



# Crystal Structure and Active Site Engineering of a Halophilic $\gamma$ -Carbonic Anhydrase

Malvina Vogler<sup>1,2†</sup>, Ram Karan<sup>1\*†</sup>, Dominik Renn<sup>1</sup>, Alexandra Vancea<sup>1</sup>, Marie-Theres Vielberg<sup>2</sup>, Stefan W. Grötzinger<sup>1</sup>, Priya DasSarma<sup>3</sup>, Shiladitya DasSarma<sup>3</sup>, Jörg Eppinger<sup>1</sup>, Michael Groll<sup>2\*</sup> and Magnus Rueping<sup>1\*</sup>

<sup>1</sup> KAUST Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, <sup>2</sup> Center for Integrated Protein Science Munich, Department of Chemistry, Technische Universität München, Garching, Germany, <sup>3</sup> Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, Baltimore, MD, United States

## OPEN ACCESS

### Edited by:

Sunil Khare,  
Indian Institute of Technology Delhi,  
India

### Reviewed by:

Amy Michele Grunden,  
North Carolina State University,  
United States  
Mohammad Ali Amoozegar,  
University of Tehran, Iran

### \*Correspondence:

Ram Karan  
ram.karan@kaust.edu.sa  
Michael Groll  
michael.groll@tum.de  
Magnus Rueping  
magnus.rueping@kaust.edu.sa

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 18 February 2020

Accepted: 30 March 2020

Published: 28 April 2020

### Citation:

Vogler M, Karan R, Renn D,  
Vancea A, Vielberg M-T,  
Grötzinger SW, DasSarma P,  
DasSarma S, Eppinger J, Groll M and  
Rueping M (2020) Crystal Structure  
and Active Site Engineering of a  
Halophilic  $\gamma$ -Carbonic Anhydrase.  
Front. Microbiol. 11:742.  
doi: 10.3389/fmicb.2020.00742

Environments previously thought to be uninhabitable offer a tremendous wealth of unexplored microorganisms and enzymes. In this paper, we present the discovery and characterization of a novel  $\gamma$ -carbonic anhydrase ( $\gamma$ -CA) from the polyextreme Red Sea brine pool Discovery Deep (2141 m depth, 44.8°C, 26.2% salt) by single-cell genome sequencing. The extensive analysis of the selected gene helps demonstrate the potential of this culture-independent method. The enzyme was expressed in the bioengineered haloarchaeon *Halobacterium* sp. NRC-1 and characterized by X-ray crystallography and mutagenesis. The 2.6 Å crystal structure of the protein shows a trimeric arrangement. Within the  $\gamma$ -CA, several possible structural determinants responsible for the enzyme's salt stability could be highlighted. Moreover, the amino acid composition on the protein surface and the intra- and intermolecular interactions within the protein differ significantly from those of its close homologs. To gain further insights into the catalytic residues of the  $\gamma$ -CA enzyme, we created a library of variants around the active site residues and successfully improved the enzyme activity by 17-fold. As several  $\gamma$ -CAs have been reported without measurable activity, this provides further clues as to critical residues. Our study reveals insights into the halophilic  $\gamma$ -CA activity and its unique adaptations. The study of the polyextremophilic carbonic anhydrase provides a basis for outlining insights into strategies for salt adaptation, yielding enzymes with industrially valuable properties, and the underlying mechanisms of protein evolution.

**Keywords:** extremophiles, halophiles, thermophiles, extremozyme, salt adaptation, mutagenesis, gamma-carbonic anhydrase

## INTRODUCTION

Recent years have seen the discoveries of extremophiles in environments previously considered uninhabitable (Madigan and Marrs, 1997; Cavicchioli et al., 2011; Antunes et al., 2017; Jorquera et al., 2019; Merino et al., 2019). To date, enzymes from extremophiles gained increasing attention because they have adapted their structure and retained their function under harsh conditions, where their mesophilic homologs are non-functional (Persidis, 1998; Akal et al., 2019). In particular, these

proteins are attractive for biotechnological and chemical industries keen on replacing traditional catalysts with enzymes. As a result, these extremozymes provide a high stereoselectivity with fewer side reactions, and lower the burden on the environment, concomitantly accelerating reaction rates (Littlechild, 2017). However, many processes require high temperatures, use of salts, organic solvents, or other demanding conditions that are incompatible with the stability and function of most proteins (Littlechild, 2015; Amoozegar et al., 2019). Conversely, extremophilic organisms, which are naturally adapted to withstand harsh conditions, provide a perspective for optimization and rational protein-engineering approaches (Liszka et al., 2012).

The Red Sea constitutes a unique habitat of several anoxic deep-sea brine pools (Gurvich, 2006; Behzad et al., 2016) and, therefore, promises the discovery of a vast number of new extremophiles and enzymes. The anoxic environment is polyextremophilic, being filled with water, a high salt concentration, metal content, and elevated temperatures. Additionally, their increased density prevents mixing with the overlying seawater (Gurvich, 2006; Antunes et al., 2011). For example, the Discovery Deep brine pool below 2038 m is characterized by a salinity close to saturation (26.2%, w/v) with a temperature of 44.7°C (Hunt et al., 1967). Therefore, Discovery Deep's microorganisms have only been scarcely investigated; nonetheless, a recent surge in interest has begun to provide glimpses of the wealth of new information waiting to be explored (Antunes et al., 2011; Mwirichia et al., 2016; Grotzinger et al., 2018).

The limited accessibility and uncultivability of the aforementioned microorganisms with current laboratory techniques hamper the investigation of these exceptional species (Stewart, 2012). Consequently, methods independent of cultivation and their further development, such as the use of Single Amplified Genomes (SAG), are required (Kvist et al., 2007; Rashid and Stingl, 2015). Here, DNA from a single cell is amplified using the Multiple Displacement Amplification (MDA) techniques (Dean et al., 2001) to generate sufficient DNA for sequencing and to avoid the need for cultivation of the respected organisms. To improve the assembly and annotation accuracy for SAG-derived samples, our group developed the Profile and Pattern Matching (PPM) algorithm method (Grotzinger et al., 2014). In this work, we use data from SAG analysis to investigate an extremophilic  $\gamma$ -carbonic anhydrase ( $\gamma$ -CA) from the Red Sea Discovery Deep brine pool. CAs (E.C. 4.2.1.1) are ubiquitous metalloenzymes that catalyze the reversible hydration of carbon dioxide to bicarbonate ( $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$ ) (Supuran, 2008). To date, seven classes have been described:  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ -,  $\zeta$ -,  $\eta$ -, and  $\theta$ -CAs. These hydrolases differ significantly in both sequence and structure, whilst all catalyzing the same reaction (Ferry, 2013; Del Prete et al., 2016; Supuran and Capasso, 2017). The  $\alpha$ -CA class was the first to be isolated and described (Meldrum and Roughton, 1933), whereas the  $\gamma$ -carbonic anhydrase from the thermophilic archaeon *Methanosarcina thermophila* (Cam) was discovered in 1994 (Alber and Ferry, 1994). It was categorized as a new class showing a left-handed parallel  $\beta$ -helix fold, and central metal coordination in the active

site by three histidine residues (Kisker et al., 1996). Subsequently, further  $\gamma$ -CAs have been investigated; however, half of them did not show measurable activity (Park et al., 2012; Herrou and Crosson, 2013), and the underlying molecular mechanisms are still controversial.

Here, we report on the first crystal structure of a halophilic  $\gamma$ -class CA (CA\_D). The gene was derived from SAG analysis of an uncultured archaeon from the Red Sea Discovery Deep brine pool (Alam et al., 2013; Mwirichia et al., 2016) and was identified using the PPM algorithm (Grotzinger et al., 2014). The gene was expressed in the bioengineered haloarchaeon *Halobacterium* sp. NRC-1. We demonstrate that CA\_D indeed encodes a  $\gamma$ -carbonic anhydrase. Activity analysis of structure-driven designed CA\_D variants provided insights into the residues constituting the catalytic site.

## MATERIALS AND METHODS

### Chemicals and Reagents

Restriction enzymes, T4 DNA ligase, and DNA polymerase were purchased from New England Biolabs (Beverly, MA, United States). Chemicals were purchased from Sigma (St. Louis, MO, United States). Water was desalted and purified using a Milli-Q<sup>®</sup> Academic system (Merck, Darmstadt, Germany).

### SAG Sampling Sites, Sample Preparation, and Genome Annotation

Samples were collected from the Discovery Deep brine pool in the Red Sea (21° 16.98'/38° 03.18'). Cells were sorted using fluorescence-activated cell sorting (FACS), lysed, the whole genome amplified and sequenced (Mwirichia et al., 2016). Genes were annotated using the INDIGO data warehouse system in combination with the profile pattern matching algorithm (PPMA) (Alam et al., 2013; Grotzinger et al., 2014).

