

Supplementary data: Gamma carbonic anhydrases from hydrothermal vent bacteria: Cases of alternating active site due to a long loop with proton shuttle residue

Colleen Varaidzo Manyumwa¹ and Özlem Tastan Bishop^{1*}

¹ Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, 6140, South Africa; colleen.manyumwa06@gmail.com; O.TastanBishop@ru.ac.za

*Correspondence: O.TastanBishop@ru.ac.za; Tel: +27-46-603-8072; Fax: +27-46-603-7576

Table S1: Motifs generated by MEME, their positions in the γ -CAs, and their respective E-values.

Organism	Motif number and E-value								
	1 1.9e ⁻¹⁷⁸	2 1.9e ⁻¹⁶⁰	3 1.3e ⁻¹⁵⁹	4 2.0e ⁻¹²⁹	5 3.1e ⁻¹³⁵	6 2.6e ⁻¹¹⁸	7 3.2e ⁻¹⁰⁷	8 6.7e ⁻⁵⁴	9 1.3e ⁻⁴¹
γ -CmCA	90-109	37-56	133-154	111-130	57-76	155-174	16-35	1-15	79-89
γ -EpCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -HtCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -LnCA	90-109	37-56	133-154	111-130	57-76	155-174	16-35	1-15	79-89
γ -NpCA	90-109	37-56	133-154	111-130	57-76	155-174	16-35	1-15	79-89
γ -NsbCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -NtCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -PhCA	84-103	39-58	127-148	105-124	59-78	149-168	18-37	3-17	—
γ -PmCA	84-103	39-58	127-148	105-124	59-78	149-168	18-37	3-17	—
γ -SaCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -SiCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -SNbcCA	89-108	37-56	132-153	110-129	57-76	154-173	16-35	1-15	78-88
γ -TtkCA	81-100	36-55	124-145	102-121	56-75	146-165	15-34	—	—

