Gamma Carbonic Anhydrases from Hydrothermal Vent Bacteria: Cases of Alternating Active Site Due to a Long Loop with Proton Shuttle Residue

Colleen Varaidzo Manyumwa and Özlem Tastan Bishop

Abstract: Accelerated CO$_2$ sequestration uses carbonic anhydrases (CAs) as catalysts; thus, there is much research on these enzymes. The γ-CA from *Escherichia coli* (EcoCA-γ) was the first γ-CA to display an active site that switches between “open” and “closed” states through Zn$^{2+}$ coordination by the proton-shuttling His residue. Here, we explored this occurrence in γ-CAs from hydrothermal vent bacteria and also the γ-CA from *Methanosarcina thermophila* (Cam) using molecular dynamics. Ten sequences were analyzed through multiple sequence alignment and motif analysis, along with three others from a previous study. Conservation of residues and motifs was high, and phylogeny indicated a close relationship amongst the sequences. All structures, like EcoCA-γ, had a long loop harboring the proton-shuttling residue. Trimeric structures were modeled and simulated for 100 ns at 423 K, with all the structures displaying thermostability. A shift between “open” and “closed” active sites was observed in the 10 models simulated through monitoring the behavior of the His proton-shuttling residue. Cam, which has two Glu proton shuttling residues on long loops (Glu62 and Glu84), also showed an active site switch affected by the first Glu proton shuttle, Glu62. This switch was thus concluded to be common amongst γ-CAs and not an isolated occurrence.

Keywords: gamma carbonic anhydrase; Campylobacteria; *Methanosarcina thermophila*; hydrothermal vents; carbon dioxide sequestration; betweenness centrality

1. Introduction

Carbonic anhydrases (CAs) are fast catalysts for the interconversion of carbon dioxide (CO$_2$) to bicarbonate ions (HCO$_3^-$) [1] and, for this reason, have been considered as ideal CO$_2$ sequestration agents. The discovery of CAs emanated in the late 1980s with the first CA having been identified from bovine blood. Since then, eight different CA classes have been discovered, and these include the α-, β-, γ-, δ-, η-, ζ-, and θ-CAs and ι-CAs [2–7]. They contain a catalytic metal ion in the active site, which is usually zinc (Zn$^{2+}$), but it has also been observed to be iron (Fe$^{2+}$) or cadmium (Cd$^{2+}$) in some γ-CAs [8–12]. γ-CAs, which are one of eight known classes, exhibit a unique trimeric biological assembly unlike in the α- and β-CAs, which are dimeric and tetrameric, respectively. Three active sites are found in the γ-CAs, with each one found between two monomers as seen in the structure of the γ-CA from the ubiquitous bacterium *Escherichia coli* (EcoCA-γ) (Figure 1A). Each monomer is composed of a long helix at the C-terminal and beta-helical prism [1,8]. Similar to the α-CAs, coordination of the Zn$^{2+}$ metal ion is tetrahedral and fulfilled by three His residues and a water molecule [8,13,14]. Unlike the α-CAs, however, two of the His residues are from one chain and the third comes from the neighboring chain. The Zn-H$_2$O is converted to Zn-OH after a proton is shuttled away by a proton-shuttling residue, normally, a His or a Glu in γ-Cas [15,16]. The resulting nucleophile attacks the CO$_2$ molecule during catalysis in the CO$_2$-binding pocket close to the active site, resulting in the formation of a bicarbonate ion [17].
The γ-CA from the ubiquitous bacterium Escherichia coli (EcoCA-γ), which has two crystal structures with Protein Data Bank (PDB) IDs of 3TIO and 3TIS, emerged as the first γ-CA to possess an active site capable of alternating between a “closed” state and an “open” state [18] (Figure 1), an occurrence common in β-CAs [19–22]. This switch in EcoCA-γ is made possible by a His residue responsible for the shuttling of a proton from the Zn-bound water molecule during catalysis. The residue is positioned close to the catalytic site on a loop that is 11 residues long. The His residue occasionally coordinates the Zn$^{2+}$, obstructing the active site opening, as shown in Figure 1C.

EcoCA-γ has been reported to maintain high catalytic activities with a $k_{cat}$ of $5.7 \times 10^5$ s$^{-1}$ [23]. In our previous study, we monitored the switch from “open” to “closed” state via in silico studies in some γ-CAs belonging to the family Aquificae from hydrothermal vents [24]. Their incapability to occupy the latter state was observed to be a result of the short loop on which the His proton shuttle residue resided. These results motivated the present study whose main objective was to monitor MD simulations of γ-CAs with a longer loop, a feature previously suggested to be necessary for this active site switch. The characteristics referred to above include a residue capable of Zn$^{2+}$ coordination (His, Glu, Cys, Asp), which is located on a relatively long loop that is flexible enough for purposes of the residue moving in and out of the active site region.

In this study, we mainly focused on γ-CAs from the class Campylobacteria (previously known as Epsilonproteobacteria), which are found in hydrothermal vent regions [25,26], and we followed similar approaches as in [24]. Hydrothermal vents are high-temperature regions that contain organisms adapted to these temperatures. Their γ-CAs are consequently contemplated to be thermostable, a characteristic that is desirable for the harsh conditions presented by the CO$_2$ sequestration process, thus their investigation in this study [2]. The Campylobacteria γ-CAs were selected because of the presence of a longer loop with the proton-shuttling His residue compared to those from the family Aquificae. The protein sequences were retrieved and aligned along with those from our previous study [24]. Their 3D structures were calculated using homology modeling, and residues that contribute to interface formation were probed using various web servers. Most of these residues were later identified as high communication residues using average betweenness centrality (BC) analysis [27] following molecular dynamics (MD) simulations. Both the multiple sequence alignment (MSA) and motif analysis revealed high residue and motif conservation across the sequences, respectively. The CAs proceeded for MD simulations, which were also performed for EcoCA-γ as a control. Additional simulations were performed for the
previously characterized γ-CA from *Methanosarcina thermophila* (Cam) to discern whether a change in active site would occur when the proton-shuttling residue is a Glu [28,29]. Simulations were performed at 423 K considering the hydrothermal vent origin of the retrieved γ-CAs as well as the elevated temperatures applied during CO₂ sequestration. Interestingly, monitoring of the active site revealed the alternation of states in all the γ-CAs, including the control EcoCA-γ validating the aforementioned hypothesis. Further, the stability of most of the γ-CA structures at the high temperature used, particularly EcoCA-γ and the γ-CAs from *Lebetimonas natsushimae* (γ-LnCA) and *Nitratiruptor tergarcus* (γ-NtCA), was observed via radius of gyration (Rg), root mean square deviation (RMSD), root mean square fluctuation (RMSF), and dynamic cross correlation (DCC).

This study hereby revealed the first group of γ-CAs, other than EcoCA-γ, that exhibits the two active site states by reason of a nearby proton-shuttling His residue. It also demonstrates this same phenomenon in the well-studied Cam, whose proton shuttling residue is a Glu unlike the His in the rest of the γ-CAs in this study. Our analysis also suggests new viable CO₂ sequestration agents.

2. Results and Discussion

A total of 10 γ-CA sequences from hydrothermal vent bacteria belonging to the class Campylobacteria were retrieved from NCBI (Table 1). The sequence lengths had a range between 174 and 179 residues with the query sequence from *Caminibacter mediatlanticus* (γ-CmCA) showing high sequence identities above 54% to the retrieved sequences. The γ-CmCA had sequence coverages greater than 96%, and significant E-values (below 1 × 10⁻³) were observed (Table 1). The numbering of residues was done according to γ-CmCA except in cases explicitly declared otherwise.