### Strains, Plasmids, Media, and Culture Conditions

*Escherichia coli* One Shot TOP10<sup>®</sup> chemical competent cells were purchased from Invitrogen (Carlsbad, United States). *E. coli* was grown at 37°C in Luria-Bertani (LB) medium supplemented with 100  $\mu\text{g}/\text{ml}$  ampicillin. *Halobacterium* strains were cultured in  $\text{CM}^+$  medium containing 4.3 M NaCl and trace metals at 42°C with shaking as previously described (DasSarma et al., 1995). For solid media, 2% (w/v) agar was added. Stock cultures were maintained in glycerol at -80°C. For short-term use, purified cultures were maintained on stock plates at 4°C.

### Construction of the *Halobacterium* Carbonic Anhydrase Knockout Strain

To eliminate background carbonic anhydrase production, *icfA* was knocked out via the *ura3*-based gene deletion method for *Halobacterium* sp. NRC-1. Approximately 500 bp regions flanking the carbonic anhydrase gene (*icfA*) were amplified by crossover PCR (using primers shown in **Supplementary Table S5**). The resulting amplified crossover PCR fragment was

cloned into the suicide vector, pBB400 using flanking *Hind*III and *Eco*RI sites incorporated in the primers (**Supplementary Table S5**) (Berquist et al., 2006). The resulting plasmid, pBB400 $\Delta icfA$  was transformed into *Halobacterium* sp. NRC-1 $\Delta ura3$  using the standard PEG-EDTA method (DasSarma et al., 1995). pBB400 $\Delta icfA$  transformants were selected by plating on CM<sup>+</sup> uracil dropout media (HURA), colonies picked and grown in liquid HURA media, and integrant candidates were plated onto 5-FOA-CM<sup>+</sup> media plates. Knockout candidates were identified by DNA extraction and PCR using flanking primers listed in **Supplementary Table S5** (DasSarma et al., 1995; Berquist et al., 2006).

## Construction of the Expression Plasmids

Synthetic genes were codon-optimized using the java codon adaptation online tool JCat (Grote et al., 2005) for *Halobacterium* sp. (strain NRC-1/ATCC 700922/JCM 11081). The optimized genes were ordered from GeneArt (Regensburg, Germany) and cloned into pRK42, which harbors an N-terminal His<sub>6</sub>-tag, *cspD2* promoter, origins of replication for *E. coli* and *Halobacterium*, and genes for ampicillin and mevinolin resistance for selection in *E. coli* and *Halobacterium*, respectively.

## Expression of the Carbonic Anhydrase Genes in *Halobacterium* sp. NRC-1 $\Delta ura3\Delta icfA$ and Purification of the Encoded Proteins

Carbonic anhydrase genes containing vectors were transformed into the *Halobacterium* sp. NRC-1 $\Delta ura3\Delta icfA$  strain using the PEG/EDTA method (DasSarma et al., 1995; Karan et al., 2013) and transformants were selected by plating on CM<sup>+</sup> agar plates using mevinolin resistance. For protein production, cells were grown to late log phase (OD<sub>600nm</sub> of 0.9–1.0) at 42°C in CM<sup>+</sup> medium supplemented with 20  $\mu$ g/ml mevinolin. To induce carbonic anhydrase expression, the cultures were further incubated at 15°C for 72 h.

Cells were harvested by centrifugation (6,000  $\times$  g, 4°C, 10 min) in a 5430R centrifuge (Eppendorf, Germany) and disrupted in binding buffer (20 mM HEPES buffer pH 7.4 containing 2.0 M NaCl, 10% v/v glycerol, protease inhibitor cocktail, cOmplete from Roche, Germany and 30 mM imidazole) using a sonicator (Model Q500, QSONICA, Newtown, CT, United States) with a 1.9 cm probe (Thermo Scientific, Waltham, United States). Cell debris were removed by centrifugation (25,000  $\times$  g, 4°C, 10 min) in an Avanti J-26 XP centrifuge (Beckman Coulter, High Wycombe, United Kingdom) and the resulting crude extract was filtered through a 0.2  $\mu$ m Nalgene membrane filter (Thermo Scientific, Rockford, IL, United States). The supernatant was loaded at a flow rate of 1.0 ml/min onto a 5-ml HiTrap Ni<sup>2+</sup> chelating column (GE Healthcare Life Sciences, Piscataway, NJ, United States) pre-equilibrated with binding buffer. The column was washed with binding buffer, and the protein was eluted by increasing concentration of imidazole (30–300 mM) in binding buffer. The purified active fractions were combined and further purified and concentrated with Amicon® Ultra-4 Centrifugal Filter Units, 10 kDa (Cat no. UFC803024, Merck Millipore, MA, United States). Protein concentration

was determined using the NanoDrop 2000c (Thermo Fisher Scientific, Pittsburgh, PA, United States) with an absorption coefficient of 9,970 M<sup>-1</sup> cm<sup>-1</sup> (Gasteiger et al., 2005).

## Polyacrylamide Gel Electrophoresis

The SDS-PAGE analysis was performed by using the precast Novex® Tris-glycine gels (4–20%, Invitrogen, Carlsbad, CA, United States). The gels were stained with Imperial Protein Stain and Invision His-Tag In-gel stain (Thermo Fisher, Rockford, IL, United States).

## Identification of Purified Protein by LC-MS/MS Analysis

The pure samples (10  $\mu$ g) were digested with trypsin/LysC mix (Promega, Madison, WI, United States) using the FASP protocol (Wiśniewski et al., 2009). Peptides were measured using a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States) and analyzed using MASCOT v2.3 (Matrix Sciences Ltd, United Kingdom).

## Activity Measurements

Carbonic anhydrase activity was measured by the SX20 Stopped-Flow Spectrometer (Applied Photophysics, Leatherhead, United Kingdom) using the pH indicator dye phenol red as described previously (Alber et al., 1999). Briefly, chamber A contained 100  $\mu$ M phenol red in 20 mM MOPS buffer pH 9.8 containing 3 M KCl, with and without protein for catalyzed and uncatalyzed reaction, respectively. Chamber B contained CO<sub>2</sub>-saturated water prepared by bubbling CO<sub>2</sub> into distilled deionized water at 25°C. The reaction was monitored spectrophotometrically by measuring the increase in absorbance at 557 nm. All slit widths were set at 0.5 mm. An attached water bath regulated the temperature as indicated. A carbonic anhydrase (0.5  $\mu$ M) from bovine erythrocytes (Sigma, St. Louis, MO, United States) was used as a positive control. As a negative control, the crude lysate of untransformed *Halobacterium* sp. NRC-1 $\Delta ura3\Delta icfA$  cells was collected, purified using a Ni-NTA column, and the activity was measured for both the crude lysate and the purified fractions.

The specific activity was calculated via the Wilbur-Anderson unit (WAU) per 1 mg of protein, with one unit of activity being defined as (T<sub>0</sub> – T)/T, where T<sub>0</sub> (uncatalyzed reaction) and T (catalyzed reaction) are recorded as the time (sec.) required for the pH to drop from 9.8 to the transition point of the dye in a buffer control and in presence of enzyme or positive/negative control, respectively. Therefore, the reaction time until the activity plateau was reached was measured for the background reaction and all measurements were averaged and analyzed via the Excel solver.

## X-ray Crystallography

For crystallization, the protein was dialyzed against 50 mM Tris-HCl buffer pH 8.0, containing 300 mM NaCl and concentrated to 17 mg/ml with Amicon® Ultra-4 Centrifugal Filter Units, 10 kDa (Cat no. UFC803024, Merck Millipore, MA, United States). Protein crystals were grown at 20°C using the hanging drop vapor diffusion method. The protein solution was mixed in a 1:1 ratio with the reservoir buffer, 0.1 M HEPES pH 7.5 containing 0.05 M



cadmium sulfate and 0.8 M sodium acetate. Crystals selected for measurement were flash-frozen in liquid nitrogen after soaking in cryobuffer (70%, v/v reservoir buffer and 30%, v/v glycerol).

