Supplementary Information

Zeta-carbonic anhydrases show CS_2 hydrolase activity: a new metabolic carbon acquisition pathway in diatoms?

Vincenzo Alterio,[‡,†] Emma Langella,[‡,†] Martina Buonanno,[‡] Davide Esposito,[‡] Alessio Nocentini,[§] Emanuela Berrino,[§] Silvia Bua,[§] Maurizio Polentarutti,[⊥] Claudiu T. Supuran,[§] Simona Maria Monti[‡,*] and Giuseppina De Simone[‡,*]

[‡] Istituto di Biostrutture e Bioimmagini-CNR, via Mezzocannone 16, 80134 Napoli, Italy

[§] NEUROFARBA Department, Pharmaceutical and Nutraceutical Section, University of Firenze, Via Ugo Schiff 6, 50019 Sesto Fiorentino, Italy

[⊥]Elettra - Sincrotrone Trieste, s.s. 14 Km 163.5 in Area Science Park, Basovizza (Trieste) 34149, Trieste, Italy.

[†]These authors contributed equally to the work

Correspondence authors. Phone: Tel: +39-081-2534583, E-mail: marmonti@unina.it (SMM) or +39-081-2534579; E-mail: giuseppina.desimone@cnr.it (GDS).

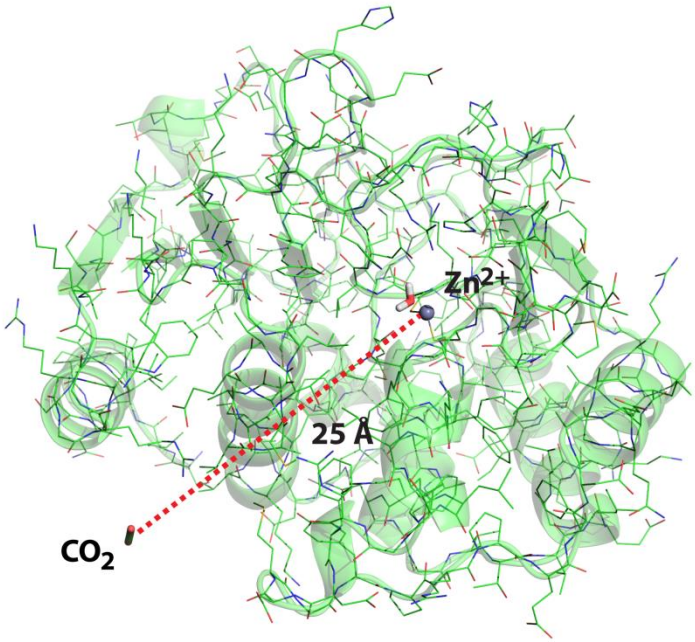

Figure S1. Starting configuration for simulations B1-B3. CO_2 molecule (stick representation) is placed out of the enzyme at ~ 5 Å from tunnel entrance and 25 Å from zinc ion (grey sphere). Distance between CO_2 and Zn^{2+} is shown as a red dashed line. Protein residues are also displayed.