| CA Abbreviation | Organism Name | Organism Taxonomic Family | CA Accession Number | CA Sequence Length | Query Sequence | Query Coverage | Sequence Identity | E-Value   |
|-----------------|---------------|---------------------------|---------------------|-------------------|----------------|----------------|------------------|-----------|
| γ-CmCA          | *Caminibacter mediatlanticus* | Nautiliaceae | WP_007473220.1 | 179 aa | γ-CmCA | 100% | 100% | 9 × 10⁻¹²⁹ |
| γ-EpCA          | Epsilonproteobacteria bacterium 4484_65 | – | OQX76055.1 | 174 aa | γ-CmCA | 97% | 68.6% | 4 × 10⁻⁸⁴ |
| γ-HtCA          | Hydrogenimonas thermophila | Hydrogenimonaceae | WP_092914162.1 | 178 aa | γ-CmCA | 99% | 69.1% | 1 × 10⁻⁸⁸ |
| γ-LnCA          | *Lebetimonas natsushimae* | Nautiliaceae | WP_096258534.1 | 179 aa | γ-CmCA | 100% | 91.1% | 3 × 10⁻¹¹⁷ |
| γ-NpCA          | *Nautilia profundicola* | Nautiliaceae | WP_012663737.1 | 179 aa | γ-CmCA | 100% | 88.3% | 1 × 10⁻¹¹⁴ |
| γ-NsbCA         | *Nitratiruptor sp. SB155-2* | Nitratiruptoraceae | WP_012081704.1 | 178 aa | γ-CmCA | 99% | 69.1% | 2 × 10⁻⁸⁷ |
| γ-NtCA          | *Nitratiruptor tergarcus* | Nitratiruptoraceae | WP_084274664.1 | 178 aa | γ-CmCA | 100% | 71.5% | 1 × 10⁻⁹² |
| γ-SaCA          | *Sulfurimonas autotrophica* | Thiouvlcaceae | WP_013326385.1 | 174 aa | γ-CmCA | 97% | 70.7% | 4 × 10⁻⁸⁶ |
| γ-SIca          | *Sulfurovum lithothrophicum* | Sulfurovacea | WP_046550225.1 | 174 aa | γ-CmCA | 97% | 68.4% | 4 × 10⁻⁸⁴ |
| γ-SNbcCA        | *Sulfurovum sp. NBC37-1* | Sulfurovacea | WP_012083921.1 | 174 aa | γ-CmCA | 97% | 68.4% | 1 × 10⁻⁸³ |

This section is divided into three subsections, which are (i) sequence analyses, (ii) structural analysis, and (iii) molecular dynamics analyses. Previously investigated hydrothermal vent γ-CA sequences from *Persephonella hydrogeniphila* (γ-PhCa), *Persephonella marina* (γ-PmCA) and *Thermosulfidibacter takaii* (γ-TtkCA) were also included in the sequence analyses.
but not in the other two as those have already been published [24]. Structure and MD analyses included crystal structures from Cam and EcoCA-γ, with the latter being used as a control in active site analysis following the simulations.

2.1. Sequence Analysis

2.1.1. Conserved Regions and Motifs Are Identified Using Multiple Sequence Alignment and Motif Analysis

Sequences that are related normally have functional residues, regions, and motifs that are conserved, and these were probed in this section using sequence alignments and motif analysis. The γ-CAs from the hydrothermal vents showed high residue and motif conservation across the sequences (Figure 2A,B, respectively). Out of the nine motifs found, seven of them were completely conserved (Figure 2B). Their positions in each γ-CA as well as their E-values are outlined in Table S1 and shown on the alignment in Figure 2A following γ-SlCA numbering given that it was the structure used to map the motifs. γ-TtkCA showed a significant difference from the other sequences by the absence of Motifs 8 and 9, the latter of which was also absent in γ-PhCA and γ-PmCA. Zn$^{2+}$-coordinating residues, His64, located on Motif 4, His90, and His95, both on Motif 1, were completely conserved in the MSA as expected. This also was true for the residues that contribute to the catalytic site, including Arg45 and Asp47 found on Motif 3 as well as Gln60 and Asp61 present on Motif 4 [1,10,16]. The hydrophobic CO$_2$ pocket previously profiled [24] was also observed as a well-conserved region in Motif 1 across all sequences. The most relevant observation for this study in the Campylobacteria sequences was the presence of His67 on an elongated loop, which was either 14 or 15 residues long and which is shown by the black box in Figure 2A. This observation is dissimilar to the members of Aquificae (γ-PhCA, γ-PmCA, and γ-TtkCA), which had a shorter loop of seven residues each. This similarity to EcoCA-γ (Figure S1), which is known to shift between “open” and “closed” active sites because of the His on a long loop, suggests that this phenomenon might also be possible for the Campylobacteria γ-CAs. His67 was found on Motif 4, which formed part of the long loop. The end of this loop was found on Motif 9, which was absent in γ-PhCA, γ-PmCA, and γ-TtkCA due to their shorter loop. Overall, MSAs and motif analysis showed high levels of conservation across the hydrothermal vent γ-CA sequences.

2.1.2. Evolutionary Relationships amongst the γ-CAs Are Investigated Using Phylogeny

The phylogeny of the sequences was inferred using the constructed tree in Figure 3 and was used to discover their evolutionary relationships. Amongst the nine trees calculated, the best tree that matched its respective consensus tree, was that generated using the LeGascuel protein model [30] with discrete gamma distribution (LGG) with 100% gap deletion. Given the high residue conservation observed in the alignment, a close relationship was deduced amongst these γ-CAs. This was confirmed by the pairwise sequence identities calculated from the MSA used for the phylogenetic tree calculations. The lowest percentage identity observed was 52%, which is considerably high. Two main branches were observed, which separated γ-CAs from the class Campylobacteria from those from Aquificae. γ-TtkCA was observed as an outgroup, which was contemplated since it had relatively lower sequence identities to all the other γ-CAs. γ-CmCA, γ-LnCA, and γ-NpCA had sequence identities above 88% and formed a subgroup in the evolutionary tree. This was anticipated given that all three come from the same family: Nautiliacea. γ-EpCA was 91.3% similar to γ-SlCA as well as γ-SNbcCA, and the branch bootstrap value indicated that, 940 out of 1000 times, the same branch was observed. It is from this that we instigated that the unclassified bacterium *Epsilonproteobacteria bacterium* 4484_65 belongs to the Sulfurovaceae family and possibly the genus *Sulfurovum*.
γ-TtkCA), which had a shorter loop of seven residues each. This similarity to EcoCA-γ (Figure S1), which is known to shift between "open" and "closed" active sites because of the His on a long loop, suggests that this phenomenon might also be possible for the Campylobacteria γ-CAs. His67 was found on Motif 4, which formed part of the long loop. The end of this loop was found on Motif 9, which was absent in γ-PhCA, γ-PmCA, and γ-TtkCA due to their shorter loop. Overall, MSAs and motif analysis showed high levels of conservation across the hydrothermal vent γ-CA sequences.

Figure 2. Multiple sequence alignment (MSA) and motif analysis of the hydrothermal γ-CA sequences. (A) The MSA was calculated using the T-Coffee alignment program, and residues that were more than 50% conserved are colored according to the extent of residue conservation. Residues boxed in blue are Zn$^{2+}$-coordinating histidines, and those with a black dot above them are important for catalysis. The black box delineates the loop containing the proton-shuttling His residue, indicated by the cyan triangle, while residues in the green box are CO$_2$-binding pocket residues. The lines under the alignment show the positions of the motifs, with the colors corresponding to those used in (C). Start and end positions for the γ-CA from Sulfurovum lithotrophicum (γ-SlCA) were used. (B) Heatmap showing the prevalence of motifs across the sequences. (C) Structure of the γ-SlCA colored according to motif positions. The spheres show residues identified in the interface. Hotspot residues are shown as red arrows on the motif logos generated by MEME, and black arrows show the interface residues. The black dotted circle shows the elongated loop in the Campylobacteria γ-CAs.
In the second tree constructed with 17 other sequences from organisms that are found in other environments (Figure S2), most sequences from hydrothermal vents clustered together showing the close relationship amongst them. However, γ-TtkCA, although it showed high sequence identities to the hydrothermal vent sequences, clustered with γ-CAs from the Clostridium genus bacteria, which are mesophilic. The γ-CA from the crystallized *Pyrococcus horikoshii* (PDB ID: 1V3W, Cap) [15], also from a hydrothermal vent, was seen clustering with that from *Brucella abortus*, which has also been crystallized with PDB ID: 4N27 [31]. They had a considerably high sequence identity of 72%. This tree revealed a distant evolutionary relationship between the hydrothermal vent γ-CAs and Cam, supported by the low sequence identities.

2.2. Structure Analysis

2.2.1. Interface and Hotspot Residues Are Probed in the Trimeric Structures

All structures modeled had good z-DOPE scores below −1.32 and passed the 80% threshold in Verify3D [32] as shown in the validation results in Table S2. These results, along with those from ProSA [33] and PROCHECK [34], reflected native-like structures. The γ-CAs were modeled as the generic trimers, whose monomers are folded into a prism of short beta-sheets with loops in between them and a long C-terminal helix, which spans the length of the prism. Structures were viewed in PyMOL [35] and superimposition with EcoCA-γ, which confirmed the elongated loop similarity observed in the sequence alignment (Figure S3). Cam’s structure was also superimposed with these structures as well as γ-PhCA and γ-PmCA (Figure S3), and the two main differences observed were the longer loop containing the proton-shuttling residue (Glu84, Cam numbering) as well as a long loop in the N-terminal, which was not present in all the other structures. An additional α-helix, absent in other structures, was also seen just before the start of the C-terminal long helix. For all subsequent experiments, crystal structures for Cam (PDB ID: 1QRG) and EcoCA-γ (PDB ID: 3TIO) were used.