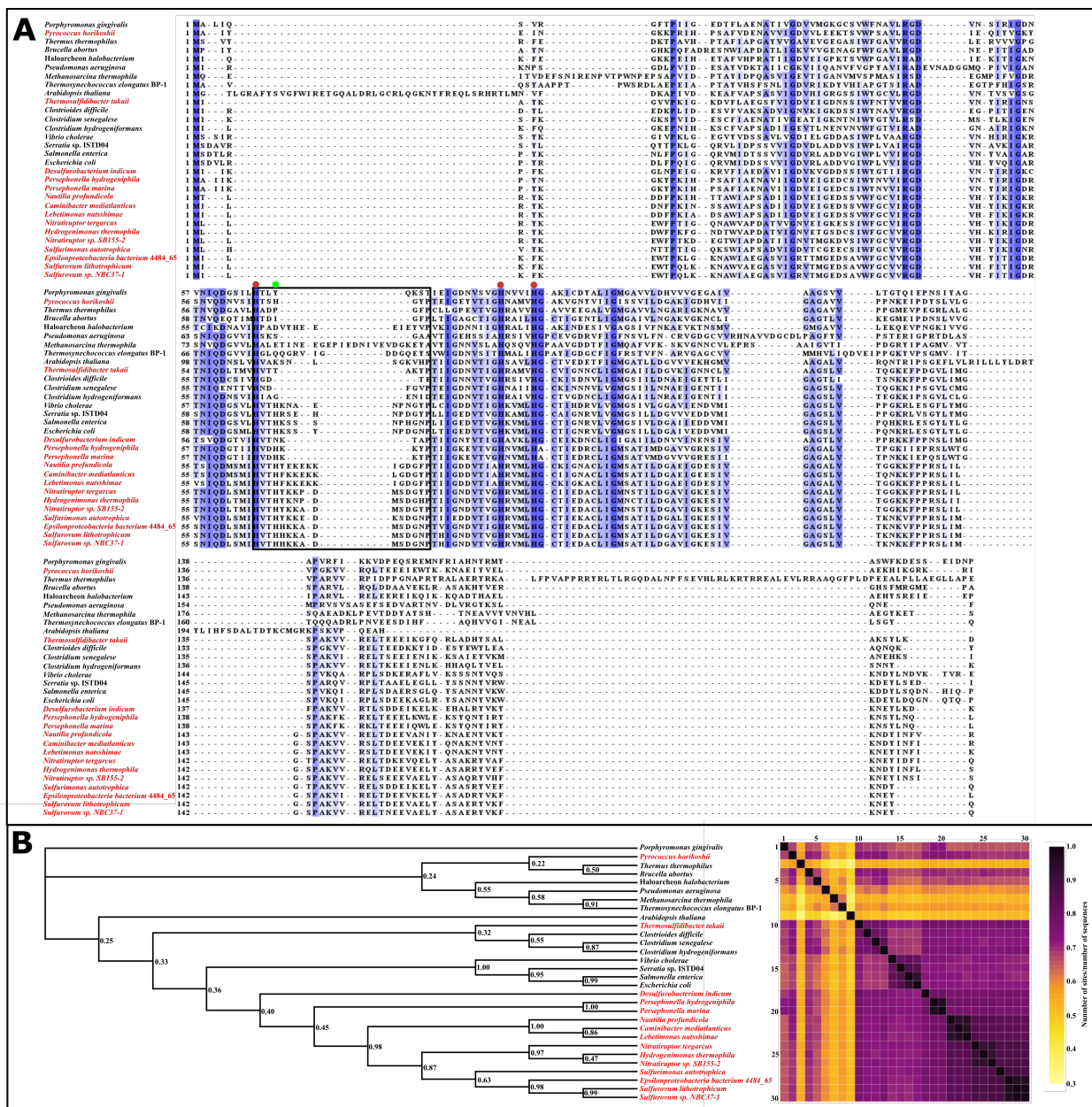


Figure S2: **A:** Multiple sequence alignment of γ -CAs from hydrothermal vent organisms as well as γ -CAs from other environments used to create the phylogenetic tree in B. The black box shows the variable loop containing the His proton shuttle. The red circles show the Zn^{2+} coordinating residues 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.

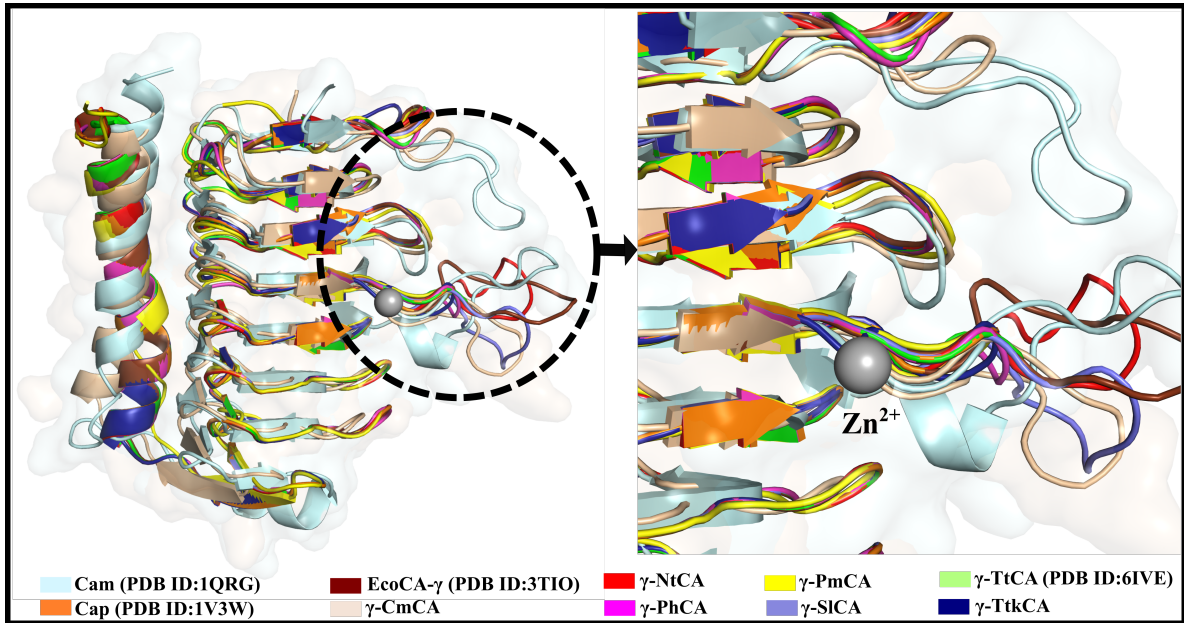


Figure S3: Monomer structure alignment of γ -CAs from *Methanosarcina thermophila* (Cam, PDB ID: 1QRL, cyan), *Pyrococcus horikoshii* (Cap, PDB ID: 1V3W, orange), *Escherichia coli* (EcoCA- γ , UniProtKB Accession number: P0A9W9.1, brick red), *Caminibacter mediatlanticus* (γ -CmCA, wheat), *Nitratiruptor tergarcus* (γ -NtCA, red), *Persephonella hydrogeniphila* (γ -PhCA, magenta), *Persephonella marina* (γ -PmCA, yellow), *Sulfurovum lithotrophicum* (γ -SICA, slate blue), *Thermus thermophilus* (γ -TtCA, 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.

