Ectoine synthase: an iron-dependent member of the cupin superfamily

Laura Czech[†,‡], Astrid Hoeppner[§], Sander HJ Smits[§,¶], and Erhard Bremer[†,‡]

[†]Department of Biology, Laboratory for Microbiology, Philipps-University Marburg, Marburg, Germany
[‡]SYNMIKRO Research Center, Philipps-University Marburg, Marburg, Germany
[§]Center for Structural Studies, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany
[¶]Institute of Biochemistry, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

FUNCTIONAL CLASS

Enzyme; ectoine synthase (EctC), EC 4.2.1.108. EctC belongs to the carbon–oxygen lyases (EC 4.2) and more specifically to the hydro-lyases (EC 4.2.1), a group of enzymes that cleaves carbon–oxygen bonds. The EctC enzyme catalyzes in an Fe^{2+}-dependent fashion the last step of ectoine [(*S*)-2-methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid] synthesis through cyclocondensation of the EctA-formed substrate *N*-γ-acetyl-L-2,4-diaminobutyric acid (*N*-γ-ADABA) via a water-elimination reaction. The reaction product, ectoine, is a member of the so-called compatible solutes and serves as a cytoprotectant against osmotic stress and growth temperature extremes. It also functions as a chemical chaperone.

The ectoine synthase EctC is a dimeric protein[1,4] and a member of the cupin superfamily.[5,6] Each of the monomers harbors a catalytically critical Fe^{2+} ion. The three residues coordinating this metal protrude into the lumen of the cupin barrel in which the EctC-promoted enzyme reaction, a water elimination, is catalyzed, using N-γ-ADABA as the substrate.[1]

OCCURENCE

EctC is the diagnostic enzyme of the ectoine biosynthetic route,[7–10] a set of three enzymes that produces this effective microbial stress-protectant and chemical chaperone.[11–13] Biosynthesis of ectoine[14] initiates from the central microbial metabolite L-aspartate-β-semialdehyde.[9,10,15–17] It is mediated by the following enzymes: L-2,4-diaminobutyrate transaminase (EctB; EC 2.6.1.76),[18] L-2,4-diaminobutyrate acetyltransferase (EctA; EC 2.3.1.178),[19] and ectoine synthase (EctC; EC 4.2.1.108),[1] with L-2,4-diaminobutyrate
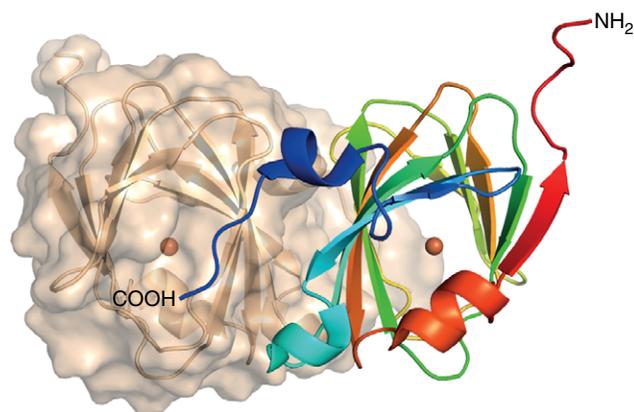
3D Structure Structure of the dimeric ectoine synthase from *Paenibacillus lautus* [(*Pl*)EctD)] (PDB code: 5onn).[1] One monomer is represented as cartoon representation, and in the other monomer the surface of the protein is represented. In both monomers, the bound Fe^{2+} ligand is shown as an orange sphere. The figures were prepared using PyMOL (www.pymol.org).[2] Adapted from Czech *et al.*[1] and Hanekop *et al.*[3]

and *N*-γ-acetyl-L-2,4-diaminobutyrate as the respective intermediates.[8–10]

Ectoine-producing microorganisms are primarily found in the members of the *Bacteria*, being mostly represented in the (α-, β-, γ-)-Proteobacteria, Actinobacteria, and Firmicutes.[11] Collectively, ectoine biosynthetic genes (*ectABC*) can be found in 10 bacterial phyla. However, they are only present in a rather restricted number of *Archaea*, representing two phyla (representatives are found in the Euryarchaeota and Thaumarchaeota).[11,20] Current evidences suggest that these *Archaea* have acquired the ectoine biosynthetic genes via lateral gene-transfer events from the members of the *Bacteria*.[11,20,21] In many ectoine producers, the tetrahydropyrimidine ring of the ectoine is further modified in a regio- and stereoselective reaction catalyzed by the ectoine hydroxylase (EctD, EC 1.14.11.55) to form 5-hydroxyectoine [(4*S*,5*S*)-5-hydroxy-2-methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid].[22–27] EctD is a member of the non-heme-containing iron(II) and 2-oxoglutarate-dependent dioxygenases[6] and is thus strictly oxygen dependent.[23,27] Ectoine/5-hydroxyectoine-producing *Bacteria* and *Archaea* can be ubiquitously found in fresh water, marine, terrestrial, and host-associated ecosystems, and some of them live in habitats with extremes in salinity, growth temperatures, or pH.[11]