The structure of the γ-CA is folded in such a way that two of the three sides of each monomeric prism have residues involved in the formation interface of the trimer. This gives rise to a considerable interface area, as observed in the results from PDBePISA [36], which revealed that between 18% and 26% of the total surface area is the interface (Table S3). The residues that form interactions in the interface were identified using various programs and are listed in Table 2, along with hotspot residues. Hotspot residues play a more significant role in the stability of the interface, which destabilizes in the event of their mutation to Ala [37]. Residues identified as hotspots whose function is known include the previously...
Table 2. Residues common to three out of five programs identified as participating in interface formation. Hotspot residues are in bold.

| CA          | Residues                                                                 |
|-------------|---------------------------------------------------------------------------|
| Interface AB |                                                                            |
| Chain A:    | F8, N10, R12, E13, N14, P15, V16, T17, P18, W19, P36, Q37, A38, S39, I41, M55, A56, S57, R59, D61, E62, G77, V79, H81, L83, E84, E88, Q118, S119, Q120, H122, Q136, A137, F138, F140, R153, S154, A155, I157, G170, V172, T174 |
| Chain B:    | F8, Y33, D35, P36, Q37, S53, P54, M55, Q75, D76, G77, H117, Q118, M135, Q136, P152, R153, A169, G170, Y191, T194, A197, V198, V201, N202, H204, L205, G208, Y209, E211, T212 |
| Interface BC |                                                                            |
| Chain A:    | F8, N10, R12, E13, N14, P15, V16, T17, P18, W19, P36, Q37, A38, S39, I41, M55, A56, S57, R59, D61, E62, G77, V79, H81, L83, E84, E88, Q118, S119, Q120, H122, Q136, A137, F138, F140, R153, S154, A155, I157, G170, V172, T174 |
| Chain C:    | F8, Y33, D35, P36, Q37, S53, P54, M55, Q75, D76, G77, H117, Q118, M135, Q136, P152, R153, A169, G170, Y191, T194, A197, V198, V201, N202, H204, L205, G208, Y209, E211, T212 |
| Interface DE |                                                                            |
| Chain D:    | D22, D23, S24, W40, P41, L42, D62, G63, H91, K92, M108, G109, A126, G127, N167, Y168, W171, E174, V175 |
| Chain E:    | E6, P7, Y8, R9, S24, V26, L42, V44, R46, D48, V49, G63, M65, H67, H70, K92, V93, M94, H96, I111, L113, L129, S145 |
| Interface EF |                                                                            |
| Chain E:    | D22, S24, W40, P41, L42, Q61, D62, G63, H91, K92, M108, G109, A126, G127, S64, N167, Y168, W171, E174, V175, Q178 |
| Chain F:    | P7, Y8, R9, S24, V26, L42, V44, R46, D48, V49, G63, M65, H67, H70, K92, V93, M94, H96, I111, L113, L129, S145 |
| Interface DF |                                                                            |
| Chain D:    | D22, S24, W40, P41, L42, D62, G63, H91, K92, M108, G109, A126, G127, N167, Y168, W171, E174, V175, Q178 |
| Chain F:    | D22, S24, W40, P41, L42, D62, G63, H91, K92, M108, G109, A126, G127, N167, Y168, W171, E174, V175, Q178 |
| Interface AC |                                                                            |
| Chain A:    | I2, L3, R4, Y5, K6, D7, D23, J25, G39, V41, R43, D45, V46, M60, S61, S62, H64, H67, K69, K70, R91, V92, M93, H95, S108, A109, T110, L112, G126, L128 |
| Chain B:    | P20, S21, W37, F38, G39, D39, M60, H90, R91, M107, S108, A125, G126, L142, P145, Q162, K165, N166, Y167, N169, Y170, D173, Y174, F177, I178, K179 |
| Chain B:    | I2, L3, R4, Y5, K6, S21, D23, J25, G39, V41, R43, D45, V46, M60, S61, S62, H64, H67, K69, K70, R91, V92, M93, H95, S108, A109, T110, L112, D113, S128, F144 |
| Interface AC |                                                                            |
| Chain A:    | S21, W37, F38, G39, D39, M60, H90, R91, M107, S108, A125, L142, S144, P145, Q162, K165, N166, Y167, Y170, D173, Y174, F177, I178, K179 |
| Chain C:    | M1, I2, L3, R4, Y5, K6, D7, N8, F9, S21, D23, J25, G39, V41, R43, D45, V46, M60, S61, S62, H64, H67, K70, R91, V92, M93, H95, A109, T110, H111, L112, D113, L128, Q131, S144, P145 |
| CA | Residues |
|---|---|
| γ-EpCA | Interface AB |
| Chain A: L3, K4, F5, K6, S23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, H67, R90, V91, M92, H94, S107, T109, L111, I110, D112, L117, S143, P144 |
| Chain B: W37, F38, G39, D59, L60, M62, H89, R90, M106, S107, A124, G125, M141, G142, R165, Y166, F169, E172, Y173 |
| γ-HiCA | Interface AB |
| Chain A: L2, L3, R4, Y5, K6, D7, P9, T23, I25, G39, V41, R43, D45, V46, L60, S61, V92, M93 H95, R91 M62, H64, H67, S108, A109, T110, L112, L127, S143, P144 |
| Chain B: P20, D21, W37, F38, D59, L60, H89, R90, M106, N107, A124, I141, R165, Y166, F169, D172, Y173, N175, F176, L177, S178 |
| γ-LnCA | Interface BC |
| Chain A: L2, L3, R4, Y5, K6, D7, D23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, H67, Y68, K69, R90, V91, M92, H94, N107, C108, T109, L111, D112, L127, S143, P144 |
| Chain B: M1, L2, R3, Y5, K6, A22, T23, I25, G39, V41, R43, D45, V46, L60, T61, M62, H64, T66, H67, Y68, K69, R90, V91, M92, H94, N107, C108, T109, L111, D112, L127, S143, P144 |
| γ-NpCA | Interface BC |
| Chain A: P20, S21, W37, F38, G39, Q58, D59, M60, H90, R91, M107, S108, A125, G126, L142, I143, K162, N166, Y167, Y170, D173, Y174, F177, V178, R179 |
| Chain B: M1, T2, L3, R4, Y5, K6, A22, D23, I25, G39, V41, R43, D45, V46, M60, S61, M62, H64, T66, H67, E69, K70, R91, V92, M93, H95, S108, A109, T110, L112, D113, S144, P145 |
| Chain C: T2, L3, R4, Y5, K6, D23, I25, G39, V41, R43, D45, V46, M60, S61, H64, T66, H67, E69, R91, V92, M93, H95, S108, A109, T110, L112, D113, G126, L128, S144, P145 |
| CA    | Residues | Interface AB                  | Interface BC                  | Interface AC                  | Interface AC                  | Interface AC                  |
|-------|----------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| γ-NsbCA | Chain A: L2, L3, R4, Y5, K6, F9, T23, I25, G39, V41, R43, D45, V46, L60, T61, M62, H64, T66, H67, Y68 R90, V91, M92, H94, N107, S108, T109, L111, D112, G125, L127, S143, P144 | Chain B: L2, L3, R4, Y5, K6, E7, F9, T23, I25, G39, V41, R43, D45, V46, L60, 61, M62, H64, V65, T66, H67, Y68, R90, V91, M92, H94, N107, S108, T109, L111, D125, L127, S143, P144 | Chain C: P20, D21, W37, F38, G39, D59, L60, H89, R90, M106, A124, G125, L141, G142, E158, R165, Y166, F169, E172, Y173, N175, S176, I177, S178 | Chain A: W17, P20, D21, W37, F38, G39, D59, L60, H89, R90, M106, N107, I121, A124, G125, G142, E154, E155, E158, A161, R165, Y166, F169, E172, Y173, S176, I177, S178 | Chain C: L2, L3, R4, Y5, K6, F9, T23, I25, G39, V41, R43, D45, V46, L60, T61, M62, H64, T66, H67, Y68, K70, R90, V91, M92, H94, N107, S108, T109, L111, L127, S143 |
| γ-Nica | Chain A: L2, L3, R4, Y5, K6, F9, T23, V25, G39, V41, R43, D45, V46, L60, S61, M62, H64, T66, H67, K69, R90, V91, M92, H94, N107, S108, T109, L111, D112, L127, T143, P144 | Chain B: M1, L2, L3, R4, Y5, K6, E7, F9, T23, V25, G39, V41, R43, D45, V46, L60, S61, M62, H64, T66, H67, Y68, R90, M92, H94, N107, S108, T109, L111, D112, G125, L127, T143, P144 | Chain C: P20, D21, A22, W37, F38, G39, D59, L60, H89, R90, M106, N107, A124, G125, L141, G142, E158, R165, Y166, F169, E172, Y173, F176, I177, Q178 | Chain A: P20, D21, W37, F38, G39, D59, L60, H89, R90, M106, N107, A124, G125, G142, K154, K164, R165, Y166, F169, E172, Y173, F176, I177, Q178 | Chain C: L2, L3, R4, Y5, K6, F9, T23, V25, G39, V41, R43, D45, V46, L60, S61, M62, H64, T66, H67, K69, P71, D72, R90, V91, M92, H94, N107, T109, L111, D112, G125, L127, T143 |
| γ-Sica | Chain A: H3, V4, Y5, K6, D23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, T66, H67, Y68, D75, R90, V91, M92, H94, N107, A108, T109, L111, H12, G125, L127, K130, T143, P144 | Chain B: P20, S21, W37, F38, G39, D59, L60, H89, R90, M106, S107, A124, G125, M141, G142, E158, A161, S162, R165, Y166, F169, E172, Y173 | Chain C: P20, S21, W37, F38, G39, D59, L60, H89, R90, M106, S107, A124, G125, M141, G142, R165, Y166, F169, E172, Y173 | Chain A: P20, S21, W37, F38, G39, D59, L60, H89, R90, M106, N107, A124, G125, M141, G142, R165, Y166, F169, D172, Y173 | Chain C: H3, V4, Y5, K6, D23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, V65, T66, H67, R90, V91, M92, H94, S107, A108, T109, L111, D112, G125, L127, T143, P144 |
| γ-Sica | Chain A: M1, L3, K4, F5, K6, S23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, V65, T66, H67, K69, R90, V91, M92, H94, S107, A108, T109, L111, D125, G127, K130, S143, P144 | Chain B: E20, G21, W37, F38, G39, D59, L60, H89, R90, M106, S107, A124, G125, M141, E158, A161, S162, R165, Y166, F169, E172, Y173 | Chain C: M1, L3, K4, F5, K6, S23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, H67, R90, V91, M92, H94, S107, A108, T109, L111, D112, G125, L127, K130, S143, P144 | Chain A: E20, G21, W37, F38, G39, D59, L60, H89, R90, M106, N107, A124, G125, M141, E158, R165, Y166, F169, E172, Y173 | Chain C: M1, L3, K4, F5, K6, S23, I25, G39, V41, R43, D45, V46, L60, S61, M62, H64, H67, R90, V91, M92, H94, S107, A108, T109, L111, D112, G125, L127, K130, S143, P144 |
2.3. Molecular Dynamics Analysis