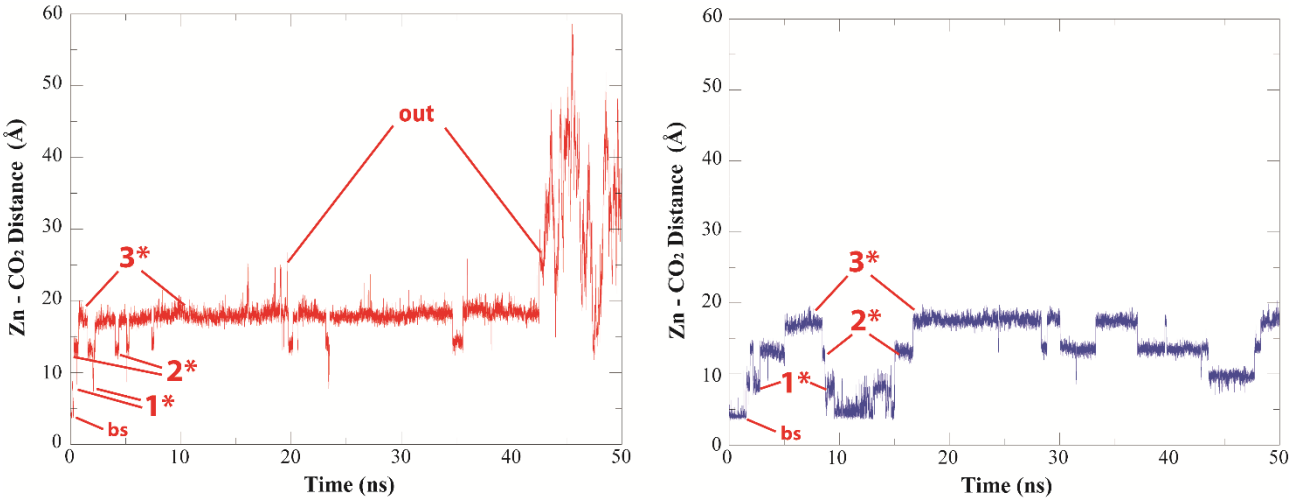

Figure S2. Distance between the Zn^{2+} ion and the CO_2 carbon atom plotted as a function of time for A2 (red, left panel) and A3 (blue, right panel) trajectories. Pockets (**1***, **2***, **3***) corresponding to CO_2 positions are indicated. **bs** corresponds to CO_2 position in its binding site, according to the crystallographic structure.

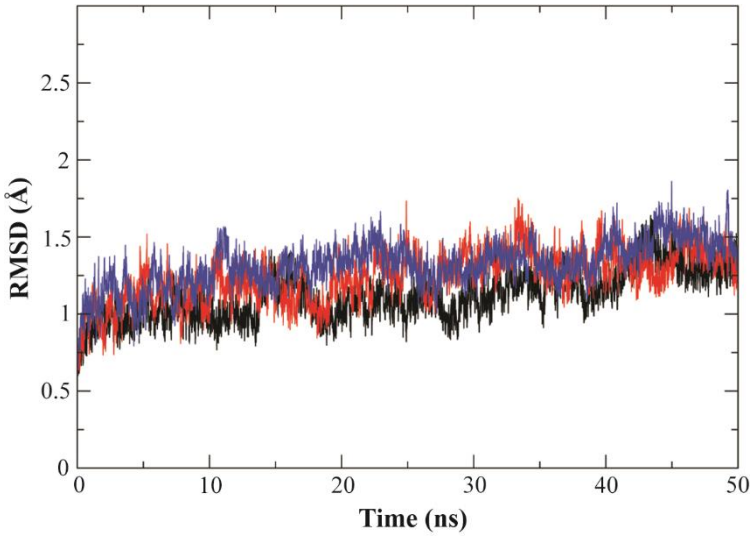

Figure S3. Root-Mean-Square Deviation (RMSD) computed on Cα atoms as a function of time for A1 (black), A2 (red) and A3 (blue) trajectories. RMSD indicates the stability of protein structure during simulation time.

