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

Malvina Vogler$^{1,2}$†, Ram Karan$^1$*, Dominik Renn$^1$, Alexandra Vancea$^1$, Marie-Theres Vielberg$^2$, Stefan W. Grötzinger$^1$, Priya DasSarma$^3$, Shiladitya DasSarma$^3$, Jörg Eppinger$^1$, Michael Groll$^2$* and Magnus Rueping$^1$*

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

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 γ-carbonic anhydrase (γ-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 (CO$_2$ + H$_2$O $\rightleftharpoons$ HCO$_3^-$ + H$^+$) (Supuran, 2008). To date, seven classes have been described: α-, β-, γ-, δ-, ζ-, η-, and θ-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 α-CA class was the first to be isolated and described (Meldrum and Roughton, 1933), whereas the γ-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 β-helix fold, and central metal coordination in the active site by three histidine residues (Kisker et al., 1996). Subsequently, further γ-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 γ-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 γ-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® 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® 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 μg/ml ampicillin. *Halobacterium* strains were cultured in 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, *icaA* was knocked out via the *ura3*-based gene deletion method for *Halobacterium* sp. NRC-1. Approximately 500 bp regions flanking the carbonic anhydrase gene (*icaA*) 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 HindIII and EcoRI sites incorporated in the primers (Supplementary Table S5) (Berquist et al., 2006). The resulting plasmid, pBB400ΔicfA was transformed into Halobacterium sp. NRC-1Δura3 using the standard PEG-EDTA method (DasSarma et al., 1995). pBB400ΔicfA transformants were selected by plating on CM+ uracil dropout media (HURA), colonies picked and grown in liquid HURA media, and integrant candidates were plated onto 5-FOA-CM+ 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 His6-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-1Δura3ΔicfA and Purification of the Encoded Proteins

Carbonic anhydrase genes containing vectors were transformed into the Halobacterium sp. NRC-1Δura3ΔicfA strain using the PEG/EDTA method (DasSarma et al., 1995; Karan et al., 2013) and transformants were selected by plating on CM+ agar plates using mevinolin resistance. For protein production, cells were grown to late log phase (OD600 of 0.9–1.0) at 42°C in CM+ medium supplemented with 20 µg/ml mevinolin. For protein production, cells were grown to late log phase (OD600 of 0.9–1.0) at 42°C in CM+ medium supplemented with 20 µg/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 × 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 × 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 µm 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 Ni2+ 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−1 cm−1 (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 µg) 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 µM 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 CO2-saturated water prepared by bubbling CO2 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 µM) 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-1Δura3ΔicfA 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 (T0 – T)/T, where T0 (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
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 α = 362.6 Å, indicating five γ-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_r-F_e 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 Rfree = 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)¹ 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) (>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.

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 γ-carbonic anhydrase (γ-CA) using the PPM algorithm (Grotzinger et al., 2014). We used the genetically modifiable extreme halarchaeon 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-1Δura3ΔicfA 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 γ-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 β-helix connected to an α-helix running antiparallel to the β-helix axis (Figure 1A). The majority of the structure comprises β-sheet (54.5%), while only a minimal amount consists of α-helix (13.5%), turn (12.9%), and coil structure (19.1%). The overall structure is highly conserved compared to reported γ-CAs (Figure 1B). Whereas the protein core is common to published γ-CA structures, differences in the connection of the helix motif is conspicuous. For example, CA_D contains a connecting β-sheet, while Cam consists of an additional α-helix instead. However, the comparison of CA_D with all 10 crystallized γ-CA structures revealed that the majority contained a β-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 β-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 γ-CAs, where the active site includes residues from the adjacent monomers.
Halophilic Adaptation of CA_D Compared to Other Non-halophilic γ-CAs

While the overall CA_D structure is conserved to other known γ-CAs, differences to non-halophilic γ-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 γ-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 α 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 change is slightly more negative (pI ~6.8). The highest negative charge is found on the surface of the flexible β10–β11 loop extending from the β-helix (Figure 2A). The presence and sequence of this loop differs between homolog γ-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* γ-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 residues 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 CO₂ 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 γ-CA in 1994 (Alber and Ferry, 1994), several γ-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 γ-CA homologs as well as literature reports based on presumptions of conserved residues of γ-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
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 CO₂ 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-1Δ*ura3ΔicfA* 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 γ-carbonic anhydrases (γ-CAs). In the most prominent γ-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
FIGURE 4 | Comparison of selected active site residues. (A) CA_D (crystal structure), (B) Cam (crystal structure) from M. thermophila (PDB ID: 1qrg), and (C) CA_D* including selected mutations such as I46E, K58Q, H166N (M2: indicates that residue belongs to the adjacent monomer). The central zinc ion is depicted as a gray sphere.

TABLE 1 | Rationalization of CA_D variants.

| Constructs                  | Residues |
|-----------------------------|----------|
| 1QRG (Cam)                  | R59      |
| CA_D                        | R43      |
| I46E                        | E46      |
| K58Q                        | Q58      |
| H166N                       | E64      |
| I46E-K58Q                   | E46      |
| K58Q-H166N                  | E46      |
| CA_D* (I46E-K58Q-H166N)     | E46      |
| CA_D*-D67E                  | E46      |

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

FIGURE 5 | Carbonic anhydrase activity assessment. CA_D WT and variants measured at 3 M KCl and 40°C.