2.3.1. Residue Fluctuation and Functionality of High Communication Residues Is Investigated

This section aimed to analyze residue fluctuation patterns during the simulations as well as to utilize BC in discovering residues important in protein communication. On account of the inverse relationship between BC and RMSF [39], these two metrics were jointly plotted together for all the γ-CAs as heatmaps in Figure 4. Generally, the trimeric structures showed high levels of rigidity. This was expected given the abundance of beta sheets in the prisms. Highly flexible regions (shown by the darker regions) were observed in the termini mostly for the γ-CAs γ-HtCA, and γ-NpCA as well as γ-NsbCA. They were also observed in the N-termini of Cam, γ-EpCA, as well as both termini in all chains of EcoCA-γ, γ-SICa, and γ-SNbcCA. The loop containing the proton-shuttling His residue (seen by the darker regions between residues 50 and 100) had distinctly high RMSF values in all the γ-CAs. All the flexible regions mentioned predictably coincided with low average BC regions. Residues that have a high average BC (shown by the darker regions in the average BC heatmaps) are known to be central in the residue network and have a considerably high amount of communication through them. Those with high fluctuations are not capable of functioning as a continuous central point in the network, thus they normally exhibit low average BC values. However, flexibility is a prerequisite for the operation of some functional residues, such as the proton-shuttling His in CAs. After narrowing down to the 5% residues with the highest average BC in each protein, these residues were plotted as heatmaps (Figure 5). As expected, His67 was not part of the list for all the proteins. Some functionally important residues, however, were identified. Most of these were identified in Section 2.2.1 as residues in the interface, highlighted in green and a considerable number were hotspot residues, highlighted in red. One or more Zn$^{2+}$-coordinating histidines (shown by the yellow dots, Figure 5) were observed as high communication residues across all γ-CAs, with the highest number being observed in EcoCA-γ. CO$_2$-binding pocket residue Met107, which was a hotspot residue in some γ-CAs, also had high average BC values. Overall, the functions of most high BC residues were successfully annotated and correlated to low fluctuating residues.
exhibit low average $BC$ values. However, flexibility is a prerequisite for the operation of some functional residues, such as the proton-shuttling His in CAs. After narrowing down to the 5% residues with the highest average $BC$ in each protein, these residues were plotted as heatmaps (Figure 5). As expected, His67 was not part of the list for all the proteins. Some functionally important residues, however, were identified. Most of these were identified in Section 2.2.1 as residues in the interface, highlighted in green and a considerable number were hotspot residues, highlighted in red. One or more $\text{Zn}^{2+}$-coordinating histidines (shown by the yellow dots, Figure 5) were observed as high communication residues across all $\gamma$-CAs, with the highest number being observed in EcoCA-$\gamma$. CO$_2$-binding pocket residue Met107, which was a hotspot residue in some $\gamma$-CAs, also had high average $BC$ values. Overall, the functions of most high $BC$ residues were successfully annotated and correlated to low fluctuating residues.

Figure 4. Root mean square fluctuation (RMSF) (top) and normalized average betweenness centrality ($BC$) (bottom) heatmaps of the $\gamma$-CAs simulated.
Figure 5. A: Heatmaps of top 5% average BC residues for all γ-CAs simulated. Residues highlighted in green are interface residues, and those highlighted red are hotspot residues. Yellow dots indicate Zn$^{2+}$-coordinating His residues.

2.3.2. Conformational Changes Are Assessed Using RMSD and R$_g$ Analysis

In order to monitor the conformational changes that occur in the γ-CAs during the MD simulations, radius of gyration (R$_g$) and root mean square deviation (RMSD) analyses were utilized and are represented using three different types of plots in Figures 6 and 7. Monitoring the evolution of these metrics using line graphs was done in order to observe when the structures attained equilibrium during the simulation as well as to see any distinct changes in R$_g$ and RMSD. KDE and violin plots were plotted to view the distribution of structures.
with different $R_g$ and RMSD values across the simulations. All $\gamma$-CAs from hydrothermal vents showed a stable RMSD, which was equilibrated for most of the simulation. This was anticipated at this temperature given that the hydrothermal vents they come from are high-temperature environments. Interestingly, the RMSD line graph for Cam showed a significant increase at 75 ns and then equilibrated up to 100 ns. Cam is moderately thermostable showing stability up to 55 °C [28]. However, unlike these organisms, _E. coli_ is a mesophilic bacterium, but EcoCA-$\gamma$ was observed to exhibit the most stable RMSD compared to the rest of the structures. Compactness was well-maintained for the $\gamma$-CAs throughout the simulations. The observations for both $R_g$ and RMSD were expected due to the rigidity of the trimeric structures.