In addition to the members of the *Bacteria* and *Archaea*, a selected group of halophilic unicellular ciliates are also ectoine/5-hydroxyectoine producers.[28–31] Like the *ect*-containing *Archaea*, these *Eukarya* have probably acquired the ectoine/5-hydroxyectoine biosynthetic genes via lateral gene-transfer events.[30,32] This is a rather likely evolutionary scenario because many of the prey bacteria hunted by these protists in their high-salinity habitats will probably be ectoine/5-hydroxyectoine producers. The acquired *ect* genes were then adjusted by the halophilic ciliates in their genetic structure (e.g. by introducing introns and changes in the genetic code) to principles governing the transcriptional and translational apparatus of the eukaryotic cell.[30,31] Strikingly, ectoine-producing microalgae were recently found among marine diatoms.[33] Because photosynthetic unicellular algae are major contributors to marine primary production by phytoplancton,[33] ectoine production by eukaryotic cells might be far more widespread than previously thought.

BIOLOGICAL FUNCTION

The major physiological function of the EctC-formed ectoine molecule (Figure 1) is to serve as a stress protectant against the detrimental effects of high osmolarity/salinity on growth and survival.[11–13,34–36] Ectoine is a prominent member of the compatible solutes;[37–39] it is highly water soluble (up to 7 M), and its physicochemical attributes make it fully compliant with the physiology and biochemistry of the cell.[40,41] Consequently, microbial
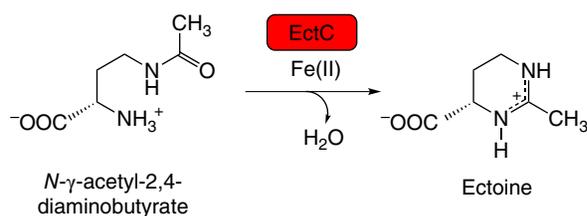
Figure 1 Scheme of the EctC-catalyzed cyclocondensation reaction. The Fe^{2+}-containing ectoine synthase converts the substrate *N*-γ-acetyl-2,4-diaminobutyrate into the stress-protectant ectoine through a cyclization reaction involving the elimination of a water molecule.

cells can accumulate ectoine via *de novo* synthesis to exceedingly high cellular levels. In this process, the degree of the imposed osmotic stress determines the intracellular concentration of ectoines in a finely tuned fashion.[42,43] Resulting from the increased ectoine pool, water efflux induced by high environmental osmolarity is counteracted, turgor is maintained in physiologically acceptable boundaries, an undue increase in molecular crowding is prevented, and the solvent properties of the cytoplasm are optimized for biochemical transformations.[40,41,44]

The ectoine biosynthetic genes are typically encoded in an operon (*ectABC*),[7,11,12] but other genetic configurations are also known.[45,46] Keeping with the osmostress-protective function of ectoine, enhanced transcription of these genes is typically triggered in response to increases in the external osmolarity.[36,43,47] In addition to serving as an effective microbial osmostress protectant, ectoine also provides cytoprotection during growth at either low-[36,42,48] or high[25,49]-temperature extremes.

Due to the function-preserving attributes of compatible solutes in general[50–52] and ectoine and 5-hydroxyectoine in particular,[53,54] these types of molecules are frequently addressed as chemical chaperones.[55,56] Ectoine can ameliorate desiccation stress,[54,57] protect the functionality of proteins against various types of challenges,[53,58–60] stabilize macromolecular protein complexes[61] and lipid bilayers,[62] protect DNA from damage by ionizing radiation,[63] and provide hydroxyl radical scavenging activity.[64] The protective effects of ectoine for proteins, cell membranes, DNA, and macromolecular complexes and its anti-inflammatory characteristics[65] attract considerable biotechnological attention[11–13,66–68] and led to the development of an industrial-scale biotechnological production process delivering ectoine on the scale of tons per annum. The highly salt-tolerant bacterium *Halomonas elongata* is used for this purpose as a natural and engineered cell factory.[12,13,46]

AMINO ACID SEQUENCE INFORMATION

EctC-type proteins are closely related in their amino acid sequence and typically comprise around 130 amino acids.[1,11]

Table 1 Kinetic parameters of biochemically studied bacterial and archaeal ectoine synthases

EctC from	Length (AA)	MW (kDa)	pI	K_m (mM ectoine)	V_{max} (U mg^{-1})	Reference
P. lautus	130	14.7	4.7	7.8	9	1
S. alaskensis	137	15.2	5.2	5	4.6	4
H. elongata	137	15.5	4.9	11	85	8
M. alcaliphilum	134	15.3	5.0	28.7	64	10
N. maritimus	128	14.7	5.7	6.4	12.8	20