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 γ-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 γ-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 γ-carboxylic anhydrase (γ-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 γ-CA revealed the molecular features of its extremophilic nature, causing 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 γ-CAs and enabled us to increase the CA_D function by 17-fold. Moreover, the applied combination of mechanistic insights from the thermophilic Cm 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., Grotzinger, S. W., et al. (2019). A polyextremophilic α-helical dehydrogenase from the Atlantis II Deep Red Sea brine pool. FERS Open Biol. 9, 194–205. doi: 10.1022/2211-3463.12557

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-carboxylic anhydrase from the methanoarchaeon Methanosaricina thermophila. Biochemistry 38, 13119–13128. doi: 10.1021/bi9828876

Alber, B. E., and Ferry, J. G. (1994). A carboxylic anhydrase from the archaeon Methanosaricina thermophila. Proc. Natl. Acad. Sci. U.S.A. 91, 6909–6913. doi: 10.1073/pnas.91.15.6909

Amoozegar, M. A., Safarpour, A., Akbari Noghabi, K., Baktiari, 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., Grotzinger, 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

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.100794-6

Behzad, H., Ibarra, M. A., Mineta, K., and Gojoebiri, T. (2016). Metagenomic studies of the Red Sea. Gene 576(2 Pt 1), 717–723. doi: 10.1016/j.gene.2015.10.034

Bergquist, B. R., Müller, J. A., and DasSarma, S. (2006). 27 genetic systems for Haloflexax 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., Ruzheimikov, S., Gilmour, D. J., Bonete, M. J., et al. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile Haloflexax mediterranei. Proc. Natl. Acad. Sci. U.S.A. 103, 4846–4851. doi: 10.1073/pnas.0508854103

Cavicchioli, R., Amils, R., Wagner, D., and McGinity, 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 Halorobium lacusprofundi, an extremely halophilic microbe from antarctica. PLoS One 8:e58587. doi: 10.1371/journal.pone.0058587

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
DasSarma, S., and DasSarma, P. (2015). Halophiles and their enzymes: negtivity 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., Cargnale, V., Osman, S. M., AlOthman, Z., et al. (2016). Cloning, expression, purification and sulfonamide inhibition profile of the complete domain of the γ-carbonic anhydrase from *Plasmiodium falciparum*. *Biorg. Med. Chem. Lett.* 26, 4184–4190. doi: 10.1016/j.bmcl.2016.07.060

Dolijn, 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

Esclapez, J., Pire, C., Bautista, V., Martínez-Espinosa, R., Ferrer, J., and Bonete, E. (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, Frolow, F., Harell, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 10), 1012–1019. doi: 10.1107/S0907444996083232

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/S0021889993005588

Karar, R., Capes, M. D., DasSarma, P., and DasSarma, S. (2013). Cloning, overexpression, purification, and characterisation of a polyextremophilic beta-galactosidase from the Antarctic halarchaeon *Halorubrum 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. Bioys.* 8:4. doi: 10.1106/2046-9063-8-4

Karan, R., Kumar, S., Sinha, R., and Khare, S. K. (2012b). *Halophil 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: 10.1002/1460-2075.1996.tb00588.x

Kohn, W. D., Kay, C. M., and Hodges, R. S. (1995). Protein destabilization by electrostatic repulsions in the two-stranded α-helical coiled-coil/leucine zipper. *Protein Sci.* 4, 237–250. doi: 10.1002/prot.306040210

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

Krisinell, 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 mining of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* 74, 926–935. doi: 10.1007/s00253-006-0725-7

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/s10529-017-2920-1925

Maculey, S. R., Zimmerman, S. A., Apolinario, E. E., Evila, C., Hou, Y. M., Ferry, J. G., et al. (2009). The archetypal gamma-class carbonic anhydrase (Cam) contains iron when synthesized in vivo. *Biochemistry* 48, 817–819. doi: 10.1021/bi802246s

Madigan, M. T., and Marrs, 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.sp03077

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

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0727-8565(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
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 homologues 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, 7453–7458. doi: 10.1073/pnas.0502829102

Ragunathan, P., Raghunath, G., Kuramitsu, S., Yokoyama, S., Kumarevel, T., 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, 7453–7458. doi: 10.1073/pnas.0502829102

Tripp, B. C., and Ferry, J. G. (2000). A structure-function study of a proton transport pathway in the gamma-class carbonic anhydrase from *Methanosaeta thermophila*. *Biochemistry* 39, 9232–9240. doi: 10.1021/bi0001877

Wilbur, K. M., and Anderson, N. G. (1948). Electrometric and colorimetric determination of carbonic anhydrase. *J. Biol. Chem.* 176, 147–154.

Wisniewski, 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, Váncza, Vielberg, Grötzinger, DasSarma, and Supuran. *Carbonic Anhydrases as Biocatalysts* (Cham: Springer), 3–13.