**Figure 6.** Root mean square deviation (RMSD) kernel density estimation (KDE) plots (top), line plots (middle), and violin plots (bottom) of $\gamma$-CAs simulated. The mean RMSD for each $\gamma$-CA is depicted as a similar colored line to the KDE plot and as a black line in the violin plots.
2.3.3. A Conformational Switch in Proton Shuttling His Close to the Active Site Is Observed

The mobility of the proton-shuttling residue in CAs was surmised due to the rotation towards and away from the active site during proton shuttling. However, coordination proximity to the metal ion and occlusion of the active site in \(\gamma\)-CAs was not contemplated until the crystallization of EcoCA-\(\gamma\). This switch was investigated in this section for all the simulated \(\gamma\)-CAs. Figure 8 shows the distances between the His residue and Zn\(^{2+}\) throughout the CA simulations performed. It was interesting to observe that in the Campylobacteria \(\gamma\)-CAs simulated, the proton-shuttling His was observed in one or more active sites to be in a coordination distance of 2.2 Å to the metal ion in a number of frames during the simulations [40]. EcoCA-\(\gamma\), which was used as a control for these observations also showed Zn\(^{2+}\) coordination by its His69, as seen in the crystal structures. It was further observed in the results for Cam that one of the Glu residues responsible for proton shuttling, Glu62, indeed was in coordination distance to the Zn\(^{2+}\) during the simulation. This corroborates with previous studies, which have indicated that Glu62 relays protons to Glu84 during catalysis [8]. Crystallized structures have shown this distance to be approximately 2.1 ± 0.3 Å [40]. Cam’s Glu84, which was found in the position equivalent of His67 in the Campylobacteria \(\gamma\)-CAs as well as EcoCA-\(\gamma\) was not
the Glu residue coordinating the Zn\(^{2+}\) during the simulation. Glu84 was instead observed maintaining a distance of at least 4 Å away from the metal ion (Figure S4) as indicated previously [8]. Overall, these results reinforce the findings from crystallization of EcoCA-γ. They also support the hypothesis that the presence proton-shuttling residue on a long loop in proximity to the catalytic site in γ-CAs indicate their ability to switch between “open” and “closed” states.

![Figure 8](image-url)  
**Figure 8.** Zn\(^{2+}\) coordination by a proton-shuttling residue. (A) Visualized frames of each γ-CA showing the fourth coordination position being fulfilled by His67 (Glu62 for Cam). Residues from chain A are highlighted in salmon, chain B in blue, and chain C in green. Zn\(^{2+}\) is shown as a grey sphere and the Zn-His (Zn-Glu for Cam) distance is indicated in the black box in the bottom right corner of each structure. (B) Bond distances between Zn and the proton-shuttling residue across the 100 ns trajectories. The red triangle shows the frame from which the respective structure in (A) was obtained.

2.3.4. High Residue Correlations Are Perceived Using Dynamic Cross Correlation Analysis

Dynamic cross correlation is an analysis used to investigate whether the pairwise movement of residues in a protein is correlated (in the same direction) or anti-correlated (in opposite directions). Residue correlated motions were investigated using dynamic cross correlation (DCC) analysis, and the matrices were plotted as heatmaps in Figure 9. Given the extent of trimer rigidity observed during RMSF analysis (Section 2.3.1) as well as the high number of residues identified in the interface, high residue correlated motions across the structures were expected. For all structures, reduced correlations were observed between residues in the loop containing the proton-shuttling His residue and those in the rest of the structure (shown by the lighter coloring between residues 50 and 100). This observation was
supported by the low communication going through these residues in average BC analysis as well as their high fluctuations. Less concerted movements were also observed in the termini regions, which showed high RMSF values in Section 2.3.1. DCC analysis for EcoCA-γ was corroborated with RMSF analysis, which showed low residue correlations across the protein. γ-HtCA also showed relatively lower positive correlated motions compared to the rest of the proteins, possibly implying a lower thermostability similar to EcoCA-γ. The highest correlations were observed for Cam and γ-LnCA as well as γ-NtCA, and the latter two are presumed to be as thermostable as the former if not more thermostable. DCC was therefore able to function as an additional analysis to assess γ-CA thermostability.

Figure 9. Dynamic cross correlation (DCC) heatmaps for γ-CAs simulated.
3. Materials and Methods

3.1. Sequence Retrieval and Alignment

The γ-CA from *Caminibacter mediatlanticus* was queried from NCBI and used based on the thermostability and efficiency of its α-CA, which has been previously studied [7,41]. One iteration of the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) in NCBI was used to obtain a list of homologous γ-CAs. CAs from this list were confirmed through literature sources to have been isolated from hydrothermal vents. The NCBI taxonomy site was used to check the taxonomic class of each of these narrowing down the list to γ-CAs from the bacterial class, Campylobacteria.

The 10 CAs selected were from *Caminibacter mediatlanticus* [42], *Hydrogenimonas thermophila* [43], *Lebetimonas natsushimae* [44], *Epsilonproteobacteria bacterium 4484_65* [45], *Nautilia profundiola* [46], *Nitratiruptor tergarcus* [47], *Sulfurovum lithotrophicum* [48], *Sulfurimonas autotrophica* [49,50], *Sulfurovum sp. NBC 37-1* [51,52], and *Nitratiruptor sp. SB155-2* [51,52].

3.2. Sequence Alignments and Motif Analysis

The γ-CA sequences for hydrothermal vent organisms *Persephonella marina*, *Persephonella hydrogeniphila*, and *Thermosulfidibacter takaii*, which were studied previously [24], were included in the sequence alignments as well as the motif analysis. The alignment programs used were ClustalOmega [53], multiple alignment using fast Fourier transform (MAFFT) FFT-NS-i [54], as well as tree-based consistency objective function for alignment evaluation (T-Coffee) [55,56] following which the result showing the best alignment of conserved residues was chosen. Conserved short repeating patterns, termed motifs, were searched across these γ-CA sequences using MEME [57] and filtered using an E-value threshold of 0.05. Those with E-values higher than the threshold were discarded.

3.3. Phylogenetic Analysis

For purposes of evaluating the evolutionary relationships amongst the retrieved sequences as well as *Persephonella marina*, *Persephonella hydrogeniphila*, and *Thermosulfidibacter takaii* γ-CAs, phylogenetic trees were constructed using the MEGA v7 [58]. Using the unaligned sequences as input, protein models for tree calculations were generated. Tree calculations proceeded using the first three models that showed the lowest Bayesian Information Criterion (BIC) values calculated using three gap deletions of 90%, 95%, and 100%. These models were the Whelan and Goldman model with gamma distribution (WAGG), LGG, and the Whelan and Goldman model with gamma distribution and invariant sites (WAGGI) with BIC values of 4040.1, 4043.3, and 4047.7, respectively, for the 100% and 95% gap deletions. For the 90% deletion, BIC values of 4069.0, 4073.8, and 4076.8 were recorded for the WAGG, LGGI, and WAGGI, respectively [30,59].

For each protein model, three trees were generated with different gap deletions of 90%, 95%, and 100%, respectively. The bootstrap method was used with a total of 1000 bootstraps for each tree. The initial tree was constructed using the neighbor joining (NJ/BioNJ) method, and the nearest-neighbor interchange (NNI) maximum likelihood statistical method was applied. Trees were viewed in Dendroscope v3.5.9 [60], and the tree which matched its respective consensus tree was chosen as the best.

Another tree was constructed with the following 17 additional sequences (making a total of 30 sequences) with their respective NCBI accession numbers: *Arabidopsis thaliana* (NP_001185039.1) [61], *Brucella abortus* (WP_002964382.1) [31], *Desulfurobacterium indicum* (GenBank Reference number: OMH40674.1), *Escherichia coli* (UniProtKB Accession number: P0A9W9.1) [18], *Clostridium difficile* (CAJ69635.1), *Clostridium senegalese* (WP_010293151.1), *Clostridium hydrogeniformans* (WP_027632952.1), *Haloarchaeon halobacterium* (KXA95168.1) [62], *Methanosarcina thermophila* (PDB ID:1QRL) [29], *Porphyromonas gingivalis* (WP_021665403.1), *Pseudomonas aeruginosa* (GenBank Reference number: Q1J92275.1), *Pyrococcus horikoshii* (PDB ID: 1V3W) [15], *Salmonella enterica* (PDB ID: 3R3R) (unpublished, Center for Structural Genomics of Infectious Diseases), *Serratia* sp. ISTD04 (ODJ21043.1), *Thermosynechococcus elongatus* BP-1 (PDB ID:3KWD) [63], *Thermus thermophilus* (PDB ID: 6IVE) [64], and *Vibrio*
cholerae (WP_053032778.1). The top three models for this tree were LGG, LGGI, and WAGGI, and tree construction proceeded as described above.

3.4. Homology Modeling, Interface, and Hotspot Residue Identification

Template identification for purposes of homology modeling was carried out for all retrieved sequences without structures using PRIMO [65]. All sequences were at least 40% identical to the template, the γ-CA from Pyrococcus horikoshii (PDB ID: 1V3W, Cap) [15], and the template covered over 97% of the sequences. Modeling of the γ-CAs in their trimeric biological assemblies was performed using MODELLER v9.20 [66], following which the structures were validated using various programs. The first structure validation scoring used was the z-DOPE score. Structures with more negative scores below a threshold of −0.5 were considered more nativelike and for each sequence, the top 5 with the lowest z-DOPE scores proceeded for validation with PROCHECK [34], ProSA [33] and Verify3D [32] web servers. For Cam and EcoCA-γ, the crystal structures were retrieved from the PDB and used for subsequent calculations.