## Data Collection, Structure Solution and Refinement

Native diffraction datasets were collected at 2.6 Å resolution using synchrotron radiation at the X06SA-beamline, SLS, Villigen, Switzerland (see **Supplementary Table S1**). Recorded reflections were processed with XDS (Kabsch, 1993). CA\_D crystallized in the cubic space group F432 with  $a = 362.6$  Å, indicating five  $\gamma$ -CA-subunits in the asymmetric unit and a solvent content of 74%. Phases were obtained by Patterson search algorithms using the coordinates 1V3W as starting model (Jeyakanthan et al., 2008). The primary sequence was placed into the  $2F_o - F_c$  electron density map using COOT (Emsley and Cowtan, 2004) and refined with REFMAC5 (Murshudov et al., 1997). The model was completed in iterative rounds where temperature factors were anisotropically refined by translation/libration/screw motion-parameters, yielding crystallographic values of  $R_{cryst} = 0.177$  and  $R_{free} = 0.204$  (see **Supplementary Table S1**). Coordinates were confirmed to have adequate stereochemistry in the Ramachandran plot with 98.0% of residues in most favored, 1.8% in additionally allowed, and 0.2% in outlier regions. The crystal structure was deposited at the RCSB Protein Data Bank under accession codes 6SC4.

## Structural Analysis

Crystal structure analysis and visualization were performed using programs PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC)<sup>1</sup> and Yasara (Krieger and Vriend, 2014), and SwissPDB Viewer 4.1 (Guex and Peitsch, 1997). Homolog searches were performed using the DALI server (Holm and Rosenstrom, 2010). For comparison of CA\_D to homolog structures, all duplicates or structure variants belonging to the same homolog protein were neglected. Interactions between protein residues were calculated using Yasara (Krieger and Vriend, 2014), except for salt bridges which were calculated using VMD (Humphrey et al., 1996) with an oxygen-nitrogen cut-off of 4 Å (between residues Arg/Lys/His and Asp/Glu) and hydrogen bonds being estimated using Chimera (Pettersen et al., 2004).

Electrostatic surface potential calculations were performed using the PDB2PQR server (Dolinsky et al., 2004) along with the PyMOL plugin APBS (Baker et al., 2001). Surface-exposed amino acids were determined using Swiss PDB viewer 4.1 (Guex and Peitsch, 1997) ( $\geq 10\%$  surface accessibility). An estimation of the oligomeric assembly was performed using the program PISA (Krissinel and Henrick, 2007).

## Mutation Library

The QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, United States) was used according to manufacturer's instructions in combination with the primers listed in the **Supplementary Table S5**. pRK.CAD vector containing the CA\_D gene was used as a DNA template.

<sup>1</sup><https://bioinformatics.chap.wordpress.com/tools/visualization-tools/pymol/>

## RESULTS AND DISCUSSION

The remarkable stability of extremozymes and potential applications under harsh operational conditions has gained increased interest (Littlechild, 2017). Nonetheless, to gain a better understanding of the halophilicity in salt adapted proteins and the underlying molecular mechanisms of halophilic enzymes, additional studies of model proteins are required.

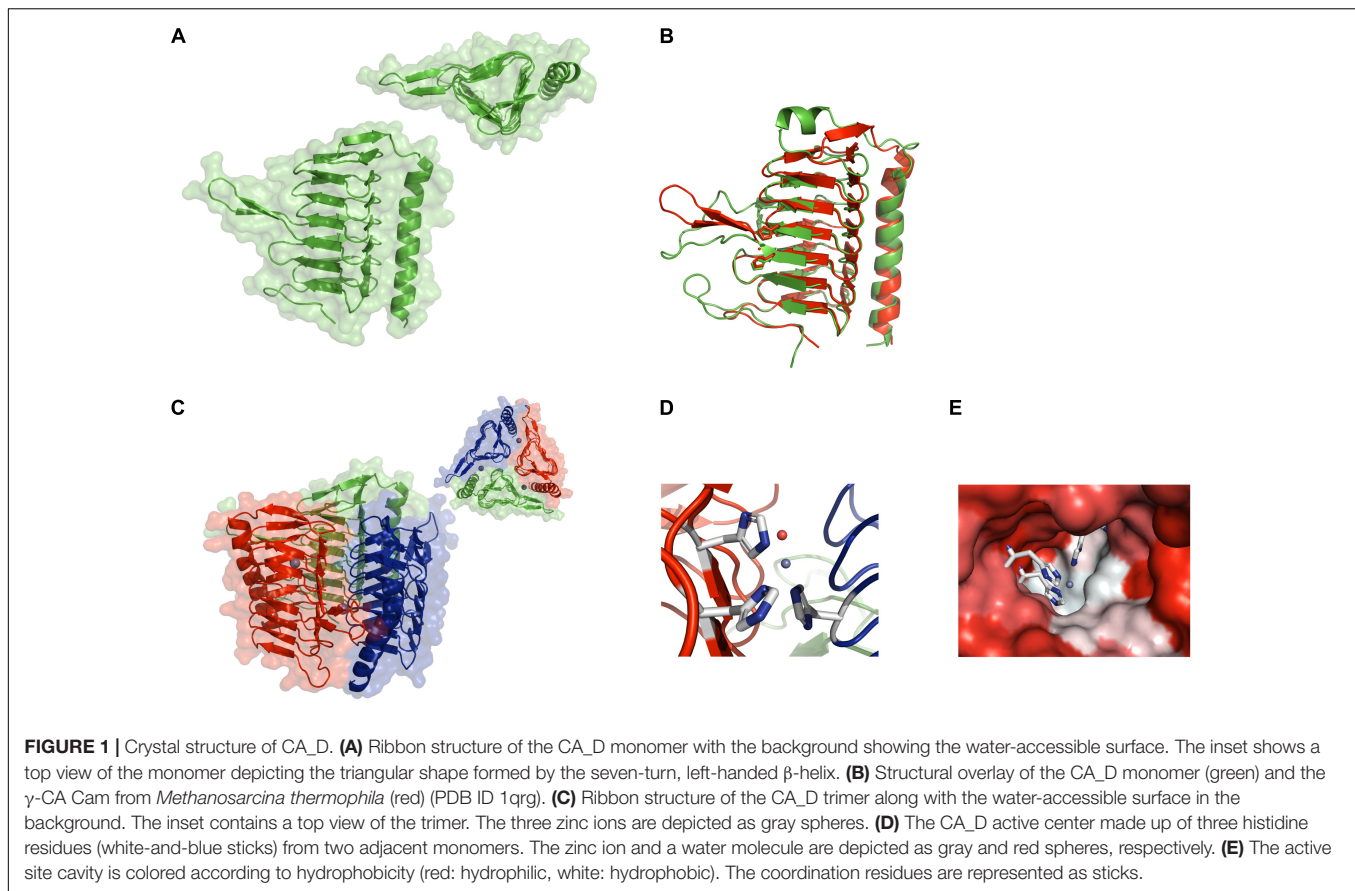
### Generation of the CA\_D Protein From SAG Analysis of a Discovery Deep Sample

The CA\_D gene originated from SAG analysis was identified and annotated as a  $\gamma$ -carbonic anhydrase ( $\gamma$ -CA) using the PPM algorithm (Grotzinger et al., 2014). We used the genetically modifiable extreme haloarchaeon *Halobacterium* sp. NRC-1 (Karan et al., 2013) as an expression system. *Haloarchaea* contain an internal salt concentration of 4-5 M and maintain an isoosmotic balance of ion concentrations in the cytosol with the surrounding medium. *Halobacterium* sp. NRC-1 harbors a carbonic anhydrase gene, *icfA*, located in the chromosome, location 638911 ← 639570 (**Supplementary Figure S1**). To eliminate background carbonic anhydrase production, *icfA* was knocked out (Berquist et al., 2006). The resulting *Halobacterium* sp. NRC-1  $\Delta$ ura3  $\Delta$ icfA deletion strain was used for the CA\_D expression. The enzyme was purified to homogeneity, with a total yield of 5 mg of protein/liter of culture (**Supplementary Figure S2A**). The identity of CA\_D was confirmed by tryptic digest and LC-MS/MS analysis (89% coverage, **Supplementary Figure S2B**).

### Crystal Structure of a Halophilic $\gamma$ -Carbonic Anhydrase From the Discovery Deep Brine Pool

The crystal structure of CA\_D was solved successfully to a resolution of 2.6 Å (PDB ID: 6SC4). Data collection and crystallographic quality statistics are shown in Table S1. The monomer contains a seven-turn, left-handed  $\beta$ -helix connected to an  $\alpha$ -helix running antiparallel to the  $\beta$ -helix axis (**Figure 1A**). The majority of the structure comprises  $\beta$ -sheet (54.5%), while only a minimal amount consists of  $\alpha$ -helix (13.5%), turn (12.9%), and coil structure (19.1%). The overall structure is highly conserved compared to reported  $\gamma$ -Cas (**Figure 1B**). Whereas the protein core is common to published  $\gamma$ -CA structures, differences in the connection of the helix motif is conspicuous. For example, CA\_D contains a connecting  $\beta$ -sheet, while Cam consists of an additional  $\alpha$ -helix instead. However, the comparison of CA\_D with all 10 crystallized  $\gamma$ -CA structures revealed that the majority contained a  $\beta$ -sheet with the thermophilic Cam and carboxysomal CcmM being an exception. Since this characteristic feature did not correspond to all investigated thermophilic homologs, we aimed to analyze whether the  $\beta$ -sheet has a role in the stability and rigidity of the overall architecture. Furthermore, it is likewise unclear if this plays a prominent role for the enzyme function.