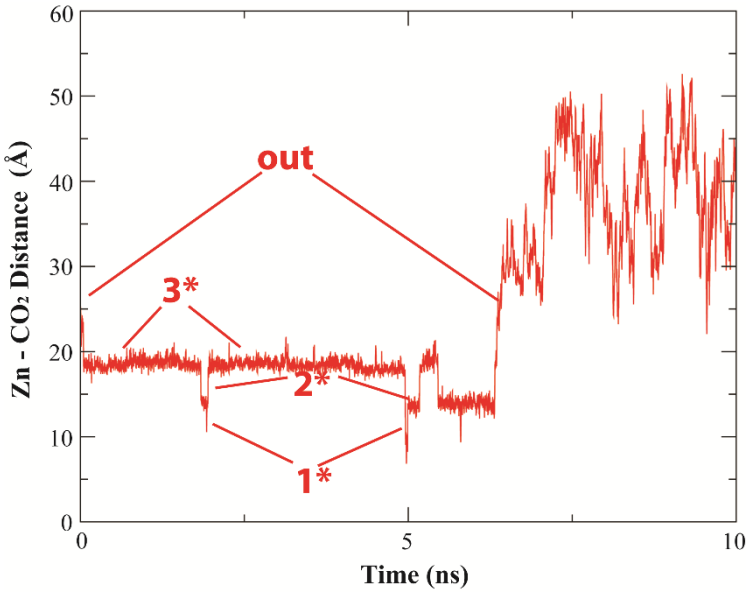

Figure S4. Distance between the Zn^{2+} ion and the CO_2 carbon atom plotted as a function of time for B2 trajectory. Pockets (**1*, 2*, 3***) corresponding to CO_2 positions are indicated.

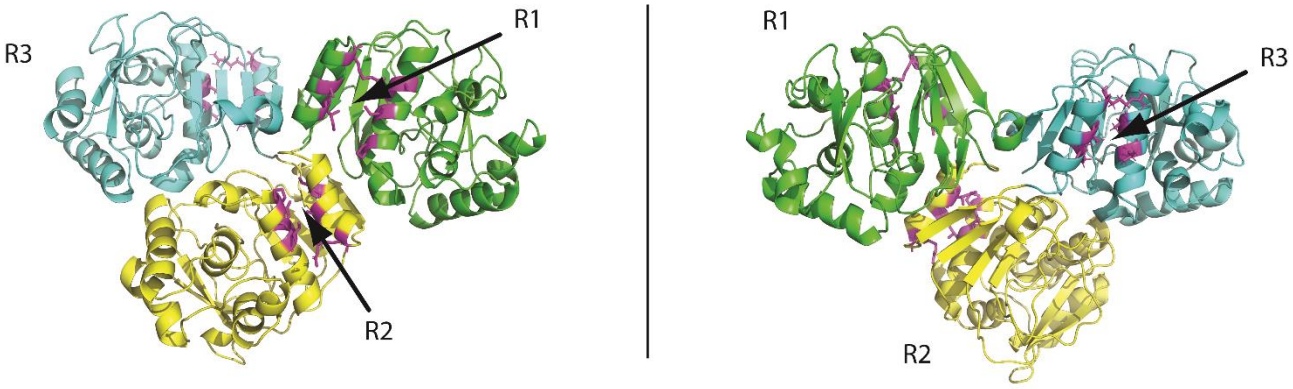

Figure S5. Model of CDCA1 full length [1]. Two views rotated by 180º around the y-axis are displayed. R1, R2 and R3 repeats are shown in green, yellow and cyan, respectively. Tunnel openings are indicated by arrows and their lining residues are highlighted in magenta.