Based on a recent comprehensive phylogenomic analysis of 6688 fully sequenced bacterial and archaeal genome sequences,[11] 437 *bona fide* EctC-type proteins were retrieved via a BLAST search.[69] Representative examples are the EctC proteins from the following *Bacteria* and *Archaea*:

- *Acidiphilium cryptum* JF-5: (*Ac*)EctC; 132 amino acids; calculated pI: 6.0; NCBI accession number: WP_012040479.1
- *Alkalilimnicola ehrlichii* MLHE-1: (*Ae*)EctC; 132 amino acids; calculated pI: 5.0; NCBI accession number: WP_011628934.1
- *Halomonas elongata* DSM 2581: (*He*)EctC; 137 amino acids; calculated pI: 4.9; NCBI accession number: WP_013332346.1
- *Methylomicrobium alcaliphilum* 20Z: (*Ma*)EctC; 134 amino acids; calculated pI: 5.0; NCBI accession number: CCE24913.1
- *Nitrosopumilus maritimus* SCM1: (*Nm*)EctC; 128 amino acids; calculated pI: 5.7; NCBI accession number: WP_012215727.1
- *Paenibacillus lautus* Y412MC10: (*Pl*)EctC; 130 amino acids; calculated pI: 4.7; NCBI accession number: WP_009589551.1
- *Pseudomonas stutzeri* A1501: (*Ps*)EctC; 133 amino acids; calculated pI: 5.3; NCBI accession number: WP_011911425.1
- *Sphingopyxis alaskensis* RB2256: (*Sa*)EctC; 137 amino acids; calculated pI: 5.2; NCBI accession number: WP_011543220.1
- *Streptomyces coelicolor* A3(2): (*Sc*)EctC; 132 amino acids; calculated pI: 5.4; NCBI accession number: NP_626133.1
- *Vibrio cholerae* sv. O1 bv. El Tor N16961: (*Vc*)EctC; 138 amino acids; calculated pI: 5.5; NCBI accession number: NP_233209.1

Using the EctC protein from the thermotolerant Gram-positive bacterium *P. lautus* as the search query, the inspected 437 EctC-type proteins possess a degree of amino acid sequence identity ranging between 90% (for *Paenibacillus glucanolyticus*) and 49% (for *Streptomyces glaucescens*).[1,11] Amino acids involved in binding of the

Fe^{2+} cofactor, the EctA-formed substrate *N*-γ-acetyl-diaminobutyrate (*N*-γ-ADABA),[19] and ectoine are highly conserved in the extended EctC protein family.[1,11]

PROTEIN PRODUCTION AND PURIFICATION

EctC proteins from *H. elongata (He)*,[8,70] *A. cryptum (Ac)*,[71] *M. alcaliphilum (Ma)*,[10,17] *S. alaskensis (Sa)*,[4] *P. lautus (Pl)*,[1] and *N. matitimus (Nm)*[20] have been purified and studied biochemically, at least to some extent (Table 1). These microorganisms live in rather different ecosystems, representing high saline (*He*), acidic (*Ac*), haloalkaline (*Ma*), cold (*Sa*), and hot (*Pl*) environments. In addition, an EctC protein from a marine Thaumarchaeon (*Np*) was also studied. With the exception of the EctC protein from *H. elongata*,[8] and *M. alcaliphilum*[17] that were purified by classical biochemical techniques, the other mentioned ectoine synthases were recombinantly produced in *Escherichia coli* and purified via affinity chromatography using either His- or Strep-tagged constructs. In the following paragraphs, we describe in more detail the purification of the *S. alaskensis* [(*Sa*)EctC] and *P. lautus* [(*Pl*)EctC] EctC proteins, as these were used to obtain high-resolution crystal structures of the ectoine synthase in its apo, Fe^{2+}-complexed, substrate- and product-bound forms.[1,4]

In order to efficiently overproduce the (*Sa*)EctC and (*Pl*)EctC proteins heterologously in *E. coli* host strains, both *ectC* genes were synthesized in codon-optimized versions; they were cloned into the expression vector pASG-IBA3 (IBA GmbH, Göttingen, Germany). This vector harbors the coding sequence for a *Strep*-tag II affinity peptide (SA-WSHPQFEK), which was fused in the corresponding recombinant constructs to the 3′-end of the *ectC* coding sequence. The recombinant *ectC* genes were placed under the transcriptional control of the *tet*-promoter, whose activity is controlled by the anhydrotetracycline (AHT)-inducible TetR repressor. The *tetR* regulatory gene is present on the backbone of the pASG-IBA3 expression plasmid, thereby allowing effective control over the timing and degree of *ectC* transcription.[1,4]

The *E. coli* B strain BL21 harboring an *ectC* expression plasmid was cultivated at 37 °C in Minimal Medium