Notably, CA\_D is organized into trimers (**Figure 1C**), resembling the reported active conformation for  $\gamma$ -CAs, where the active site includes residues from the adjacent monomers



(Ferry, 2010). Closer inspection of the active site reveals a zinc ion, coordinated by three histidine residues (His64 chain A, His89 chain B, and His94 chain A). A well-defined water molecule is coordinated to the zinc ion, which acts as the nucleophile in the reaction mechanism of these enzymes (Supuran, 2016) (**Figure 1D**). Interestingly, one half of the active site pocket exhibits a more hydrophobic character whereas the facing side is more hydrophilic (**Figure 1E**) (Supuran and De Simone, 2015). Taken together, the overall structure of CA\_D reveals a strong conservation to reported  $\gamma$ -CAs, supporting the  $\gamma$ -CA class gene annotation. Contrary to the sequence conservation, where the alignment of the CA\_D sequence with structural homologs shows a sequence identity of less than 40% for alignments with a query coverage larger than 95%, the structure is, consequently, well conserved (**Supplementary Table S2**).

### Halophilic Adaptation of CA\_D Compared to Other Non-halophilic $\gamma$ -CAs

While the overall CA\_D structure is conserved to other known  $\gamma$ -CAs, differences to non-halophilic  $\gamma$ -CAs must provide the observed stability under high salt concentrations. Thus, the structural elucidation provides interesting insights into halophilic adaptation by comparing CA\_D with meso- or thermophilic  $\gamma$ -CAs (**Supplementary Table S2**).

### Comparison of Stabilizing Interactions

Compared to the average values for meso- and thermophilic CAs, the CA\_D monomer contained an increased number of salt bridges (12 vs. an average of 10 and 9 for meso- and thermophilic homologs, respectively) (**Supplementary Table S3**), which is seen as a characteristic of haloadaptation (Frolow et al., 1996; Britton et al., 2006; Karan et al., 2012a). Despite the higher average number of salt bridges for the CA\_D monomer, several homologs showed a comparable or increased amount of salt bridges. In fact, the discrepancy between the homologs within the group is large. Conversely, thermophilic homologs averaged a higher number of pi-pi interactions and an increased number of hydrogen bonds and hydrophobic interactions within the monomer that displayed an increased rigidity to maintain their structure (calculated numbers are presented in **Supplementary Table S3**). Oligomerization is expected to have a stabilizing effect and the calculation of the interactions within the trimers for CA\_D and homologs demonstrates higher amount of interactions compared to monomers alone. Interestingly, while the CA\_D monomer does not show a statistically significant increase in interactions compared to single homologs, the trimerization of CA\_D stabilizes the enzyme by additional interactions on the monomer interfaces to a higher extent than for homologs. The CA\_D trimer displayed an increased number of hydrogen bonds and salt bridges compared to all of the individual meso- and thermophilic homologs being additionally added to the enzyme

upon trimerization (hydrogen bonds: 69 vs. homolog average of 42; salt bridges: 10 vs. a homolog average of 5 or 5.5). As the  $\gamma$ -CA\_D trimer constitutes the active form, this increased stabilization is critical to assemble the active enzyme under high salt conditions.

## Comparison of the Surface-Charge

Halophilic proteins are typically characterized by a highly negative charged electrostatic surface (DasSarma and DasSarma, 2015). Interestingly, while the CA\_D monomer shows both positive and negative charges on its surface, the overall charge is slightly more negative ( $pI \sim 6.8$ ). The highest negative charge is found on the surface of the flexible  $\beta 10$ – $\beta 11$  loop extending from the  $\beta$ -helix (Figure 2A). The presence and sequence of this loop differs between homolog  $\gamma$ -CAs (Park et al., 2012). For CA\_D it shows a high acidic surface charge. A highly acidic surface is an important and common trait of halophilic proteins, enabling protein hydration under high salt conditions (Frolow et al., 1996; Britton et al., 2006; Grotzinger et al., 2018).

The CA\_D trimer contains an overall negative surface potential, while a positive charge is located only in two concentrated locations. The positive charge is possibly involved in the fast release of the bicarbonate product. The overall negative charge indicates the burying of positive patches found on the monomer surface (Figure 2B). Compared to mesophilic homologs, CA\_D showed a slightly higher negative surface charge except for *E. coli* and *C. difficile*  $\gamma$ -CAs (PDB ID: 3tio, 4mfg) (Figure 2C and Supplementary Figure S3). Still, this negative surface-charge is lower than often described for halophilic proteins and possibly explains the flexibility that enables stability even at comparatively low salt concentrations.

## Comparison of the Surface-Exposed Amino Acids

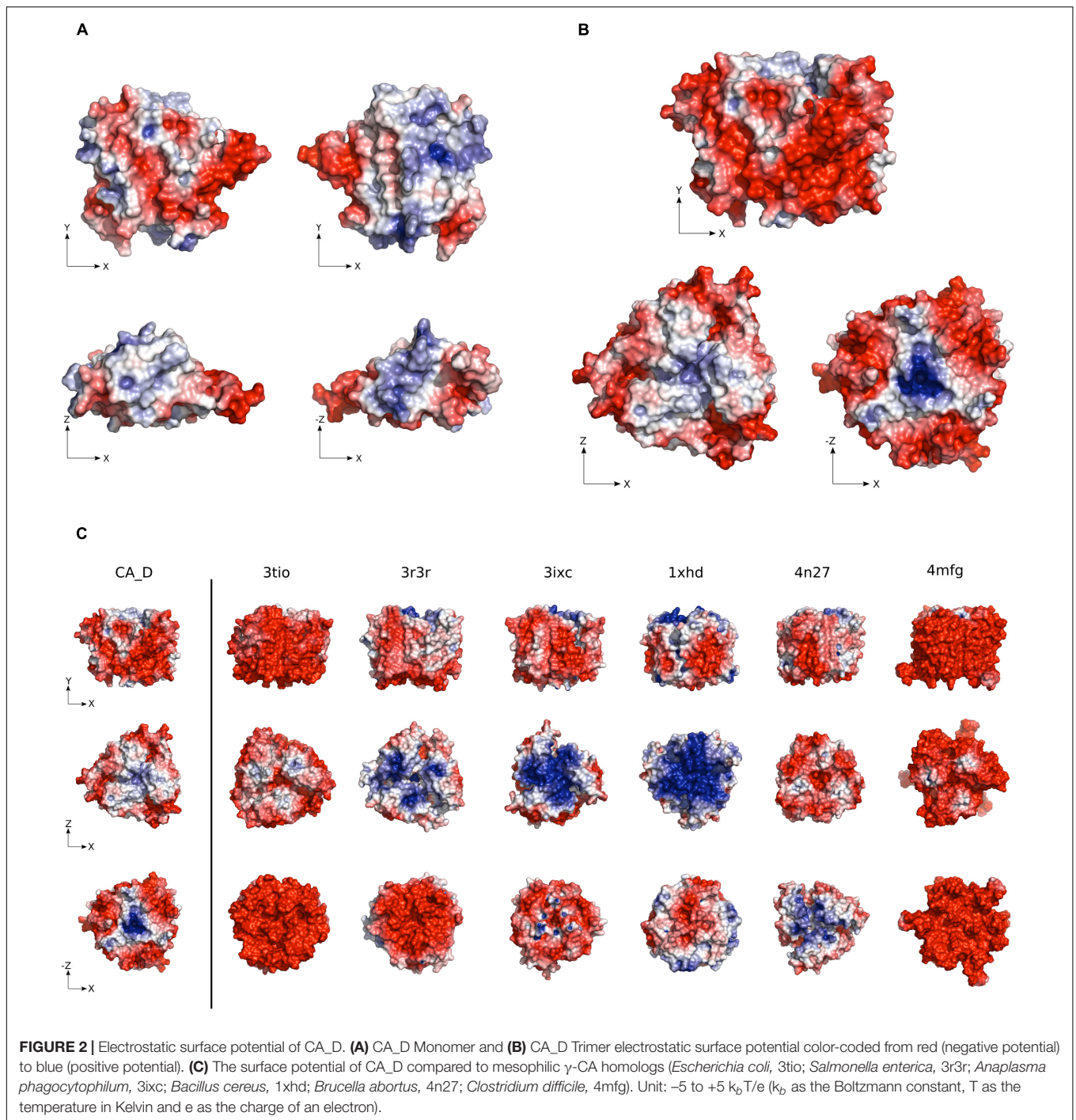
The most pronounced difference found for CA\_D compared to homologs is the surface amino acid composition. The CA\_D trimer protein surface contains a decreased number of hydrophobic and polar amino acid residues (Supplementary Figure S4) whilst showing an increased number of charged amino acid residues on the surface as compared to homologs. 56% of the amino acids on the CA\_D protein surface are charged, compared to 38 and 39% in meso- and thermophilic homologs, respectively (Figure 3A). This decrease in hydrophobic amino acids leads to a decrease of the hydrophobic surface patch that facilitates the aggregation of the protein in CA\_D (hydrophobic amino acids constitute 29% of the surface amino acids compared to an average of 36 and 39% for meso- and thermophilic homologs, respectively). An increased amount of charged amino acids is observed that form a stable hydration shell. This is essential for stability under high salt conditions. A closer look at positively and negatively charged amino acids on the trimer surface of CA\_D reveals the dominance of glutamate (23.6 vs. 10.5 and 14.4% on the surface of mesophilic and thermophilic homologs, respectively) (Figure 3B). Interestingly, CA\_D also contains a comparatively high amount of lysines.