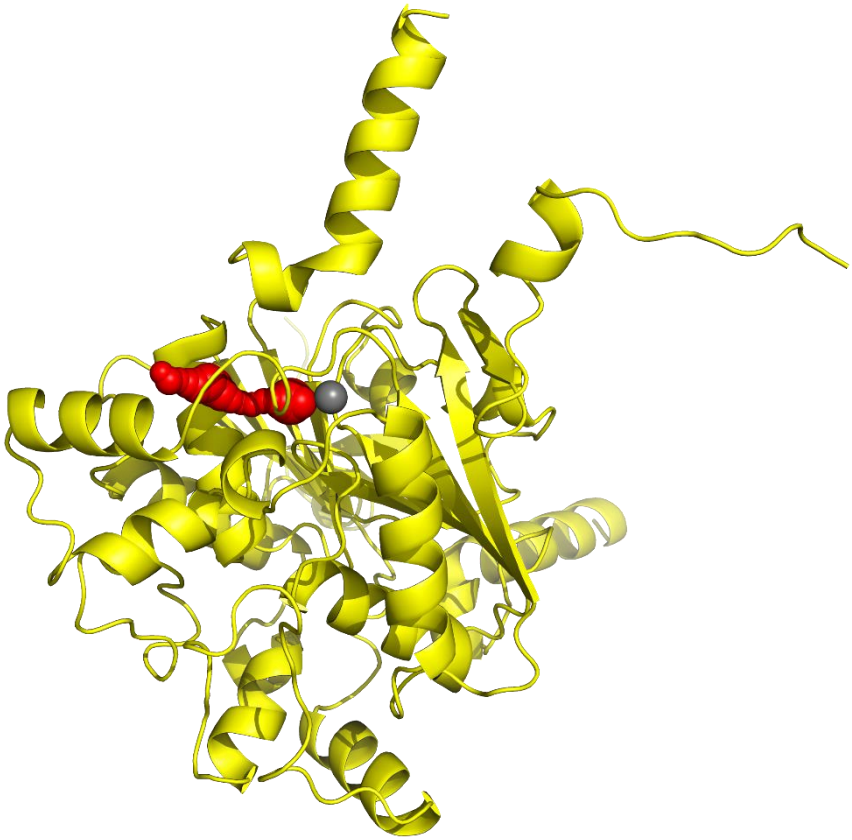

Figure S6. Tunnel identified in CS_2 hydrolase from *Acidianus* A1-3 (pbd code 3TEN) [2] using CAVER [3]. The protein is shown as yellow cartoon and the long hydrophobic tunnel as red spheres. Zinc ion is displayed as a grey sphere.

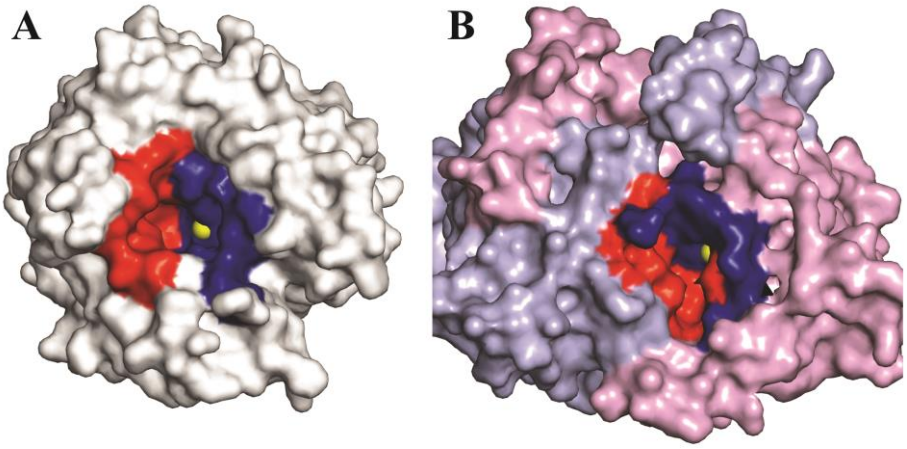

Figure S7. Solvent accessible surface of (**A**) hCA II (PDB code 1CA2) [4] and (**B**) the β–CA Nce103 from *Saccharomyces cerevisiae* (PDB code 3EYX) [5]. Putative CO_2 access routes are highlighted in red (hydrophobic region), blue (hydrophilic region). The catalytic zinc ion is showed as a yellow sphere. β-CA dimer chains A and B are reported in pink and light blue, respectively.

Table S1. Data collection and refinement statistics.