A (MMA)[72] supplemented with 0.5% glucose, 1 mg l^{-1} thiamine, 1 mM MgSO$_4$, and 0.5% casamino acids. Expression of an *ectC-Strep*-tag II hybrid gene was induced by the addition of the synthetic TetR inducer AHT (at a final concentration of 0.2 mg l^{-1}) when bacterial cultures reached an OD$_{578}$ of 0.7. The cultures were then further incubated for 2 h to allow EctC production before the cells were harvested by centrifugation. By passing them several times through a French pressure cell (at 1000 psi), the EctC-*Strep*-tag II-producing cells were disrupted, and a cleared cell lysate was prepared by ultracentrifugation (100 000 × g for 45 min) at 4 °C. The EctC-*Strep*-tag II proteins were purified from the supernatant by affinity chromatography on *Strep*-Tactin resin as detailed by Widderich *et al.*[4] and Czech *et al.*[1] for the (*Sa*)EctC and (*Pl*)EctC proteins, respectively. The concentration of the EctC proteins in the individual fractions eluted from the purification column was measured with the Pierce BCA Protein Assay Kit (Thermo Scientific, Waltham, MA, USA) or using a NanoDrop spectrophotometer (Peqlab, Erlangen, Germany). SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis; 15% polyacrylamide) was used to assess the purity of the affinity-purified EctC proteins. For crystallization trials,[73] the EctC proteins were concentrated to approximately 10 mg ml^{-1} using Vivaspin 6 columns with a 10-kDa molecular-weight cutoff.[1,4]

ACTIVITY ASSAY AND KINETIC CHARACTERIZATION

The limited availability of the substrate (*N*-γ-ADABA) for the ectoine synthase complicated full kinetic characterization of this enzyme in the past, as *N*-γ-ADABA had to be generated via alkaline hydrolysis of ectoine,[4,74] purified from bacteria carrying an *ectC* gene disruption mutation,[75] or chemically synthesized.[17] However, these constraints have now been relieved, since highly purified *N*-γ-ADABA became recently commercially available (Merck, Darmstadt, Germany).

Enzyme activities of the purified (*Sa*)EctC and (*Pl*)EctC proteins were determined by HPLC-based (high-performance liquid chromatography) enzyme assays that monitor ectoine formation from the substrate *N*-γ-ADABA with either a UV- or a diode array detector (DAD)-detector (set to 210 nm).[36] For the (*Sa*)EctC protein, the initial enzyme activity assays were performed in a 30-μl reaction volume for 20 min at 20 °C. The used standard buffer (20 mM Tris [tris(hydroxymethyl)aminomethane], pH 8.0) contained 150 mM NaCl, 1 mM FeCl$_2$, and 10 mM *N*-γ-ADABA. To determine optimal enzyme assay conditions for the (*Sa*)EctC protein, assay parameters and buffer conditions (e.g. the salt concentrations, temperature, and pH) were individually changed. The optimized assay buffer for ectoine synthase activity of the (*Sa*)EctC protein contained 20 mM Tris (pH 8.5), 200 mM NaCl, 1 mM FeCl$_2$, and 10 mM *N*-γ-ADABA. Activity assays were incubated for 20 min at 15 °C. Usually, 10 μg of the purified (*Sa*)EctC protein was added to start the enzymatic reaction. To assess the kinetic parameters of the (*Sa*)EctC ectoine synthase, varying concentrations of the substrates were used in the optimized assay buffer with a constant amount (10 μg) of the (*Sa*)EctC protein. The concentration of the substrate was varied between 0 and 40 mM. Enzyme reactions were stopped by adding 30 μl of acetonitrile (100%) to the assay solution. The samples were centrifuged (16 060 × g at room temperature for 5–10 min) to remove denatured proteins; the supernatant was subsequently analyzed for the formation of ectoine using HPLC. Usually, 5–10-μl samples were injected into the HPLC system, and ectoine was analytically separated on a GROM-SIL Amino-1PR column (125 × 4 mm with a particle size of 3 μm; purchased from GROM, Rottenburg-Hailfingen, Germany, or subsequently from Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Synthesis of ectoine by the purified (*Sa*)EctC protein was monitored using an Infinity 1260 DAD (Agilent, Waldbronn, Germany) integrated into an Agilent 1260 Infinity LC system (Agilent). The ectoine content of the samples was quantified via the OpenLAB software suite (Agilent) using commercially available ectoine (bitop AG, Witten, Germany) for the preparation of quantification standards.[1,4]

Enzyme activity assays using the (*Pl*)EctC protein followed basically the same protocol, but the assay conditions were slightly varied: 1 μg of purified (*Pl*)EctC protein was used in a 30-μl reaction volume containing 20 mM HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid], (pH 8.5) 50 mM NaCl, 0.1 mM (NH$_4$)$_2$Fe(SO$_4$), and 10 mM *N*-γ-ADABA. The enzyme reaction was run for 2.5 min at 30 °C in a water bath. When (*Pl*)EctC mutants were assayed for their enzymatic activity, the buffer conditions optimized for the wild-type enzyme were used as well, but 10 μg of purified (*Pl*)EctC protein was employed and the reaction time was extended to 30 min in order to detect even low enzyme activities of the (*Pl*)EctC variants.[1]

In addition to the (*Sa*)EctC and (*Pl*)EctC enzymes,[1,4] only a few other ectoine synthases have been analyzed with respect to their kinetic properties. These data are summarized in Table 1.