The increase of surface glutamic acid residues described for CA\_D is a characteristic hallmark of halophilic enzymes (DasSarma et al., 2006), as these residues have a high water-binding capacity via interaction with  $Na^+$  or  $K^+$  ions, and thereby attract the bound hydrating water (Trevino et al., 2007). This explains how the strong increase of glutamic acid on the surface of CA\_D has such a pronounced effect on maintaining stability and activity at high salt concentrations. However, contradictory to reported halo-adaptation strategies, CA\_D also shows increased surface lysine residues, compared to homologs. This explains why the negative surface charge is not as pronounced as for other reported halophilic proteins (Premkumar et al., 2005; Karan et al., 2012a; DasSarma et al., 2013). Lysine tends to disrupt the formation of an ordered hydration shell under elevated salt concentrations (Britton et al., 2006; Esclapez et al., 2007). At low ionic strength, the higher amount of glutamic acid on the surface causes electrostatic repulsion and, therefore, destabilizes the protein (Kohn et al., 1995; Frolow et al., 1996; Britton et al., 2006). Moreover, the long hydrophobic part of the lysine residues potentially plays a role in attracting hydrophobic  $CO_2$  substrate for the catalytic reaction. Noteworthy, the negatively charged glutamate residues are located around the trimer surface, except for the top and bottom view of the multimer where positive patches stem from a network of arginine residues which are possibly initiating the rapid release of the formed product (Smith et al., 2002; Tripp et al., 2002).

## CA\_D Variant Library Design

Since the discovery of the first  $\gamma$ -CA in 1994 (Alber and Ferry, 1994), several  $\gamma$ -CAs have been characterized: some with a high reported activity and some with no detectable activity, raising the question whether essential residues are missing or an alternative function is appropriate (Macauley et al., 2009). We bioengineered the CA\_D active center based on the CA\_D crystal structure. To investigate the role of selected residues in the active site of CA\_D, several variants were expressed, purified, and assayed for enzymatic activity. The selection of residues for mutagenesis was performed based on structural comparisons to  $\gamma$ -CA homologs as well as literature reports based on presumptions of conserved residues of  $\gamma$ -CAs (Smith et al., 1999; Iverson et al., 2000; Jeyakanthan et al., 2008; Ferry, 2010; Pena et al., 2010; Park et al., 2012; Ragunathan et al., 2013; Frost and McKenna, 2014). Thereby, the main comparison was focused on the Cam structure (Kisker et al., 1996). A structural comparison of the active center of CA\_D (Figure 4A) with a simplified view of the Cam homolog (Figure 4B) revealed that CA\_D residue I46 corresponds to Cam residue E62. This part of the enzyme plays an important role in product release, relaying protons during hydroxide formation from the zinc-bound water as well as forming hydrogen bonds with the bicarbonate. CA\_D residues K58 and H166 are substituted by Q75 and N202 in Cam. These residues presumably orchestrate the orientation of the carbon dioxide for the nucleophilic attack in Cam. Moreover, Cam N202 together with E62 is thought to form hydrogen bonds with the product bicarbonate (Ferry, 2010; Pena et al., 2010).



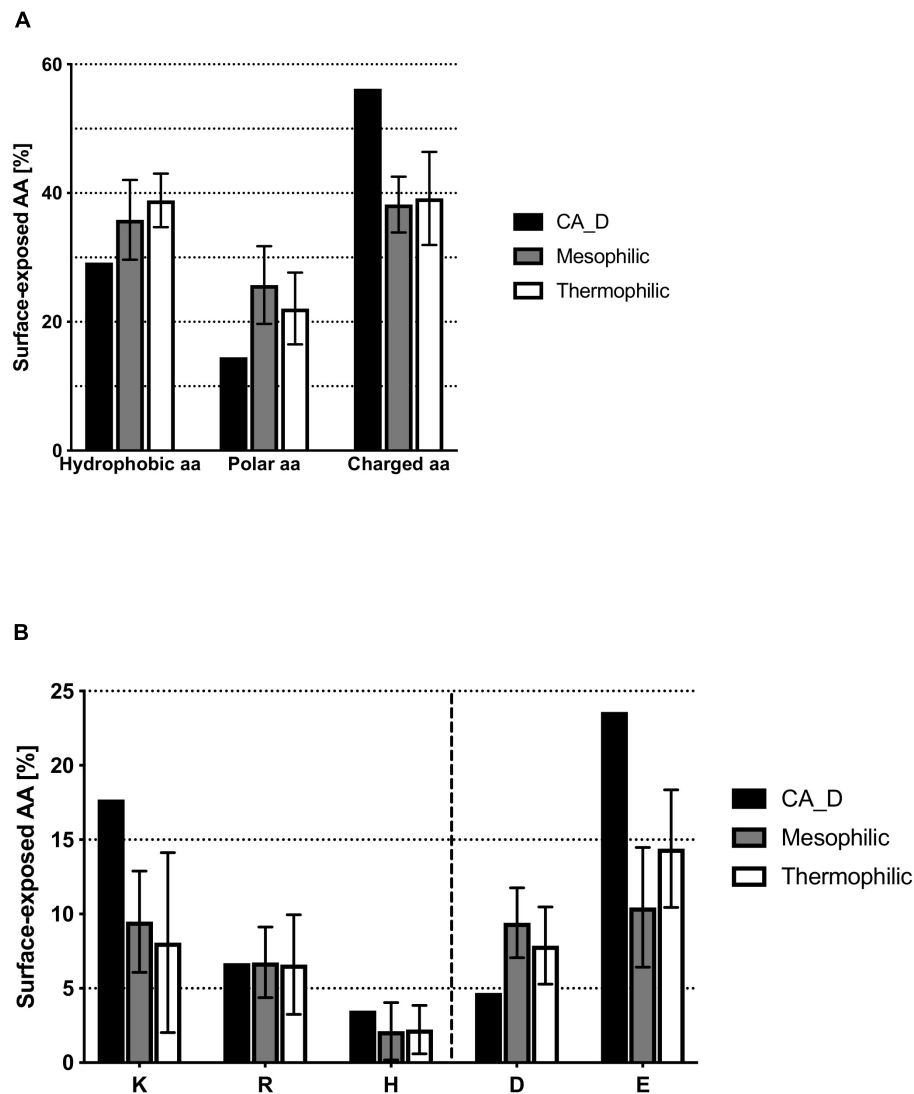


The strict conservation of the E84 position is a matter of debate, as the CamH subclass has lost this residue (Ferry, 2010) and a mutagenesis study exhibits activity for the D84 and H84 substituted Cam variants (Tripp and Ferry, 2000). The presence of D residues instead of E is explained by the different abundance of aspartic and glutamic acid in thermophilic proteins, compared to mesophilic proteins (Lee et al., 2004). Thus, the selected point mutations were I46E, K58Q, H166N, I46E-K58Q, K58Q-H166N, I46E-H166N as well as a triple

CA\_D\* (I46E-K58Q-H166N) (Figure 4C) and a quadruple CA\_D\*-D67E (I46E-K58Q-H166N-D67E) variant (Table 1 and Supplementary Figures S5, S6).