The interfaces of the proteins were probed for residues that contribute to stability of the trimers using five different web servers. Each γ-CA was individually queried in HotRegion [67]; the Knowledge-based Fade and Contacts (KFC) server [68]; PPCheck [69]; Protein Interfaces, Surfaces, and Assemblies (PDBePISA) [36]; and Robetta [37] web servers. Residue consideration for the interface was contingent on it appearing in at least three out of the five servers. HotRegion, KFC, PPCheck, and Robetta were further queried for hotspot residues, and a residue was considered if it appeared in at least three web server results.

3.5. Molecular Dynamics Simulations

Protein preparation for MD simulations involved protonation of the structures at pH 8 using H++ web server [70] as well as inferring previously generated and validated Zn$^{2+}$ parameters to ensure its maintenance in the active site [71]. An alkaline pH is required for metal carbonate precipitation during CO$_2$ sequestration, justifying the pH used in this study. The conversion of topology and coordinate files from AMBER- to GROMACS-compatible files was achieved by tleap [74] from the AmberTools20 [72]. The systems were solvated using the TIP3P cubic water box with a 10 Å box boundary, and the AMBER ff14SB force field [75] was applied. Minimization was performed using the steepest descent method, where termination was effectuated when the system reached a maximum force below 1000 kJ mol$^{-1}$ nm$^{-1}$. Canonical ensemble (NVT) and isothermal/isobaric (NPT) equilibration were performed one after the other at 423 K using a 2 fs time step for 100 ps each. Simulations were then run on CHPC clusters using GROMACS v2016.1 [73] for 100 ns at 423 K for each protein, using up a total of 27,565 CPU hours. GROMACS was also used for calculation of $R_g$ and RMSD as well as RMSF for all trajectories. Visual molecular dynamics (VMD) [76] were utilized to monitor the bond distances between Zn$^{2+}$ and the proton-shuttling His residue (Glu residues in Cam) in all active sites for all trajectories, and Gnuplot v5.2 [77] was used to plot the distances as a function of time.

3.6. Average Betweenness Centrality Analysis

The dynamic residue networks (DRNs) in a protein are created using nodes, which are the residues and edges, which connect one node within a particular distance to another. For each protein, DRNs were generated in MD-TASK [27] using a 10-ps time step interval in the last 15 ns of the 100 ns trajectories. The C$_\alpha$ atoms (C$_\beta$ for Gly) were utilized as nodes, and these were required to be within 6.7 Å of each other for it to be considered an edge. The shortest path between two residues/nodes was calculated using the calc_network.py script in MD-TASK for every residue pair, and in each frame, the number of shortest paths going through each node were summed up to provide it with a betweenness centrality (BC) value. The higher the BC, the more the paths passing through the node thus the more central it is in the DRN. BC values for each node were averaged using the avg_network.py script giving rise to the average BC, which was normalized between 0 and 1 for each γ-CA.
3.7. Dynamic Cross Correlation Analysis

MD-TASK [27] was utilized in the monitoring of the residue motion correlations in the γ-CAs using the `calc_correlation.py` script. Calculations were performed using the Ca atoms (Cβ for Gly) for each residue over the 100 ns trajectories, generating pairwise correlation matrices. These were plotted as heatmaps using Python scripts.

4. Conclusions

The intriguing discovery of an alternative conformation of the active site in the γ-CA from *E. coli* motivated the present study, whose main aim was to detect a similar occurrence in γ-CAs from hydrothermal vent systems. The experiments around this investigation were built on the postulation that two features be found present in a γ-CA. These are (i) a proton-shuttling residue close to the active site and (ii) a long, flexible loop on which the proton shuttle residue resides. Similar to previously investigated γ-PhCA, γ-PmCA, and γ-TtkCA, a proton-shuttling residue was found in the Campylobacteria γ-CAs, but, unlike the three γ-CAs, the retrieved γ-CAs had a longer loop of between 14 and 15 residues. This was observed in the MSA analysis showing high conservation across the sequences, which were closely related, as witnessed in the evolutionary tree as well as the sequence identity heat map. Cam, a well-characterized and crystallized γ-CA, also had the above-mentioned features. Following structure calculations, the retrieved γ-CAs as well as Cam and EcoCA-γ were simulated at 423 K, and, as anticipated, a transitioning from “open” to “closed” was observed in one or more chains in all the proteins. Zn²⁺ coordination by the first proton-shuttling residue, Glu62 in Cam was unexpected given that Glu84 was the proton-shuttling residue close to the active site and (ii) a long, flexible loop on which the proton-shuttling residue resides. Similar to previously investigated

Further research involving in vitro testing of the thermostability properties of EcoCA-γ through the discovery of more γ-CAs using the phylogenetic tree constructed for γ-CAs from hydrothermal vent systems. The experiments around this investigation was motivated the present study, whose main aim was to detect a similar occurrence in γ-CAs at the elevated temperature of 423 K was monitored post-simulation using Rg, RMSD, RMSF as well as DCC analyses. EcoCA-γ displayed remarkable potential thermostability in all the analyses closely followed by γ-NtCA and γ-LnCA. The rest of the γ-CAs were relatively stable, although γ-HtCA and γ-NpCa showed lowered thermostability particularly with lowered residue correlations and slightly higher fluctuations in DCC and RMSF analyses, respectively. Motif analysis enabled the discovery of nine highly conserved motifs, eight of which harbored hotspot and interface residues. The importance of these residues was further substantiated by average BC analysis of the trajectories, where most of the top 5% residues were found in the interface. Zn²⁺-coordinating residues were also found as high communication residues. Overall, this work contributes to the knowledge pool of CAs through the discovery of more γ-CAs that exhibit an alternating active site like EcoCA-γ. Further research involving in vitro testing of the thermostability properties of EcoCA-γ as well as the simulated hydrothermal vent γ-CAs as prospective CO₂ sequestration agents would be highly crucial.

**Supplementary Materials:** The following are available online at https://www.mdpi.com/article/10.390/catal11101177/s1. Figure S1: Multiple sequence alignment of γ-CAs from *Methanosarcina thermophila* (Cam, PDB ID: 1QRL), *Pyrococcus horikoshii* (Cap, PDB ID: 1V3W), *Escherichia coli* (EcoCA-γ, UniProtKB Accession number: P0A9W9.1), *Caminitabacter mediatlanticus* (γ-CmCA), *Epsilonproteobacteria bacterium 4484* (γ-EPCA), *Hydrogenimonas thermophila* (γ-HtCA), *Lebetimonas natsushimae* (γ-LnCA), *Nautilia profundicola* (γ-NpCA), *Nitratiruptor* sp. SB155-2 (γ-NsbCA) *Nitratiruptor tergaricus* (γ-NtCA), *Persephonella hydrogeniphila* (γ-PhCA), *Persephonella marina* (γ-PmCA), *Sulfurimonas autotrophica* (γ-SaCA), *Sulfurovum lithotrophicum* (γ-SICA), *Sulfurovum* sp. NBC 37-1 (γ-SNBCA), *Thermus thermophilus* (γ-TtCA, PDB ID: 6IVE) and *Thermosulfidibacter takaii* (γ-TtkCA). The magenta box shows the variable loop containing the His proton shuttle. The first cyan triangle indicates proton shuttling residue Glu62 in Cam, and the second cyan shows the second proton shuttling residue Glu84 in Cam and His in the rest of the γ-CAs except γ-TtCA and γ-TtkCA, Figure S2: A: Multiple sequence alignment of γ-CAs from hydrothermal vent organisms as well as γ-CAs from other environments used to create the phylogenetic tree in B. The black box shows the variable loop containing the His proton shuttle. The red circles show the Zn²⁺ coordinating residues and the green circle shows the position of the His proton shuttle. Sequences from hydrothermal vents are labelled in red. B: Phylogenetic tree constructed for γ-CA sequences from hydrothermal vent organisms as well as γ-CAs.
from other environments, alongside the pairwise sequence identity heatmap. The Maximum Likelihood method was applied using the LGGI protein model with 95% gap deletion and the bootstrap values are displayed as percentages of 1000 bootstraps on the branch nodes, Figure S3: Monomer structure alignment of γ-CAs from Methanosarcina thermophila (Cam, PDB ID: 1QRL, cyan), Pyrococcus horikoshii (Cap, PDB ID: 1V3W, orange), Escherichia coli (EcoCA-γ, UniProtKB Accession number: P0A9W9.1, brick red), Caminibacter mediatlanticus (γ-CmCA, wheat), Nitratiruptor tergarcus (γ-NtCA, red), Persephonella hydrogeniphila (γ-PhCA, magenta), Persephonella marina (γ-PmCA, yellow), Sulfurovum lithotrophicum (γ-SICA, slate blue), Thermus thermophilus (γ-TiCA, PDB ID: 6IVE, green) and Thermosulfidibacter takaii (γ-TtkCA, navy blue). The variable loop region containing the proton shuttling residue is shown in the black dotted circled and zoomed in the image pointed to by the black arrow, Figure S4: Bond distance between Zn$^{2+}$ and proton shuttling residues Glu62 (top) and Glu84 (bottom) in Cam’s three active sites during the 100 ns simulation. Distances were measured for both oxygen atoms, OD1 and OD2, which are capable of coordinating the Zn$^{2+}$. Table S1: Motifs generated by MEME, their positions in the γ-CAs, and their respective E-values, Table S2: Template coverage, sequence identity and model validation for γ-CA sequences, Table S3: Interface characteristics obtained from PDBePISA of the trimeric γ-CA structures