Table S2: Template coverage, sequence identity and model validation for γ -CA sequences.

CA	Template PDB ID	Sequence identity (%)	Sequence coverage (%)	z-Dope score	Verify3D (%)	ProSA (z-score)	PROCHECK	
							Most favoured region (%)	Disallowed region(%)
γ -CmCA	1V3W	46	98	-1.32	85.3	-6.35	88.56	0.6
γ -EpCA		42	99	-1.42	84.2	-6.19	89.0	0.2
γ -HtCA		41	98	-1.44	83.2	-6.35	88.1	0.9
γ -LnCA		46	98	-1.23	84.7	-6.55	90.2	0.7
γ -NpCA		47	99	-1.24	83.2	-6.21	90.2	0
γ -NsbCA		40	98	-1.31	87.1	-6.53	88.6	1.3
γ -NtCA		41	99	-1.33	80.15	-6.29	87.6	1.1
γ -SaCA		41	100	-1.53	86.8	-5.95	88.5	0
γ -SICA		44	98	-1.34	81.2	-6.72	87.4	0.7
γ -SNbcCA		43	100	-1.33	83.3	-6.57	88.3	0.2

Table S3: Interface characteristics obtained from PDBePISA of the trimeric γ -CA structures.

Carbonic anhydrase	Number of hydrogen bonds	Salt bridges	Buried surface area/ \AA^2	Total surface area/ \AA^2	% BSA
Cam	60	18	7,458.6	28,398	26.2
EcoCA- γ	26	18	4,805.4	25,496	18.8
γ -CmCA	23	11	5,461.2	26,737	20.4
γ -EpCA	14	8	4,578.6	25,602	17.9
γ -HtCA	30	12	5,886.4	26,261	22.4
γ -LnCA	28	9	5,495.4	26,405	20.8
γ -NpCA	22	12	5,630.8	26,678	21.1
γ -NsbCA	27	12	6,031.0	25,834	23.3
γ -NtCA	24	9	6,023.8	26,829	22.5
γ -SaCA	18	10	5,072.6	25,176	20.1
γ -SICA	13	10	4,847.6	25,537	19.0
γ -SNbcCA	15	11	5,087.2	26,007	19.6

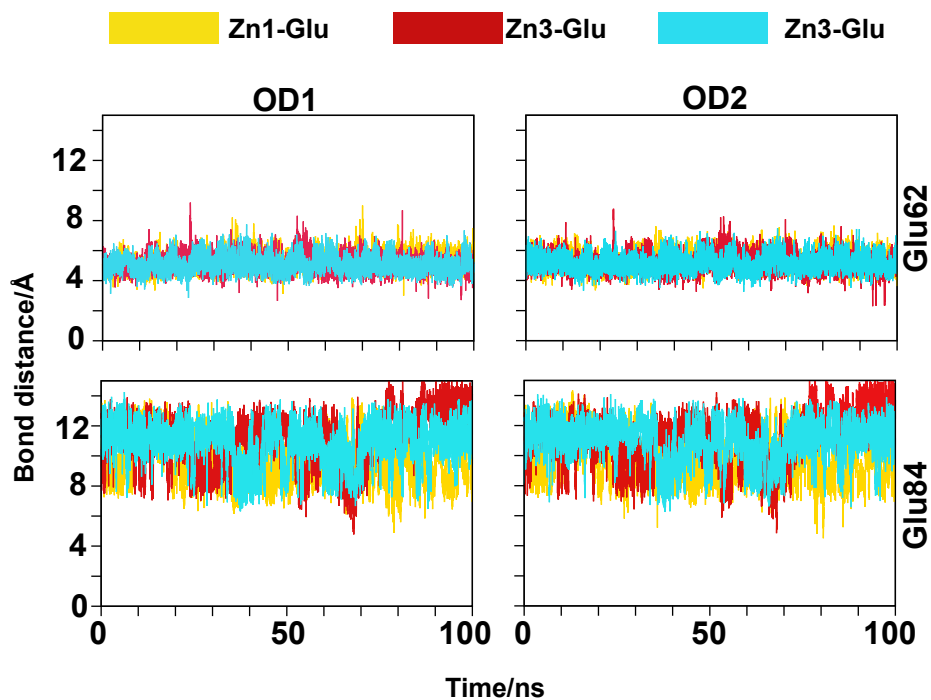


Figure S4: Bond distance between Zn^{2+} and proton shuttling residues Glu62 (top) and Glu84 (bottom) in Cam's three active sites during the 100 ns simulation. Distances were measured for both oxygen atoms, OD1 and OD2, which are capable of coordinating the Zn^{2+} .