### Carbonic Anhydrase Activity Measurement of CA\_D and Variants

To evaluate the potential impact of the mutations, we examined the CA\_D variants for their activities expressed



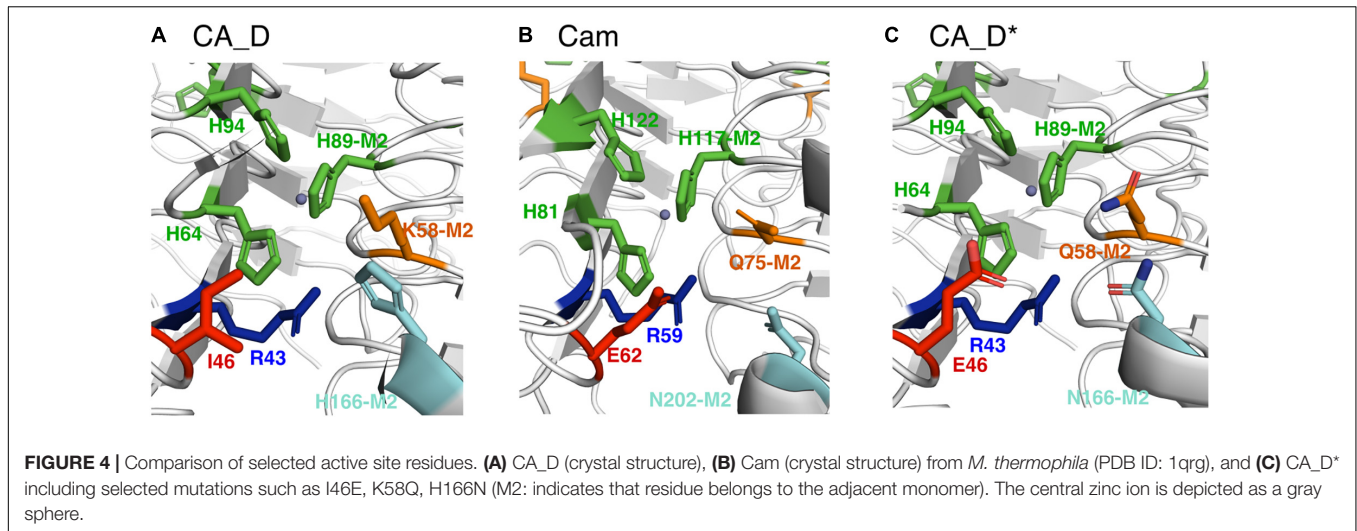
**FIGURE 3** | Comparison of surface-exposed amino acids between CA\_D and meso- and thermophilic homologs. **(A)** hydrophobic, polar or charged amino acids residues, **(B)** depicts the amount per individual charged amino acid.

in Wilbur-Anderson Unit per 1 mg of protein (WAU/mg) (Wilbur and Anderson, 1948; Hou et al., 2019). The colorimetric carbonic anhydrase activity assay by Wilbur-Anderson measures the time required for a saturated  $\text{CO}_2$  solution to lower the pH of a specific buffer. Respective controls were: (i) a carbonic anhydrase from bovine erythrocytes as a positive control; and (ii) the crude lysate of untransformed *Halobacterium* sp. NRC-1  $\Delta\text{ura3}\Delta\text{icfA}$  cells as negative controls (Supplementary Figure S7). Purified lysate and the crude lysate did not show any enzyme activity. Therefore, these measurements confirmed that the observed activity resulted from the purified CA\_D variants. As CA\_D is from an uncultured archaeon from the Discovery Deep brine pool, the enzyme activity was measured at 40°C in a solution containing 3 M KCl.

The activity assays displayed a distinct profile for CA\_D with a rather low activity of ~33 WAU/mg. In contrast, the engineered variants showed a decreased activity for I46E, K58Q, and H166N, and undetectable activity for the double variants, I46E-K58Q, K58Q-H166N, and I46E-H166N (Figure 5 and Supplementary Table S4). However, the triple variant CA\_D\* (I46E-K58Q-H166N) displayed a 17-fold increased activity of 566 WAU/mg (Figure 5 and Supplementary Table S4) compared to CA\_D.

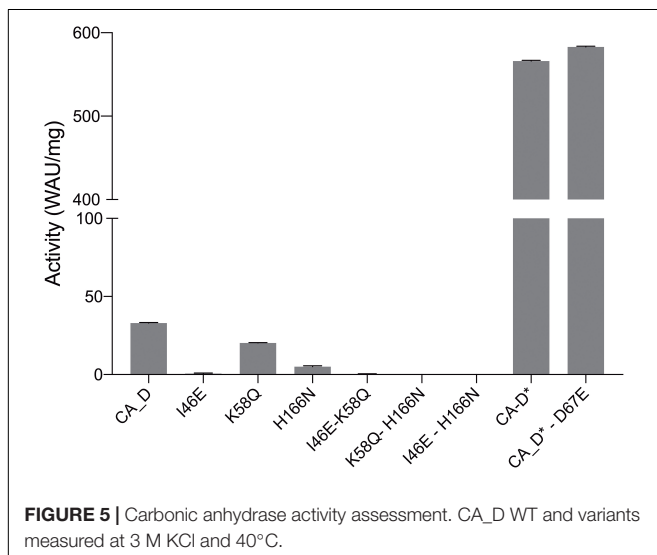
These results are in close alignment with previous findings of related  $\gamma$ -carbonic anhydrases ( $\gamma$ -CAs). In the most prominent  $\gamma$ -CA Cam, residue E84 has been proposed as part of a proton shuttling network along with E62, E88, and E89 on the acidic loop (Tripp and Ferry, 2000). Conversely, when fully mimicking the expected proton shuttling network, activity of the quadruple



**TABLE 1** | Rationalization of CA\_D variants.

Constructs	Residues							
1QRG (Cam)	R59	E62	Q75	E84	N202	H81	H117	H122
CA_D	R43	<b>I46</b>	<b>K58</b>	D67	<b>H166</b>	H64	H89	H94
I46E	R43	<b>E46</b>	K58	D67	H166	H64	H89	H94
K58Q	R43	I46	<b>Q58</b>	D67	H166	H64	H89	H94
H166N	R43	I46	K58	D67	<b>N166</b>	H64	H89	H94
I46E-K58Q	R43	<b>E46</b>	<b>Q58</b>	D67	H166	H64	H89	H94
K58Q-H166N	R43	I46	<b>Q58</b>	D67	<b>N166</b>	H64	H89	H94
I46E-H166N	R43	<b>E46</b>	K58	D67	<b>N166</b>	H64	H89	H94
CA_D* (I46E-K58Q-H166N)	R43	<b>E46</b>	<b>Q58</b>	D67	<b>N166</b>	H64	H89	H94
CA_D*-D67E	R43	<b>E46</b>	<b>Q58</b>	<b>E67</b>	<b>N166</b>	H64	H89	H94

The residues in bold were modified based on the alignment with Cam shown in the first row.



variant CA\_D\*-D67E (**Figure 5**) is slightly higher as for the triple variant CA\_D\*, which is in agreement with the observations made in  $\gamma$ -CA Cam (Tripp and Ferry, 2000; Ferry, 2010).

Our findings support the proposed mechanism for Cam, in which the probed amino acids play a decisive role. Some conclusions, as to why CA\_D is lacking activity could be drawn by comparing the altered residues in CA\_D\* to  $\gamma$ -CA Cam (Ferry, 2010). With the CA\_D mutation I46E, a hydrophobic residue was replaced with a negatively charged one, the corresponding residue in the Cam homolog presumably is essential in water activation to enable the reaction (Ferry, 2010). K58Q and H166N are further substitutions based on Cam active residues and facilitate the orientation of the carbon dioxide, while the H166N substitution facilitates hydrogen bonding and release of the bicarbonate (Ferry, 2010). We, therefore, assume that CA\_D is regulated by a similar proton shuttling network such as Cam (Ferry, 2010), but due to its halophilic nature CA\_D behaves slightly differently.

## CONCLUSION

High cytoplasmic salt concentrations critically affect the folding and activity of proteins and other macromolecules as they may induce protein aggregation due to enhanced hydrophobic interactions, increased hydration of ions, decreased

availability of unbound water molecules, and prevention of intra- and intermolecular electrostatic interactions (Karan et al., 2012a,b). Halophilic proteins are adapted to maintain their native conformation under high salt concentrations. They are functionally active in the presence of high salt concentrations, following halo-adaptation strategies such as high acidic amino acid content on the surface, low hydrophobicity at the core of the protein, and an increased number of salt bridges (Madern et al., 2000; Fukuchi et al., 2003; Bolhuis et al., 2008; Tadeo et al., 2009; DasSarma et al., 2013).