**Author Contributions:** Conceptualization, Ö.T.B.; formal analysis, C.V.M. and Ö.T.B.; funding acquisition, Ö.T.B.; methodology, C.V.M.; project administration, Ö.T.B.; resources, Ö.T.B.; supervision, Ö.T.B.; validation, C.V.M.; visualization, C.V.M.; writing—original draft, C.V.M.; writing—review & editing, C.V.M. and 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 Ph.D. This research 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 funders.

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

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

**Abbreviations**

| Abbreviation | Definition |
|--------------|------------|
| BC           | Betweenness centrality |
| CA           | Carbonic anhydrase |
| Cam          | Methanosarcina thermophila gamma carbonic anhydrase |
| EcoCA-γ      | Escherichia coli gamma carbonic anhydrase |
| DCC          | Dynamic cross correlation |
| γ-CmCA       | Caminibacter mediatlanticus γ-carbonic anhydrase |
| γ-EpCA       | Epsilonproteobacteria bacterium 4484_65 γ-carbonic anhydrase |
| γ-HtCA       | Hydrogenimonas thermophilus γ-carbonic anhydrase |
| γ-LnCA       | Lebetimonas natshishimae γ-carbonic anhydrase |
| γ-NpCA       | Nautilia profundicola γ-carbonic anhydrase |
| γ-NsbCA      | Nitratiruptor sp. SB155-2 γ-carbonic anhydrase |
| γ-NtCA       | Nitratiruptor tergarcus γ-carbonic anhydrase |
| γ-PhCA       | Persephonella hydrogeniphila γ-carbonic anhydrase |
| γ-PmCA       | Persephonella marina γ-carbonic anhydrase |
| γ-SaCA       | Sulfurovum autotrophica γ-carbonic anhydrase |
| γ-SICA       | Sulfurovum lithotrophicum γ-carbonic anhydrase |
| γ-SNbcCA     | Sulfurovum sp. NBC37-1 γ-carbonic anhydrase |
| γ-TtkCA      | Thermosulfidibacter takaii gamma carbonic anhydrase |
| MD           | Molecular dynamics |
| MSA          | Multiple sequence alignment |
| Rg           | Radius of gyration |
| RMSD         | Root mean square deviation |
| RMSF         | Root mean square fluctuation |
References

1. Tripp, B.C.; Smith, K.; Ferry, J.G. Carbonic anhydrase: New insights for an ancient enzyme. J. Biol. Chem. 2001, 276, 48615–48618. [CrossRef] [PubMed]
2. 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, 15456–15460. [CrossRef] [PubMed]
3. DiMario, R.J.; Machingira, 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]
4. Jensen, E.L.; Clement, R.; Kosta, A.; Maberly, S.C.; Gonterto, B. A new widespread subclass of carbonic anhydrase in marine phytoplankton. ISME J. 2019, 13, 2094–2106. [CrossRef] [PubMed]
5. Kikutani, S.; Nakajima, K.; Nagasato, C.; Tsuji, Y.; Miyatake, A.; Matsuda, Y. Thylakoid luminal β-carbonic anhydrase critical for growth and photosynthesis in the marine diatom Phaeodactylum tricornutum. Proc. Natl. Acad. Sci. USA 2016, 113, 9828–9833. [CrossRef] [PubMed]
6. 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 through use of a highly thermostable carbonic anhydrase. Acta Crystallogr. Sect. D Biol. Crystallogr. 2012, 68, 6683–6687. [CrossRef] [PubMed]
7. Manyumwa, C.V.; Emameh, 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]
8. Ferry, J.G. The γ class of carbonic anhydrases. Biochim. Biophys. Acta—Proteins Proteom. 2010, 1804, 374–381. [CrossRef]
9. Lane, T.W.; Saito, M.A.; George, G.N.; Pickering, I.J.; Prince, R.C.; Morel, F.M. Biochemistry: A cadmium enzyme from a marine diatom. Nature 2005, 435, 42. [CrossRef]
10. 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]
11. Alterio, V.; Langella, E.; Viparelli, F.; Vullo, D.; Ascione, G.; Dathan, N.A.; Morel, F.M.; Supuran, C.T.; De Simone, G.; Monti, S.M. Structural and inhibition insights into carbonate anhydrase CDCA1 from the marine diatom Thalassiosira weissflogii. Biochimie 2012, 94, 1232–1241. [CrossRef]
12. Xu, Y.; Feng, L.; Jeffrey, P.D.; Shi, Y.; Morel, F.M. Structure and metal exchange in the cadmium carbonate anhydrase of marine diatoms. Nature 2008, 452, 56–61. [CrossRef]
13. Liljas, A.; Kannan, K.K.; Bergsten, P.C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carlbom, U.; Järup, L.; Lövgren, S.; Petef, M. Crystal structure of human carbonic anhydrase C. Nat. New Biol. 1972, 235, 131–137. [CrossRef]
14. Cox, E.H.; McLendon, G.L.; Morel, F.M.; Lane, T.W.; Prince, R.C.; Pickering, I.J.; George, G.N. The active site structure of Thalassiosira weissflogii carbonic anhydrase 1. Biochemistry 2000, 39, 12128–12130. [CrossRef]
15. Jeyakanthan, J.; Rangarajan, S.; Sriraj, P.; Kanaujia, S.P.; Lo, K.H.; Ng, I.S. Observation of a calcium-binding site in the γ-class carbonic anhydrase from Pyrococcus horikoshii Horikoshii. Acta Crystallogr. Sect. D Biol. Crystallogr. 2008, 64, 1012–1019. [CrossRef]
16. Zimmerman, S.A.; Tomb, J.-F.; Ferry, J.G. Characterization of CamH from Methanosarcina thermophila, founding member of a subclass of the γ class of carbonic anhydrases. J. Bacteriol. 2010, 192, 1353–1360. [CrossRef]
17. Coleman, J.E. Mechanism of action of carbonic anhydrase substrate, sulfonamide, and anion binding. J. Biol. Chem. 1967, 242, 5212–5219. [CrossRef]
18. Park, H.M.; Park, J.H.; Choi, J.W.; Lee, J.; Kim, B.Y.; Jung, C.H.; Kim, J.S. Structures of the γ-class carbonic anhydrase homologue YrdA suggest a possible allosteric switch. Acta Crystallogr. Sect. D Biol. Crystallogr. 2012, 68, 920–926. [CrossRef] [PubMed]
19. Sawaya, M.R.; Cannon, G.C.; Heinhorst, S.; Tanaka, S.; Williams, E.B.; Yeates, T.O.; Kerfeld, C.A. The structure of β-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two. J. Biol. Chem. 2006, 281, 7546–7555. [CrossRef] [PubMed]
20. Covarrubias, A.S.; Larsson, A.M.; Högblom, M.; Lindberg, J.; Bergfors, T.; Björkelid, C.; Mowbray, S.L.; Unge, T.; Jones, T.A. Structure and function of carbonic anhydrases from Mycobacterium tuberculosis. J. Biol. Chem. 2005, 280, 18782–18789. [CrossRef] [PubMed]
21. Covarrubias, A.S.; Bergfors, T.; Jones, T.A.; Högblom, M. Structural mechanics of the pH-dependent activity of β-carbonic anhydrase from Mycobacterium tuberculosis. J. Biol. Chem. 2006, 281, 4993–4999. [CrossRef] [PubMed]
22. Ferraroni, M.; Del Prete, S.; Vullo, D.; Capasso, C.; Supuran, C.T. Crystal structure and kinetic studies of a tetrameric type II β-carbonic anhydrase from the pathogenic bacterium Vibrio cholerae. Acta Crystallogr. Sect. D Biol. Crystallogr. 2015, 71, 2449–2456. [CrossRef] [PubMed]
23. Del Prete, S.; Bua, S.; Supuran, C.T.; Capasso, C. Escherichia coli γ-carbonic anhydrase: Characterisation and effects of simple aromatic/heterocyclic sulphonamide inhibitors. J. Enzym. Inhib. Med. Chem. 2020, 35, 1545–1554. [CrossRef] [PubMed]
24. Manyumwa, C.V.; Tastan Bishop, Ö. In Silico Investigation of Potential Applications of Gamma Carbonic Anhydrases as Catalysts of CO2 Bomineralization Processes: A Visit to the Thermophilic Bacteria Persephonella hydrogeniphila, Persephonella marina, Thermosulfidobacter takai, and Thermus thermophilus. Int. J. Mol. Sci. 2021, 22, 2861. [CrossRef] [PubMed]
25. Waite, D.W.; Van Wonterghem, I.; Rinke, C.; Parks, D.H.; Zhang, Y.; Takai, K.; Sievert, S.M.; Simon, J.; Campbell, B.J.; Hanson, T.E.; et al. Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteraeota (phyl. nov.). Front. Microbiol. 2017, 8, 682. [CrossRef] [PubMed]
Catalysts 2021, 11, 1177