X-RAY STRUCTURES
Crystallization of the (*Sa*)EctC and (*Pl*)EctC proteins

Ectoine synthase has been crystallized from the cold-adapted Gram-negative bacterium *S. alaskensis* [(*Sa*)EctC][4] and the thermo-tolerant Gram-positive bacterium *P. lautus* [(*Pl*)EctC].[1] Crystallization trials with (*Sa*)EctC were performed using the sitting-drop vapor-diffusion and microbatch method at 12 °C; 0.1 μl homogeneous protein solution [10–12 mg ml^{-1} in 20 mM Tris (pH 8.0) and 200 mM NaCl] was mixed with 0.1 μl reservoir solution

and equilibrated against 50-µl reservoir solution. Two promising conditions were found: the first contained 0.05 M calcium acetate, 0.1 M sodium acetate, pH 4.5, and 40% (v v^{-1}) 1,2-propanediol, and the second one 20% (w v^{-1}) PEG 6000 (polyethylene glycol), 0.9 M lithium chloride, and 0.1 M citric acid (pH 5.0). Crystals were optimized by grid screening around the initial condition by mixing 1 µl protein and 1 µl reservoir solution and equilibrated against 300-µl reservoir solution. Large crystals were obtained with or without tert-butanol, which was used as an additive. The crystals reached their maximum size of $50 \times 50 \times 70$ µm after 3–10 weeks. Additionally, methylmercury(II) chloride (final concentration 0.5 mM) was added to some of the crystals to generate heavy atom-derivatized crystals.

The first condition resulted in a good diffracting (*Sa*)EctC crystal (1.2 Å; form A) exhibiting $P3_221$ symmetry with a unit cell of $a = 72.71$ Å, $b = 72.71$ Å, and $c = 52.33$ Å and α = 90°, β = 90°, and γ = 120°. Crystals obtained from the second condition (form B) diffracted to 2.0 Å and had a unit cell of $a = 97.52$ Å, $b = 43.96$ Å, and $c = 138.54$ Å and α = 90°, β = 101.5°, and γ = 120° and displayed a C2 symmetry. Attempts to solve the crystal structure by molecular replacement failed, but the mercury-derivatized crystal from the second condition diffracted to 2.8 Å and was used to solve the phases. This initial structure of (*Sa*)EctC was then used as a template for the analysis of the native datasets. One monomer was present in the asymmetric unit (ASU) of form A (it was referred to as the 'open' form of (*Sa*)EctC) and four monomers in form B (it was described as the 'semiclosed' form of (*Sa*)EctC).[4]

No (*Sa*)EctC crystals containing either the substrate *N*-γ-ADABA or the enzyme reaction product ectoine and/or the catalytically important iron could be obtained.[4] Therefore, cocrystallization trials with (*Pl*)EctC were conducted in the hope that the ectoine synthase from the thermotolerant bacterium *P. lautus*[76] would yield crystals suitable for structural analysis containing either the substrate or the product.[1] Homogeneous (*Pl*)EctC protein [8–12 mg ml^{-1} in 20 mM Tris (pH 7.5) 200 mM NaCl] was premixed with just Fe(II)Cl$_2$ (added to a final concentration of 4 mM), ectoine (added to a final concentration of 40 mM), or *N*-γ-ADABA (added to a final concentration of 20 mM), or a combination of Fe(II) plus ectoine or Fe(II) plus *N*-γ-ADABA was used. The different samples were incubated on ice for 1 h. Crystallization trials were set up using the sitting-drop vapor-diffusion method by mixing 0.1 µl (*Pl*)EctC with 0.1 µl reservoir solution and equilibrated against 50 µl reservoir solution. First crystals were observed after 12 h, and they were subsequently optimized by grid screens around the initially observed condition by mixing 1 µl protein and 1 µl reservoir solution and equilibrated against 300-µl reservoir solution. Best diffracting crystals appeared in a condition consisting of 0.2 M ammonium sulfate, 0.1 M phosphate citrate (pH 4.2), 20% (v v^{-1}) PEG 300, and 10% (v v^{-1}) glycerol,

and they reached their maximum size of $120 \times 45 \times 30$ µm (with ectoine) and $250 \times 45 \times 35$ µm (with *N*-γ-ADABA). All crystals were cryoprotected by overlaying the drop with mineral oil before crystals were harvested and flash-frozen in liquid nitrogen.