We used single amplified genomes to resuscitate a  $\gamma$ -carbonic anhydrase ( $\gamma$ -CA) from an uncharacterized haloarchaeon collected from a brine pool at the bottom of the Red Sea. The detailed structural analysis and comparison with previously solved structures of mesophilic  $\gamma$ -CAs revealed the molecular features of its extremophilic nature, caused by the unique habitat. The most prominent features of extremophilicity are the increased charged residues on the protein surface and an increased number of hydrogen bonds as well as salt bridges. Investigation of CA\_D, therefore, demonstrates potential for further development and implementation of SAG analysis to generate extremozymes from previously inaccessible environments. This approach has to date only been scarcely used to characterize specific proteins of interest (Grotzinger et al., 2018). Further, mutation analysis provided an interesting insight into active site residue conservation for  $\gamma$ -CAs and enabled us to increase the CA\_D function by 17-fold. Moreover, the applied combination of mechanistic insights from the thermophilic Cam into the scaffold of the halophilic CA\_D, resulting in the CA\_D\* variant which demonstrated high activity and stability, underlines the potential of protein evolution for extremophilic proteins for industrial applications and the design of novel catalysts for industry.

## REFERENCES

- Akal, A. L., Karan, R., Hohl, A., Alam, I., Vogler, M., Grötzinger, S. W., et al. (2019). A polyextremophilic alcohol dehydrogenase from the Atlantis II Deep Red Sea brine pool. *FEBS Open Biol.* 9, 194–205. doi: 10.1002/2211-5463.12557
- Alam, I., Antunes, A., Kamau, A. A., Ba Alawi, W., Kalkatawi, M., Stingl, U., et al. (2013). INDIGO – Integrated data warehouse of microbial genomes with examples from the red sea extremophiles. *PLoS One* 8:e82210. doi: 10.1371/journal.pone.0082210
- Alber, B. E., Colangelo, C. M., Dong, J., Stalhandske, C. M., Baird, T. T., Tu, C., et al. (1999). Kinetic and spectroscopic characterization of the gamma-carbonic anhydrase from the methanoarchaeon *Methanosarcina thermophila*. *Biochemistry* 38, 13119–13128. doi: 10.1021/bi9828876
- Alber, B. E., and Ferry, J. G. (1994). A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6909–6913. doi: 10.1073/pnas.91.15.6909
- Amoozgar, M. A., Safarpour, A., Akbari Noghabi, K., Bakhtiary, T., and Ventosa, A. (2019). Halophiles and their vast potential in biofuel production. *Front. Microbiol.* 10:1895. doi: 10.3389/fmicb.2019.01895
- Antunes, A., Ngugi, D. K., and Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* 3, 416–433. doi: 10.1111/j.1758-2229.2011.00264.x
- Antunes, A., Simões, M. F., Grötzinger, S. W., Eppinger, J., Bragança, J., and Bajic, V. B. (2017). “Bioprospecting archaea: focus on extreme halophiles,” in *Bioprospecting*, eds R. Paterson and N. Lima (Cham: Springer), 81–112. doi: 10.1007/978-3-319-47935-4\_5

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the PDB ID: 6SC4.

## AUTHOR CONTRIBUTIONS

MR and JE conceived and supervised the study. MV and RK designed and performed the experiments with the help of DR and AV. SG, PD, and SD provided plasmid and support. M-TV and MG solved the crystal structure. MV and RK wrote the manuscript. All the authors contributed to and commented on this manuscript.

## FUNDING

The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia. The staff of the Beamline X06SA at the Paul Scherrer Institute, SLS, Villigen, Switzerland, is acknowledged for assistance during data collection. Work in the DasSarma laboratory was supported by grant 80NSSC19K0463.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00742/full#supplementary-material>

- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi: 10.1073/pnas.181342398
- Behzad, H., Ibarra, M. A., Mineta, K., and Gojobori, T. (2016). Metagenomic studies of the Red Sea. *Gene* 576(2 Pt 1), 717–723. doi: 10.1016/j.gene.2015.10.034
- Berquist, B. R., Müller, J. A., and DasSarma, S. (2006). 27 genetic systems for *Haloflex volcanii*. *Methods Microbiol.* 35, 649–680. doi: 10.1007/s00792-015-0794-6
- Bolhuis, A., Kwan, D., and Thomas, J. (2008). “Halophilic adaptations of proteins,” in *Protein Adaptation in Extremophiles*, eds K. S. Siddiqui, T. Thomas (New York: Nova Science Publishers Inc).
- Britton, K. L., Baker, P. J., Fisher, M., Ruzhenikov, S., Gilmour, D. J., Bonete, M. J., et al. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloflex mediterranei*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4846–4851. doi: 10.1073/pnas.0508854103
- Cavicchioli, R., Amils, R., Wagner, D., and McGenity, T. (2011). Life and applications of extremophiles. *Environ. Microbiol.* 13, 1903–1907. doi: 10.1111/j.1462-2920.2011.02512.x
- DasSarma, S., Berquist, B. R., Coker, J. A., DasSarma, P., and Muller, J. A. (2006). Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1. *Saline Syst.* 2:3. doi: 10.1186/1746-1448-2-3
- DasSarma, S., Capes, M. D., Karan, R., and DasSarma, P. (2013). Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from antarctica. *PLoS One* 8:e58587. doi: 10.1371/journal.pone.0058587