	Zn-R3	**Zn-R3/CO_2**
Crystal parameters		
Space group	C2	C2
a (Å)	117.8	119.6
b (Å)	62.5	62.7
c (Å)	75.4	75.0
β (°)	120.1	119.8
Data collection statistics		
Resolution (Å)	50.0-1.98 (2.02-1.98)	50.0-1.60 (1.63-1.60)
Temperature (K)	100	100
Total reflections	147785	710837
Unique reflections	31341	61450
Completeness (%)	95.2 (75.7)	97.0 (78.0)
$\langle I \rangle / \langle \sigma(I) \rangle$	13.3 (2.1)	35.8 (3.9)
Redundancy (%)	4.7 (2.7)	11.6 (7.1)
R_{merge}[a]	0.093 (0.454)	0.066 (0.443)
R_{meas}[a]	0.103 (0.555)	0.068 (0.474)
R_{pim}[a]	0.043 (0.311)	0.019 (0.162)
CC1/2[b]	0.996 (0.794)	0.999 (0.949)
Refinement statistics		
Resolution (Å)	33.24-1.98	50.0-1.60
R_{work}[c] (%)	20.8	21.5
R_{free}[c] (%)	24.3	24.6
r.m.s.d. from ideal geometry:		
Bond lengths (Å)	0.009	0.009
Bond angles (°)	1.5	1.5
Number of protein atoms	3188	3183
Number of ligand atoms		6
Number of water molecules	152	171
Average B factor (Å^2)		
All atoms	28.1	26.4
Protein atoms	28.0	26.1
Ligand atoms		32.8
Water molecules	29.8	30.5
PDB Code	**7BEZ**	**7BF0**

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{meas} = \Sigma_{hkl}\{n(hkl)/[n(hkl)-1]\}^{1/2}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{pim} = \Sigma_{hkl}\{1/[n(hkl)-1]\}^{1/2}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl)\rangle$ is the mean value for its unique reflection; summations are over all "n" reflections.

[b] $CC1/2 = [\Sigma_i(a_i-\langle a\rangle) / \Sigma_i(b_i-\langle b\rangle)] / [\Sigma_i(a_i-\langle a\rangle)^2 \Sigma_i(b_i-\langle b\rangle)^2]^{1/2}$; where a_i and b_i are the intensities of unique reflections merged across the observations randomly assigned to subsets A and B, respectively, and $\langle a\rangle$ and $\langle b\rangle$ are their averages.

[c] $R_{factor} = \Sigma_h||F_o(h)|-|F_c(h)|| / \Sigma_h|F_o(h)|$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. R_{free} was calculated with 4.1% of the data excluded from the refinement.

Table S2. Tunnel residues interacting with CO_2 into the main pockets (**1***, **2***, **3***) detected along the migration pathway.

Pocket **1***	Pocket **2***	Pocket **3***
Val628, Val584, Ala624 Val474, Phe603	Trp448, Val584, Ile586, Val605, Ala624, Ala625, Ala637	Ile439, Ala442, Leu443, Arg446, Leu620, Val618, Val622

References

[1] Alterio V, Langella E, Viparelli F, Vullo D, Ascione G, Dathan NA, et al. Structural and inhibition insights into carbonic anhydrase CDCA1 from the marine diatom Thalassiosira weissflogii. Biochimie 2012;94:1232-1241.
[2] Smeulders MJ, Pol A, Venselaar H, Barends TR, Hermans J, Jetten MS, et al. Bacterial CS2 hydrolases from Acidithiobacillus thiooxidans strains are homologous to the archaeal catenane CS2 hydrolase. J Bacteriol 2013;195:4046-4056.
[3] Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol 2012;8:e1002708.
[4] Eriksson AE, Jones TA, Liljas A. Refined structure of human carbonic anhydrase II at 2.0 A resolution. Proteins 1988;4:274-282.
[5] Teng YB, Jiang YL, He YX, He WW, Lian FM, Chen Y, et al. Structural insights into the substrate tunnel of Saccharomyces cerevisiae carbonic anhydrase Nce103. BMC Struct Biol 2009;9:67.