCA,  $\gamma$ -PmCA, and  $\gamma$ -TtkCA, a hydrophobic region for CO<sub>2</sub> binding was observed in the catalytic pocket, containing a set of residues whose hydrophobicity was conserved in all sequences in the MSA. This region was clearly defined and illustrated using the structure from  $\gamma$ -PmCA. One of these hydrophobic residues, Met101, was identified as a hotspot residue during interface analysis, which was supported by average BC results identifying it as a high usage residue. Other hotspot residues which had high average BCs had functional significance, such as the Zn<sup>2+</sup> coordinating His66, some CO<sub>2</sub> binding pocket as well as proton transfer residues. This was expected considering the catalytic pocket is shared in the interface of two chains, which was dominated by hydrophobic interactions. Except for the termini regions, RMSF analysis of the simulations revealed the rigidity of the structures at all four temperatures simulated (300 K, 363 K, 393 K and 423 K), particularly in high communication residues, advocating the use of these  $\gamma$ -CAs at high temperatures. Given that  $\gamma$ -TtCA showed exceptional stability properties, the residues identified in this study that are in place of the proton transfer residues, could be mutated in future research to permit and enhance catalytic activity. By doing so,  $\gamma$ -TtCA could be an excellent option for sequestration. This in silico work concludes that  $\gamma$ -CAs from *P. hydrogeniphila* and *P. marina* are viable prospects as CO<sub>2</sub> sequestration agents. It also adds to biotechnology advances in identification of important features for catalysis and thermostability in the  $\gamma$ -CAs as catalysts for biomineralization of CO<sub>2</sub>.

**Supplementary Materials:** Supplementary Materials can be found at <https://www.mdpi.com/1422-0067/22/6/2861/s1>.

**Author Contributions:** Conceptualization, Ö.T.B.; Formal analysis, C.V.M.; Funding acquisition, Ö.T.B.; Methodology, C.V.M. and Ö.T.B.; Project administration, Ö.T.B.; Resources, Ö.T.B.; Supervision, Ö.T.B.; Visualization, C.V.M.; Writing—original draft, C.V.M.; Writing—review & editing, Ö.T.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** Colleen Varaidzo Manyumwa was funded by the DSI-CSIR Inter-bursary Support (IBS) for her PhD. Özlem Tastan Bishop was financially supported by the grant number 111212 from the National Research Foundation (NRF) of South Africa under the Iran–South Africa scientific collaboration agreement. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funder.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data generated in the present study are included in this published article. Protein models are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would like to acknowledge the Centre for High Performance Computing (CHPC), Cape Town, South Africa for computing resources.

**Conflicts of Interest:** Authors declare no conflict of interest.

## Abbreviations

BC	betweenness centrality
CA	Carbonic anhydrase
Cam	Methanosarcina thermophila gamma carbonic anhydrase
Cap	Pyrococcus horikoshii gamma carbonic anhydrase
CcmM	Thermosynechococcus elongatus BP-1 gamma carbonic anhydrase
EcoCA- $\gamma$	Escherichia coli gamma carbonic anhydrase
$\gamma$ -PhCA	Persephonella hydrogeniphila gamma carbonic anhydrase
$\gamma$ -PmCA	Persephonella marina gamma carbonic anhydrase
$\gamma$ -TtkCA	Thermosulfidibacter takaii gamma carbonic anhydrase
MD	Molecular dynamics
MSA	Multiple sequence alignment
PA5540	Pseudomonas aeruginosa gamma carbonic anhydrase
R <sub>g</sub>	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation

## References

- Pachauri, R.K.; Allen, M.R.; Barros, V.R.; Broome, J.; Cramer, W.; Christ, R.; Church, J.A.; Clarke, L.; Dahe, Q.; Dasgupta, P.; et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2014.
- Harrison, A.L.; Power, I.M.; Dipple, G.M. Accelerated carbonation of brucite in mine tailings for carbon sequestration. *Environ. Sci. Technol.* **2013**, *47*, 126–134. [[CrossRef](#)]
- Mirjafari, P.; Asghari, A.K.; Mahinpey, N. Investigating the application of enzyme carbonic anhydrase for CO<sub>2</sub> sequestration purposes. *Ind. Eng. Chem. Res.* **2007**, *46*, 921–926. [[CrossRef](#)]
- Silverman, D.N.; Lindskog, S. The catalytic mechanism of carbonic anhydrase: Implications of a rate-limiting protolysis of water. *Acc. Chem. Res.* **1988**, *21*, 30–36. [[CrossRef](#)]
- Di Fiore, A.; Alterio, V.; Monti, S.M.; De Simone, G.; D’Ambrosio, K. Thermostable carbonic anhydrases in biotechnological applications. *Int. J. Mol. Sci.* **2015**, *16*, 15456–15480. [[CrossRef](#)] [[PubMed](#)]
- Kikutani, S.; Nakajima, K.; Nagasato, C.; Tsuji, Y.; Miyatake, A.; Matsuda, Y. Thylakoid luminal  $\theta$ -carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9828–9833. [[CrossRef](#)]
- Dimario, R.J.; Machingura, M.C.; Waldrop, G.L.; Moroney, J.V. The many types of carbonic anhydrases in photosynthetic organisms. *Plant Sci.* **2018**, *268*, 11–17. [[CrossRef](#)] [[PubMed](#)]
- Tan, S.-I.; Han, Y.-L.; Yu, Y.-J.; Chiu, C.-Y.; Chang, Y.-K.; Ouyang, S.; Fan, K.-C.; Lo, K.-H.; Ng, I.-S. Efficient carbon dioxide sequestration by using recombinant carbonic anhydrase. *Process. Biochem.* **2018**, *73*, 38–46. [[CrossRef](#)]
- Jensen, E.L.; Clement, R.; Kosta, A.; Maberly, S.C.; Gontero, B. A new widespread subclass of carbonic anhydrase in marine phytoplankton. *ISME J.* **2019**, *13*, 2094–2106. [[CrossRef](#)] [[PubMed](#)]
- Coleman, J.E. Mechanism of action of carbonic anhydrase substrate, sulfonamide, and anion binding. *J. Biol. Chem.* **1967**, *242*, 5212–5219. [[CrossRef](#)]
- Silverman, D.N.; McKenna, R. Solvent-mediated proton transfer in catalysis by carbonic anhydrase. *Accounts Chem. Res.* **2007**, *40*, 669–675. [[CrossRef](#)]
- Ferry, J.G. The  $\gamma$  class of carbonic anhydrases. *Biochim. Biophys. Acta Proteins Proteom.* **2010**, *1804*, 374–381. [[CrossRef](#)]
- Alber, B.E.; Ferry, J.G. Characterization of heterologously produced carbonic anhydrase from *Methanosarcina thermophila*. *J. Bacteriol.* **1996**, *178*, 3270–3274. [[CrossRef](#)]