The (*Pl*)EctC::Fe complex (PDB code: 5onm) diffracted to a 1.6-Å resolution and exhibited a $P3_121$ symmetry with a unit cell of $a = 71.13$ Å, $b = 71.13$ Å, and $c = 68.66$ Å and α = 90°, β = 90°, and γ = 120°. The corresponding dataset was phased using molecular replacement with the previously solved structure of (*Sa*)EctC[4] as a search model. The iron–substrate complexes (*Pl*)EctC::Fe::*N*-γ-ADABA (PDB code: 5onn) diffracted to 2.0 Å, while the (*Pl*)EctC::Fe::ectoine (PDB code: 5ono) complex diffracted to 2.5 Å. These latter complexes crystallized in $P3_121$ symmetry and similar unit cells [(*Pl*)EctC::Fe::*N*-γ-ADABA: $a = 70.91$ Å, $b = 70.91$ Å, and $c = 68.54$ Å and α = 90°, β = 90°, and γ = 120°; the (*Pl*)EctC::Fe::ectoine: $a = 71.41$ Å, $b = 71.41$ Å, and $c = 68.83$ Å and α = 90°, β = 90°, and γ = 120°]. One monomer was present in the ASU in all three crystal structures obtained for the (*Pl*)EctC protein.

Overall structure of ectoine synthase

The overall structure of the (*Pl*)EctC protein consists of 11 β-strands (βI-βXI) and 2 α-helices (α-a and α-b) (Figure 2(a)). The β-strands form two antiparallel β-sheet regions consisting of βII, βIII, βX, βV, βI, βVI, βIX, and βXI. These sets of antiparallel β-sheets are packed against each other, forming a cup-shaped β-sandwich (Figure 2(a)) with a topology characteristic for the evolutionarily conserved cupin-fold of proteins.[5,77–80] The amino acid sequences of the (*Sa*)EctC[4] and (*Pl*)EctC[1] proteins whose crystal structures were determined possess an amino acid sequence identity of 43%. An overall structural comparison of the five determined EctC structures [PDB codes of (*Sa*)EctC: 5bxx and 5by5; PDB codes of (*Pl*)EctC: 5onm, 5ono, and 5onn][1,4] revealed a high degree of structural similarity with root-mean-square deviation (r.m.s.d) in the range between 0.6 and 0.58 Å over 90 Cα atoms. These data indicate that the (*Sa*)EctC and the (*Pl*)EctC ectoine synthases display a highly similar three-dimensional structure despite the fact that these proteins were derived from microorganisms living in ecophysiologically rather different habitats, the effluent of a hot spring (*P. lautus*)[1,76] [for (*Pl*)EctC)] and permanently cold ocean waters (*S. alaskensis*)[4,81] [for (*Sa*)EctC)].

The structure of the psychrophilic (*Sa*)EctC protein has been solved in two different conformations, which were coined the 'open' and 'semiclosed' states.[4] In the latter state, only part of the carboxy-terminus of the (*Sa*)EctC protein is visible in the electron density map, and it folds into a small helix (α-b) that closes the active site of the enzyme.[4] The formation of the helix α-b induces a reorientation and shift
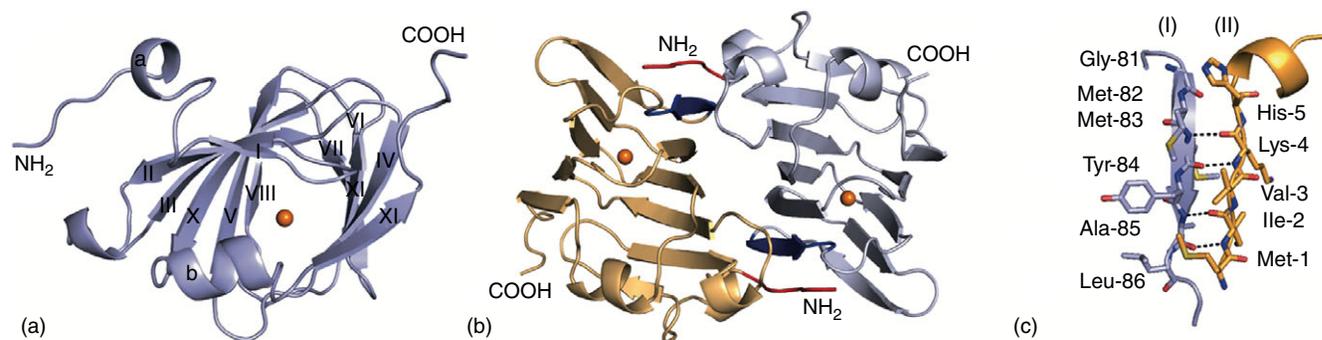
Figure 2 Crystal structure of the Fe^{2+}-containing (Pl)EctC protein and analysis of its dimer interface. The overall fold of the (Pl)EctC monomer in its iron-bound state (PDB code: 5onm) is shown in (a), and the head-to-tail dimer of the (Pl)EctC protein in (b) is presented in a side view. An orange sphere indicates the position of the iron ion in the (Pl)EctC protein. The dimer interface formed by β-strand VI of one monomer (shown in blue) and the N-terminal region (shown in red) of the other monomer is highlighted. (a, b) The (Pl)EctC protein in a cartoon representation. (c) The structural details of the dimer interface with hydrogen bonds formed between the backbone functional groups of Met1 to Lys4 (N-terminal segment of monomer II) and those of Met83 to Ala85 (β-strand VI from monomer I). The figures were prepared using PyMOL (www.pymol.org).[2] Adapted from Hanekop et al.[3]