26. Waite, D.W.; Vanwonterghem, I.; Rinke, C.; Parks, D.H.; Zhang, Y.; Takai, K.; Sievert, S.M.; Simon, J.; Campbell, B.J.; Hanson, T.E.; et al. Addendum: Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteriaceae (phyl. nov.). Front. Microbiol. 2018, 9, 772. [CrossRef] [PubMed]

27. Brown, D.K.; Penkler, D.L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. Complete genome sequence of Methanosarcina thermophila. J. Bacteriol. 1996, 178, 3270–3274. [CrossRef]

28. Alber, B.E.; Ferrer, J.G. Characterization of heterologously produced carbonic anhydrase from Methanosarcina thermophila. J. Bacteriol. 1996, 178, 1307–1320. [CrossRef]

29. Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc coordination spheres in protein structures. J. Chem. Inf. Model. 2018, 52, 9020–9028. [CrossRef]

30. Eisenberg, D.; Lüthy, R.; Bowie, J.U. VERIFY3D: Assessment of protein models with three-dimensional profiles. In Methods in Enzymology; Academic Press: Cambridge, MA, USA, 1997; Volume 277, pp. 396–404. [CrossRef]

31. Herrou, J.; Crosson, S. Molecular structure of the Brucella abortus metalloprotein RicA, a Rab2-binding virulence effector. Biochemistry 2013, 52, 9020–9028. [CrossRef]

32. Inagaki, F.; Takai, K.; Nealson, K.H.; Horikoshi, K. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing epsilonproteobacterium from deep-sea hydrothermal vents. Int. J. Syst. Evol. Microbiol. 2004, 55, 773–779. [CrossRef]

33. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. Mol. Biol. Evol. 2008, 25, 1307–1320. [CrossRef]

34. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 2007, 372, 774–797. [CrossRef] [PubMed]

35. Kortemme, T.; Kim, D.E.; Baker, D. Computational alanine scanning of protein-protein interfaces. Protein Sci. 2004, 13, 219–22. [CrossRef] [PubMed]

36. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. Acta Crystallogr. D Biol. Crystallogr. 1993, 49, 288–299. [CrossRef]

37. 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, 328–338. [CrossRef] [PubMed]

38. Nagata, R.; Takaki, Y.; Tame, A.; Nunoura, T.; Muto, H.; Mino, S.; Sawayama, S.; Takai, K.; Nakagawa, S. Sulfurimonas autotrophica gen. nov., sp. nov., a novel sulfur-oxidizing epsilonproteobacterium from deep-sea hydrothermal sediments. Int. J. Syst. Evol. Microbiol. 2004, 55, 773–779. [CrossRef]

39. Nakagawa, S.; Takai, B.E.; Rice, J.G.; Rees, D.C. A closer look at the active site of γ-class carbonic anhydrases: High-resolution crystallographic studies of the carbonic anhydrase from Methanosarcina thermophila. Biochemistry 2000, 39, 9222–9231. [CrossRef]

40. Nakagawa, S.; Takaki, Y.; Shimamura, S.; Reysenbach, A.L.; Takai, K.; Horikoshi, K. Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. Proc. Natl. Acad. Sci. USA 2007, 104, 12146–12150. [CrossRef]

41. Nakagawa, S.; Takai, B.E.; Riberayama, H.; Nunoura, T.; Horikoshi, K.; Sako, Y. Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. Environ. Microbiol. 2005, 7, 1619–1632. [CrossRef]
53. 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. [CrossRef]

54. Katoh, K.; Misawa, K.; Kuma, K.I.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3859–3866. [CrossRef]

55. Notredame, C.; Higgins, D.G.; Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment1. *J. Mol. Biol.* **2000**, *302*, 205–217. [CrossRef] [PubMed]

56. Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobbitg, M.; Montanya, A.; Chang, J.M.; Taly, J.F.; Notredame, C. T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **2011**, *39*, W13–W17. [CrossRef] [PubMed]

57. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]

58. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]

59. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699. [CrossRef]

60. Huson, D.H.; Richter, D.C.; Rausch, C.; Dezulian, T.; Franz, M.; Rupp, R. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinf.* **2007**, *8*, 460. [CrossRef]

61. Parisi, G.; Perales, M.; Fornasari, M.; Colaneri, A.; Schain, N.; Casati, D.; Zimmermann, S.; Brennicke, A.; Araya, A.; Ferry, J.; et al. Gamma carbonic anhydrases in plant mitochondria. *Plant Mol. Biol.* **2004**, *55*, 193–207. [CrossRef]

62. Vogler, M.; Karan, R.; Renn, D.; Vancea, A.; Vööhrmet, M.; Grätzinger, S.W.; DasSarma, P.; Eppinger, J.; Groll, M.; et al. Crystal structure and active site engineering of a halophilic γ-carbonic anhydrase. *Front. Microbiol.* **2020**, *11*, 742. [CrossRef]

63. Pena, K.L.; Castel, S.E.; de Araujo, C.; Espie, G.S.; Kimber, M.S. Structural basis of the oxidative activation of the carboxysomal γ-carbonic anhydrase. *CemM. Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2455–2460. [CrossRef]

64. Wang, W.; Zhang, Y.; Wang, L.; Jing, Q.; Wang, X.; Xi, X.; Zhao, X.; Wang, H. Molecular structure of thermostable and zinc-ion-binding γ-class carboxy anhydrases. *BioMetals* **2019**, *32*, 317–328. [CrossRef] [PubMed]

65. Hatherley, R.; Brown, D.K.; Glenister, M.; Bishop, O.T. PRIMO: An interactive homology modeling pipeline. *PLoS ONE* **2016**, *11*, e0166698. [CrossRef]

66. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform.* **2006**, *15*, 5–6. [CrossRef]

67. Cukuroglu, E.; Gursoy, A.; Keskin, O. HotRegion: A database of predicted hot spot clusters. *Nucleic Acids Res.* **2012**, *40*, D829–D833. [CrossRef] [PubMed]

68. Darnell, S.J.; LeGault, L.; Mitchell, J.C. KFC Server: Interactive forecasting of protein interaction hot spots. *Nucleic Acids Res.* **2008**, *36*, W265–W269. [CrossRef] [PubMed]

69. Sukhwal, A.; Sowdhamini, R. PPCheck: A webserver for the quantitative analysis of protein-protein interfaces and prediction of residue hotspots. *Bioinformatics*. **2015**, *31*, S259–S259. [CrossRef]

70. Gordon, J.C.; Myers, J.B.; Folta, T.; Shoja, V.; Heath, L.S.; Onufriev, A. H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33* (Suppl. 2), W368–W371. [CrossRef]

71. Sanyanga, T.A.; Nizami, B.; Tastan Bishop, Ö. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with -Carbonic Anhydrase II Deficiency. *PLoS ONE* **2011**, *6*, e117746. [CrossRef] [PubMed]

72. Case, D.A.; Belfon, K.; Ben-Shalom, I; Brozell, S.R.; Cerutti, D.; Cheatham, T.; Cruzeiro, V.W.; Darden, T.; Duke, R.E.; Giambasu, G.; et al. Amber 2020. 2020. Available online: https://ambermd.org/doc12/Amber20.pdf (accessed on 14 March 2021).

73. 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]

74. Schafmeister, C.; Ross, W.; Romanovski, V. LeAP; University of California: San Francisco, CA, USA, 1995.

75. 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]

76. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

77. 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 25 March 2021).