14. White, S. Enzymatic and Structural Analyses of PA5540: A Gamma-Carbonic Anhydrase of *Pseudomonas aeruginosa* PAO1. Ph.D. Thesis, University of Guelph, Guelph, ON, Canada, 2014. Available online: <http://hdl.handle.net/10214/7815> (accessed on 12 November 2020).
15. Parisi, G.; Perales, M.; Fornasari, M.S.; Colaneri, A.; González-Schain, N.; Gómez-Casati, D.; Zimmermann, S.; Brennicke, A.; Araya, A.; Ferry, J.G.; et al. Gamma carbonic anhydrases in plant mitochondria. *Plant Mol. Biol.* **2004**, *55*, 193–207. [[CrossRef](#)] [[PubMed](#)]
16. Enameh, R.Z.; Kuuslahti, M.; Nosrati, H.; Lohi, H.; Parkkila, S. Assessment of databases to determine the validity of  $\beta$ - and  $\gamma$ -carbonic anhydrase sequences from vertebrates. *BMC Genom.* **2020**, *21*, 1–8. [[CrossRef](#)]
17. Martin, V.; Villarreal, F.; Miras, I.; Navaza, A.; Haouz, A.; González-Lebrero, R.M.; Kaufman, S.B.; Zabaleta, E. Recombinant plant gamma carbonic anhydrase homotrimers bind inorganic carbon. *FEBS Lett.* **2009**, *583*, 3425–3430. [[CrossRef](#)]
18. Fromm, S.; Braun, H.; Peterhansel, C. Mitochondrial gamma carbonic anhydrases are required for complex I assembly and plant reproductive development. *New Phytol.* **2016**, *211*, 194–207. [[CrossRef](#)] [[PubMed](#)]
19. Peters, K.; Dudkina, N.V.; Jänsch, L.; Braun, H.-P.; Boekema, E.J. A structural investigation of complex I and I+III<sub>2</sub> supercomplex from *Zea mays* at 11–13 Å resolution: Assignment of the carbonic anhydrase domain and evidence for structural heterogeneity within complex I. *Biochim. Biophys. Acta Bioenerg.* **2008**, *1777*, 84–93. [[CrossRef](#)] [[PubMed](#)]
20. Tripp, B.C.; Bell, C.B.; Cruz, F.; Krebs, C.; Ferry, J.G. A role for iron in an ancient carbonic anhydrase. *J. Biol. Chem.* **2004**, *279*, 6683–6687. [[CrossRef](#)] [[PubMed](#)]
21. Zimmerman, S.A.; Tomb, J.-F.; Ferry, J.G. Characterization of CamH from *Methanosarcina thermophila*, founding member of a subclass of the  $\gamma$  class of carbonic anhydrases. *J. Bacteriol.* **2009**, *192*, 1353–1360. [[CrossRef](#)]
22. Del Prete, S.; Bua, S.; Supuran, C.T.; Capasso, C. *Escherichia coli*  $\gamma$ -carbonic anhydrase: Characterisation and effects of simple aromatic/heterocyclic sulphonamide inhibitors. *J. Enzym. Inhib. Med. Chem.* **2020**, *35*, 1545–1554. [[CrossRef](#)]
23. Iverson, T.M.; Alber, B.E.; Kisker, C.; Ferry, J.G.; Rees, D.C. A closer look at the active site of  $\gamma$ -class carbonic anhydrases: High-Resolution Crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* **2000**, *39*, 9222–9231. [[CrossRef](#)]
24. Del Prete, S.; De Luca, V.; Vullo, D.; Scozzafava, A.; Carginale, V.; Supuran, C.T.; Capasso, C. Biochemical characterization of the  $\gamma$ -carbonic anhydrase from the oral pathogen *Porphyromonas gingivalis*, PgiCA. *J. Enzym. Inhib. Med. Chem.* **2013**, *29*, 532–537. [[CrossRef](#)] [[PubMed](#)]