- DasSarma, S., and DasSarma, P. (2015). Halophiles and their enzymes: negativity put to good use. *Curr. Opin. Microbiol.* 25, 120–126. doi: 10.1016/j.mib.2015.05.009
- DasSarma, S., Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., and Fleischmann, E. M. (1995). *Archaea: A Laboratory Manual – Halophiles*. Cham: Springer.
- Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095–1099. doi: 10.1101/gr.180501
- Del Prete, S., Vullo, D., De Luca, V., Carginale, V., Osman, S. M., AlOthman, Z., et al. (2016). Cloning, expression, purification and sulfonamide inhibition profile of the complete domain of the  $\eta$ -carbonic anhydrase from *Plasmodium falciparum*. *Bioorg. Med. Chem. Lett.* 26, 4184–4190. doi: 10.1016/j.bmcl.2016.07.060
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res. (Suppl.)* 2, W665–W667.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2126–2132. doi: 10.1107/S0907444904019158
- Escalpez, J., Pire, C., Bautista, V., Martínez-Espinoza, R., Ferrer, J., and Bonete, M. (2007). Analysis of acidic surface of *Haloferax mediterranei* glucose dehydrogenase by site-directed mutagenesis. *FEBS Lett.* 581, 837–842. doi: 10.1016/j.febslet.2007.01.054
- Ferry, J. G. (2010). The gamma class of carbonic anhydrases. *Biochim. Biophys. Acta* 1804, 374–381. doi: 10.1016/j.bbapap.2009.08.026
- Ferry, J. G. (2013). Carbonic anhydrases of anaerobic microbes. *Bioorg. Med. Chem.* 21, 1392–1395. doi: 10.1016/j.bmc.2012.12.005
- Frolow, F., Harell, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.* 3:452. doi: 10.1038/nsb0596-452
- Frost, S. C., and McKenna, R. (2014). *Carbonic Anhydrase Mechanism, Regulation, Links to Disease, and Industrial Applications*. Cham: Springer.
- Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327, 347–357. doi: 10.1016/s0022-2836(03)00150-5
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). “Protein identification and analysis tools on the ExpASY server,” in *The Proteomics Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-890-0:571
- Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., et al. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–W531. doi: 10.1093/nar/gki376
- Grotzinger, S. W., Alam, I., Ba Alawi, W., Bajic, V. B., Stingl, U., and Eppinger, J. (2014). Mining a database of single amplified genomes from Red Sea brine pool extremophiles-improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA). *Front. Microbiol.* 5:134. doi: 10.3389/fmicb.2014.00134
- Grotzinger, S. W., Karan, R., Strillinger, E., Bader, S., Frank, A., Al Rowaihi, I. S., et al. (2018). Identification and experimental characterization of an extremophilic brine pool alcohol dehydrogenase from single amplified genomes. *ACS Chem. Biol.* 13, 161–170. doi: 10.1021/acscchembio.7b00792
- Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. doi: 10.1002/elps.1150181505
- Gurvich, E. G. (2006). *Metalliferous Sediments of the World Ocean*. Berlin: Springer.
- Herrou, J., and Crosson, S. (2013). Molecular structure of the *Brucella abortus* metalloprotein RicA, a Rab2-binding virulence effector. *Biochemistry* 52, 9020–9028. doi: 10.1021/bi401373r
- Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–W549. doi: 10.1093/nar/gkq366
- Hou, J., Li, X., Kaczmarek, M. B., Chen, P., Li, K., Jin, P., et al. (2019). Accelerated CO<sub>2</sub> hydration with thermostable sulfurhydrogenibium azorene carbonic anhydrase-chitin binding domain fusion protein immobilised on chitin support. *Int. J. Mol. Sci.* 20:1494. doi: 10.3390/ijms20061494
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Hunt, J. M., Hays, E. E., Degens, E. T., and Ross, D. A. (1967). Red sea: detailed survey of hot-brine areas. *Science* 156, 514–516. doi: 10.1126/science.156.3774.514
- Iverson, T. M., Alber, B. E., Kisker, C., Ferry, J. G., and Rees, D. C. (2000). A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39, 9222–9231. doi: 10.1021/bi000204s
- Jeyakanthan, J., Rangarajan, S., Mridula, P., Kanaujia, S. P., Shiro, Y., Kuramitsu, S., et al. (2008). Observation of a calcium-binding site in the gamma-class carbonic anhydrase from *Pyrococcus horikoshii*. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 10), 1012–1019. doi: 10.1107/S0907444908024323
- Jorquera, M. A., Graether, S. P., and Maruyama, F. (2019). Bioprospecting and biotechnology of extremophiles. *Front. Bioeng. Biotechnol.* 7:204.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26, 795–800. doi: 10.1107/s0021889893005588
- Karan, R., Capes, M. D., DasSarma, P., and DasSarma, S. (2013). Cloning, overexpression, purification, and characterization of a polyextremophilic beta-galactosidase from the Antarctic haloarchaeon *Haloalobosstratum lacusprofundi*. *BMC Biotechnol.* 13:3. doi: 10.1186/1472-6750-13-3
- Karan, R., Capes, M. D., and DasSarma, S. (2012a). Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* 8:4. doi: 10.1186/2046-9063-8-4
- Karan, R., Kumar, S., Sinha, R., and Khare, S. K. (2012b). *Halophilic Microorganisms as Sources of Novel Enzymes*. Cham: Springer, 555–579.
- Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996). A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J.* 15, 2323–2330. doi: 10.1002/j.1460-2075.1996.tb00588.x
- Kohn, W. D., Kay, C. M., and Hodges, R. S. (1995). Protein destabilization by electrostatic repulsions in the two-stranded  $\alpha$ -helical coiled-coil/leucine zipper. *Protein Sci.* 4, 237–250. doi: 10.1002/pro.5560040210
- Krieger, E., and Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* 30, 2981–2982. doi: 10.1093/bioinformatics/btu426
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022
- Kvist, T., Ahring, B. K., Lasken, R. S., and Westermann, P. (2007). Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* 74, 926–935. doi: 10.1007/s00253-006-0725-727
- Lee, D. Y., Kim, K.-A., Yu, Y. G., and Kim, K.-S. (2004). Substitution of aspartic acid with glutamic acid increases the unfolding transition temperature of a protein. *Biochem. Biophys. Res. Commun.* 320, 900–906. doi: 10.1016/j.bbrc.2004.06.031
- Liszka, M. J., Clark, M. E., Schneider, E., and Clark, D. S. (2012). Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* 3, 77–102. doi: 10.1146/annurev-chembioeng-061010-114239
- Littlechild, J. A. (2015). Archaeal enzymes and applications in industrial biocatalysts. *Archaea* 2015:147671. doi: 10.1155/2015/147671
- Littlechild, J. A. (2017). Improving the ‘tool box’ for robust industrial enzymes. *J. Ind. Microbiol. Biotechnol.* 44, 711–720. doi: 10.1007/s10295-017-1920-1925
- Macauley, S. R., Zimmerman, S. A., Apolinario, E. E., Evilia, C., Hou, Y. M., Ferry, J. G., et al. (2009). The archetype gamma-class carbonic anhydrase (Cam) contains iron when synthesized in vivo. *Biochemistry* 48, 817–819. doi: 10.1021/bi802246s
- Madern, D., Ebel, C., and Zaccari, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142
- Madigan, M. T., and Mairs, B. L. (1997). Extremophiles. *Sci. Am.* 276, 82–87. doi: 10.1038/scientificamerican0497-82
- Meldrum, N. U., and Roughton, F. J. (1933). Carbonic anhydrase. Its preparation and properties. *J. Physiol.* 80, 113–142. doi: 10.1113/jphysiol.1933.sp003077
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780



- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53(Pt 3), 240–255. doi: 10.1107/S0907444996012255
- Mwirichia, R., Alam, I., Rashid, M., Vinu, M., Ba-Alawi, W., Anthony Kamau, A., et al. (2016). Metabolic traits of an uncultured archaeal lineage—MSBL1—from brine pools of the Red Sea. *Sci. Rep.* 6:19181. doi: 10.1038/srep19181
- Park, H. M., Park, J. H., Choi, J. W., Lee, J., Kim, B. Y., Jung, C. H., et al. (2012). Structures of the gamma-class carbonic anhydrase homologue YrdA suggest a possible allosteric switch. *Acta Crystallogr. D Biol. Crystallogr.* 68(Pt 8), 920–926. doi: 10.1107/S0907444912017210
- Pena, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2455–2460. doi: 10.1073/pnas.0910866107
- Persidis, A. (1998). Extremophiles. *Nat. Biotechnol.* 16, 593–594. doi: 10.1038/nbt0698-593
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Premkumar, L., Greenblatt, H. M., Bageshwar, U. K., Savchenko, T., Gokhman, I., Sussman, J. L., et al. (2005). Three-dimensional structure of a halotolerant algal carbonic anhydrase predicts halotolerance of a mammalian homolog. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7493–7498. doi: 10.1073/pnas.0502829102
- Ragunathan, P., Raghunath, G., Kuramitsu, S., Yokoyama, S., Kumarevel, T., and Ponnuraj, K. (2013). Crystallization, characterization and preliminary X-ray crystallographic analysis of GK2848, a putative carbonic anhydrase of *Geobacillus kaustophilus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 69(Pt 2), 162–164. doi: 10.1107/S1744309112051913
- Rashid, M., and Stingl, U. (2015). Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnol. Adv.* 33, 1755–1773. doi: 10.1016/j.biotechadv.2015.09.005
- Smith, K. S., Ingram-Smith, C., and Ferry, J. G. (2002). Roles of the conserved aspartate and arginine in the catalytic mechanism of an archaeal  $\beta$ -class carbonic anhydrase. *J. Bacteriol.* 184, 4240–4245. doi: 10.1128/jb.184.15.4240-4245.2002
- Smith, K. S., Jakubzick, C., Whittam, T. S., and Ferry, J. G. (1999). Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 15184–15189. doi: 10.1073/pnas.96.26.15184
- Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160. doi: 10.1128/JB.00345-312
- Supuran, C. T. (2008). Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* 7, 168–181. doi: 10.1038/nrd2467
- Supuran, C. T. (2016). Structure and function of carbonic anhydrases. *Biochem. J.* 473, 2023–2032. doi: 10.1042/BCJ20160115
- Supuran, C. T., and Capasso, C. (2017). Carbonic anhydrase from porphyromonas gingivalis as a drug target. *Pathogens* 6:30. doi: 10.3390/pathogens6030030
- Supuran, C. T., and De Simone, G. (eds). (2015). “Carbonic anhydrases: an overview,” in *Carbonic Anhydrases as Biocatalysts* (Cham: Springer), 3–13.
- Tadeo, X., Lopez-Mendez, B., Trigueros, T., Lain, A., Castano, D., and Millet, O. (2009). Structural basis for the amino acid composition of proteins from halophilic archaea. *PLoS Biol.* 7:e1000257. doi: 10.1371/journal.pbio.1000257
- Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* 366, 449–460. doi: 10.1016/j.jmb.2006.10.026
- Tripp, B. C., and Ferry, J. G. (2000). A structure-function study of a proton transport pathway in the gamma-class carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39, 9232–9240. doi: 10.1021/bi0001877
- Tripp, B. C., Tu, C., and Ferry, J. G. (2002). Role of arginine 59 in the  $\gamma$ -class carbonic anhydrases. *Biochemistry* 41, 669–678. doi: 10.1021/bi010768b
- Wilbur, K. M., and Anderson, N. G. (1948). Electrometric and colorimetric determination of carbonic anhydrase. *J. Biol. Chem.* 176, 147–154.
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Vogler, Karan, Renn, Vancea, Vielberg, Grötzinger, DasSarma, DasSarma, Eppinger, Groll and Rueping. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.