of a long unstructured loop connecting the beta-sheets βIV and βVI of the (Sa)EctC protein, resulting in the formation of the stable β-strand βV. In contrast to the (Sa)EctC crystal structure, the COOH-terminus of the thermotolerant *P. lautus* EctC protein was completely resolved in the electron density map (Figure 2(a)). The remaining segment of the previously unresolved part of the carboxy-terminus of the (Sa)EctC protein flanks the cupin fold of the (Pl)EctC protein and protrudes out of the main body of the protein (Figure 2(a)).

A DALI search that identifies and assesses the structurally closely related proteins to the query crystal structure[82] was used to search for (Pl)EctC-related proteins. The 10 top hits with Z-scores[82,83] ranging between 28.1 and 13.0 are listed in Table 2. Each of these proteins is a member of the cupin superfamily,[5,6,77,78,80] and most of them contain divalent ions. As expected, the crystal structures of the *S. alaskensis* and *P. lautus* ectoine synthases are found with the highest Z-scores in this dataset. In addition to the ectoine synthases, three functionally characterized proteins are represented. These are the KdgF enzyme and two DddK-type proteins (Table 2). KdgF from *Halomonas* sp. is an enzyme forming 2-keto-3-deoxygluconate (Kdg) catalyzing the degradation of pectin and alginate-derived 4,5-unsaturated monouronates to linear ketonized forms.[84] The DddK protein from *Pelagibacter ubique*[85,86] is a dimethylsulfoniopropionate (DMSP) lyase, an enzyme involved in the catabolism of this major marine compatible solute[87,88]; DMSP can also function as an effective microbial osmostress protectant.[89]

Oligomeric state: EctC is a dimeric protein

To determine the oligomeric states of both the (Sa)EctC and (Pl)EctC proteins, size-exclusion chromatography (SEC)

analysis and high-performance liquid chromatography coupled to multiangle light-scattering (HPLC-MALS) experiments were conducted. The latter method allows the calculation of the size of the particle (in this case the EctC proteins) after running through the sizing column and yields the absolute molecular mass of the studied protein independent of the shape and chromatographic behavior of reference proteins.[90] The normalized elution profiles from the UV, refractive index, and light scattering detectors revealed a homogeneous and monodisperse protein solution with a molecular mass of 33.0 ± 2.3 kDa for (Sa)EctC and 32.2 ± 0.2 kDa for (Pl)EctC. These values correspond very well with the theoretically calculated molecular masses of EctC dimers [molecular mass of the monomer including the *Strep*-tag: 16.3 kDa for the (Sa)EctC protein and 15.9 kDa for the (Pl)EctC protein]. Hence, both ectoine synthases described here in greater detail are dimers in solution.[1,4] Dimer formation of the ectoine synthases from *N. maritimus*[4] and *M. alacaliphilum*[10] has also been reported, and the corresponding protein from *H. elongata* forms in all likelihood also dimers in solution.[8] Taken together, these data indicate that dimer formation is a common characteristic of ectoine synthases.

The crystal structures of the (Sa)EctC and (Pl)EctC proteins exhibited a different crystal packing due to different symmetries of the crystals and therefore a diverse composition of the ASU. Nevertheless, the inspection of the crystal packing and analysis of the respective monomer–monomer interactions revealed the functional dimer within the crystal structure in all solved EctC structures. Two monomers are arranged in a head-to-tail orientation, which is mainly stabilized through strong interactions between the N-terminal β-strand βI from monomer A and β-strand βVIII from monomer B. The interactions between these two β-strands rely primarily on backbone contacts (Figure 2(b)

Table 2   Structures of known EctC enzymes and other cupin proteins as deduced from a DALI search