25. De Araujo, C.; Arefeen, D.; Tadesse, Y.; Long, B.M.; Price, G.D.; Rowlett, R.S.; Kimber, M.S.; Espie, G.S. Identification and characterization of a carboxysomal  $\gamma$ -carbonic anhydrase from the cyanobacterium *Nostoc* sp. PCC 7120. *Photosynth. Res.* **2014**, *121*, 135–150. [[CrossRef](#)]
26. Peña, K.L.; Castel, S.E.; De Araujo, C.; Espie, G.S.; Kimber, M.S. Structural basis of the oxidative activation of the carboxysomal  $\gamma$ -carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2455–2460. [[CrossRef](#)] [[PubMed](#)]
27. Soltés-Rak, E.; Mulligan, M.E.; Coleman, J.R. Identification and characterization of a gene encoding a vertebrate-type carbonic anhydrase in cyanobacteria. *J. Bacteriol.* **1997**, *179*, 769–774. [[CrossRef](#)] [[PubMed](#)]
28. Wang, W.; Zhang, Y.; Wang, L.; Jing, Q.; Wang, X.; Xi, X.; Zhao, X.; Wang, H. Molecular structure of thermostable and zinc-ion-binding  $\gamma$ -class carbonic anhydrases. *BioMetals* **2019**, *32*, 317–328. [[CrossRef](#)]
29. Nakagawa, S.; Takai, K.; Horikoshi, K.; Sako, Y. *Persephonella hydrogeniphila* sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium from a deep-sea hydrothermal vent chimney. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 863–869. [[CrossRef](#)]
30. Götz, D.; Banta, A.; Beveridge, T.J.; Rushdi, A.I.; Simoneit, B.R.T.; Reysenbach, A.L. *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 1349–1359. [[CrossRef](#)]
31. Nunoura, T.; Oida, H.; Miyazaki, M.; Suzuki, Y. *Thermosulfidibacter takaii* gen. nov., sp. nov., a thermophilic, hydrogen-oxidizing, sulfur-reducing chemolithoautotroph isolated from a deep-sea hydrothermal field in the Southern Okinawa Trough. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 659–665. [[CrossRef](#)]
32. Manyumwa, C.V.; Enameh, R.Z.; Tastan Bishop, Ö. Alpha-carbonic anhydrases from hydrothermal vent sources as potential carbon dioxide sequestration agents: In silico sequence, structure and dynamics analyses. *Int. J. Mol. Sci.* **2020**, *21*, 8066. [[CrossRef](#)]
33. Parra-Cruz, R.; Jäger, C.M.; Lau, P.L.; Gomes, R.L.; Pordea, A. Rational design of thermostable carbonic anhydrase mutants using molecular dynamics simulations. *J. Phys. Chem. B* **2018**, *122*, 8526–8536. [[CrossRef](#)] [[PubMed](#)]
34. Parra-Cruz, R.; Lau, P.L.; Loh, H.-S.; Pordea, A. Engineering of *Thermovibrio ammonificans* carbonic anhydrase mutants with increased thermostability. *J. CO<sub>2</sub> Util.* **2020**, *37*, 1–8. [[CrossRef](#)]
35. Jeyakanthan, J.; Rangarajan, S.; Mridula, P.; Kanaujia, S.P.; Shiro, Y.; Kuramitsu, S.; Yokoyama, S.; Sekar, K. Observation of a calcium-binding site in the  $\gamma$ -class carbonic anhydrase from *Pyrococcus horikoshii*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64*, 1012–1019. [[CrossRef](#)] [[PubMed](#)]
36. Kortemme, T.; Kim, D.E.; Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. Signal.* **2004**, *2004*, pl2. [[CrossRef](#)] [[PubMed](#)]
37. Brown, D.K.; Penkler, D.L.; Amamuddy, O.S.; Ross, C.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: A software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **2017**, *33*, 2768–2771. [[CrossRef](#)] [[PubMed](#)]

38. Pei, J.; Kim, B.-H.; Grishin, N.V. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300. [[CrossRef](#)]
39. Tripp, B.C.; Smith, K.; Ferry, J.G. Carbonic anhydrase: New insights for an ancient enzyme. *J. Biol. Chem.* **2001**, *276*, 48615–48618. [[CrossRef](#)]
40. Kisker, C.; Schindelin, H.; Alber, B.E.; Ferry, J.G.; Rees, D.C. A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J.* **1996**, *15*, 2323–2330. [[CrossRef](#)]
41. Park, H.-M.; Park, J.-H.; Choi, J.-W.; Lee, J.; Kim, B.Y.; Jung, C.-H.; Kim, J.-S. Structures of the  $\gamma$ -class carbonic anhydrase homologue YrdA suggest a possible allosteric switch. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 920–926. [[CrossRef](#)]
42. Rowlett, R.S. Structure and catalytic mechanism of the  $\beta$ -carbonic anhydrases. *Biochim. Biophys. Acta Proteins Proteom.* **2010**, *1804*, 362–373. [[CrossRef](#)]
43. Aspatwar, A.; Kairys, V.; Rala, S.; Parikka, M.; Bozdog, M.; Carta, F.; Supuran, C.T.; Parkkila, S. Mycobacterium tuberculosis  $\beta$ -carbonic anhydrases: Novel targets for developing antituberculosis drugs. *Int. J. Mol. Sci.* **2019**, *20*, 5153. [[CrossRef](#)]
44. Eisenberg, D.; Lüthy, R.; Bowie, J.U. VERIFY3D: Assessment of protein models with three-dimensional profiles. In *Methods in Enzymology*; Abelson, J., Simon, M., Verdine, G., Pyle, A., Eds.; Elsevier: Amsterdam, The Netherlands, 1997; Volume 277, pp. 396–404.
45. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. [[CrossRef](#)]
46. Alexander, R.S.; Nair, S.K.; Christianson, D.W. Engineering the hydrophobic pocket of carbonic anhydrase II. *Biochemistry* **1991**, *30*, 11064–11072. [[CrossRef](#)]
47. West, D.; Kim, C.U.; Tu, C.; Robbins, A.H.; Gruner, S.M.; Silverman, D.N.; McKenna, R. Structural and kinetic effects on changes in the CO<sub>2</sub> binding pocket of human carbonic anhydrase II. *Biochemistry* **2012**, *51*, 9156–9163. [[CrossRef](#)] [[PubMed](#)]
48. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774–797. [[CrossRef](#)]
49. Tina, K.G.; Bhadra, R.; Srinivasan, N. PIC: Protein interactions calculator. *Nucleic Acids Res.* **2007**, *35*, W473–W476. [[CrossRef](#)] [[PubMed](#)]
50. Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc coordination spheres in protein structures. *Inorg. Chem.* **2013**, *52*, 10983–10991. [[CrossRef](#)]
51. Tamames, B.; Sousa, S.F.; Tamames, J.; Fernandes, P.A.; Ramos, M.J. Analysis of zinc-ligand bond lengths in metalloproteins: Trends and patterns. *Proteins Struct. Funct. Bioinform.* **2007**, *69*, 466–475. [[CrossRef](#)] [[PubMed](#)]
52. Song, W.J.; McCormick, M.S.; Behan, R.K.; Sazinsky, M.H.; Jiang, W.; Lin, J.; Krebs, C.; Lippard, S.J. Active site threonine facilitates proton transfer during dioxygen activation at the diiron center of toluene/o-Xylene monooxygenase hydroxylase. *J. Am. Chem. Soc.* **2010**, *132*, 13582–13585. [[CrossRef](#)]
53. Penkler, D.L.; Atilgan, C.; Tastan Bishop, Ö. Allosteric modulation of human Hsp90 $\alpha$  conformational dynamics. *J. Chem. Inf. Model.* **2018**, *58*, 383–404. [[CrossRef](#)] [[PubMed](#)]
54. Kim, S.; Sung, J.; Yeon, J.; Choi, S.H.; Jin, M.S. Crystal structure of a highly thermostable  $\alpha$ -carbonic anhydrase from *Persephonella marina* EX-H1. *Mol. Cells* **2019**, *42*, 460–469. [[PubMed](#)]
55. 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<sub>2</sub>-sequestration applications. *Process. Biochem.* **2014**, *49*, 2114–2121. [[CrossRef](#)]
56. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bolton, E.; Bryant, S.H.; Canese, K.; Chetvermin, V.; Church, D.M.; DiCuccio, M.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2010**, *39*, D38–D51. [[CrossRef](#)]
57. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]
58. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)] [[PubMed](#)]
59. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* **2006**, *15*, 5.6.1–5.6.30. [[CrossRef](#)] [[PubMed](#)]
60. Cukuroglu, E.; Gürsoy, A.; Keskin, O. HotRegion: A database of predicted hot spot clusters. *Nucleic Acids Res.* **2011**, *40*, D829–D833. [[CrossRef](#)] [[PubMed](#)]
61. Lise, S.; Buchan, D.; Pontil, M.; Jones, D.T. Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS ONE* **2011**, *6*, e16774. [[CrossRef](#)]
62. Sukhwal, A.; Sowdhamini, R. PPCheck: A webserver for the quantitative analysis of protein-protein interfaces and prediction of residue hotspots. *Bioinform. Biol. Insights* **2015**, *9*, BBI.S25928. [[CrossRef](#)]
63. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.
64. Gordon, J.C.; Myers, J.B.; Foltz, T.; Shoja, V.; Heath, L.S.; Onufriev, A. H++: A server for estimating pK<sub>a</sub>s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–W371. [[CrossRef](#)]
65. Sanyanga, T.A.; Nizami, B.; Tastan Bishop, Ö. Mechanism of action of non-synonymous single nucleotide variations associated with  $\alpha$ -carbonic anhydrase II deficiency. *Molecules* **2019**, *24*, 3987. [[CrossRef](#)] [[PubMed](#)]
66. Schafmeister, C.; Ross, W.; Romanovski, V. *LEaP Analysis Tools*; University of California: San Francisco, CA, USA, 1995.
67. Case, K.D.A.; Belfon, I.Y.; Ben-Shalom, S.R.; Brozell, D.S.; Cerutti, T.E.; Cheatham, V.W.D., III; Cruzeiro, T.A.; Darden, R.E.; Duke, G.; Giambasu, G.; et al. *AMBER 2020*; University of California: San Francisco, CA, USA, 2020.

68. Da Silva, A.W.S.; Vranken, W.F. ACPYPE—Antechamber python parser interface. *BMC Res. Notes* **2012**, *5*, 367. [[CrossRef](#)] [[PubMed](#)]
69. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)] [[PubMed](#)]
70. 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*, 19–25. [[CrossRef](#)]
71. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
72. Williams, T.; Kelley, C.; Broker, H.B.; John Campbell, R.; Cunningham, D.; Denholm, G.; Elber, R.; Fearick, C.; Grammes, L.; Hart, L.; et al. Gnuplot 5.2. 2: An Interactive Plotting Program. 2017. Available online: <http://www.gnuplot.info> (accessed on 20 January 2020).
73. Roe, D.R.; Cheatham, I.T.E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [[CrossRef](#)]