Enzyme	Function	PDB code	Z-score	Ligand(s)	Metal-binding site	Binding geometry
(*Pl*)EctC	Ectoine synthesis	5onm	28.8	Fe^{2+}	Glu75, Tyr84, and His92	Tetrahedral
(*Pl*EctC	Ectoine synthesis	5onn	28.1	Fe^{2+}/ADABA	Glu75, Tyr84, and His92	Tetrahedral
(*Pl*)EctC	Ectoine synthesis	5ono	27.5	Fe^{2+}/ectoine	Glu75, Tyr84, and His92	Tetrahedral
(*Sa*)EctC	Ectoine synthesis	5bxx	21.1	–	–	–
(*Sa*)EctC	Ectoine synthesis	5by5	18.5	S-1,2-propanediol	–	–
(*Ha*)KdgF	Uronate metabolism	5fq0	13.4	Ni^{2+}/citrate anion	His48, His50, Gln55, and His89	Octahedral
(*Sw*)Cupin	–	2pfw	13.2	1,2-Ethanediol	–	–
(*Pu*)DddK$_{mut}$	DMSP metabolism	6a55	13.2	Mn^{2+}/3-(dimethyl-lambda-sulfanyl) propanoic acid	His56, His58, Glu62, and His96	Octahedral
(*St*)He2323	–	4e2g	13.1	Ni^{2+}/acetate	His54, His56, Gln60, and His94	Tetrahedral/Octahedral
(*Pu*)DddK	DMSP metabolism	5tfz	13.0	Ni^{2+}/di(hydroxyl)ether/3-(acryloyloxy) propanoic acid	His56, His58, Glu62, and His96	Trigonal pyramidal

and (c)).[1] Additionally, some weaker hydrophobic interactions between the two monomers are also observed in some loop regions connecting the β-strands (Figure 2(b)). Since the (*Sa*)EctC and (*Pl*)EctC proteins are both head-to-tail dimers,[1,4] the interactions between the monomers occur twice in the dimer assembly (Figure 2(b)). As determined by PISA (Proteins, Interfaces, Structures, and Assemblies) analysis,[91] the (*Pl*)EctC monomers interact through a surface area of $1501\,\text{Å}^2$ in their dimer assembly.[1] The corresponding value for the (*Sa*)EctC protein is $1462\,\text{Å}^2$, which is roughly 20% of the total accessible surface of one monomer.[4]

## Metal content and architecture of the iron-binding site in EctC

Considerations based on bioinformatics of the amino acid sequences of EctC-type proteins suggested that the ectoine synthase belongs to the cupin superfamily.[5,6,77–80] Most of these proteins contain catalytically important transition state metals such as iron, copper, zinc, manganese, cobalt, or nickel, thereby allowing them to impose different types of chemistry onto an evolutionarily conserved protein fold. Cupins form barrel-like structures in which the metal-binding site and the catalytically important residues protrude into the lumen of the barrel.[5,6,77–80] As revealed by the crystal structures of (*Sa*)EctC[4] and (*Pl*)EctC,[1] the overall fold of ectoine synthase closely confers to the characteristic fold of cupins (Figure 2(a)).

Cupins contain two conserved motifs: $G(X)_5HXH(X)_{3,4}$ $E(X)_6G$ and $G(X)_5PXG(X)_2H(X)_3N$ (the letters in bold represent those residues that typically coordinate the metal).

Inspection of an alignment of the amino acid sequences of 437 EctC-type proteins[1] revealed that the cupin signature motifs are somewhat varied in ectoine synthase in comparison with the canonical amino acid motif found in other members of the cupin superfamily. The amino acid sequences of the two cupin motifs in EctC are as follows: motif-I is $G(X)_5WY(X)_4E(X)_6G$, while motif-II is $G(X)_6PG(X)_2Y(X)_3G(X)_3H$ (letters in bold indicate metal-binding residues). These variations in the novel cupin motif of EctC do not affect those residues that coordinate the catalytically critical metal ion.[1] The position of the two cupin motifs within the EctC amino acid chain is shown in Figure 3(a) for an abbreviated alignment of 10 EctC proteins.

To investigate the presence and nature of the metal that might be present in (*Sa*)EctC, inductively coupled plasma mass spectrometry (ICP-MS) experiments were performed. For this analysis, recombinant (*Sa*)EctC preparations from three independent protein overproduction and purification experiments were employed. The ICP-MS analyses yielded an iron content of 0.66 ± 0.06 mol iron per mol of (*Sa*)EctC protein with a minor amount of zinc (0.08 mol zinc per mol of protein). All other assayed metals (copper and nickel) were only present in trace amounts (0.01 mol metal per mol of protein). The presence of iron in these (*Sa*)EctC protein preparations was further confirmed by a colorimetric method that is based on an iron-complexing reagent;[92] this procedure yielded an iron content of 0.84 ± 0.05 mol per mol of (*Sa*)EctC protein. Hence, both ICP-MS and the colorimetric method clearly established that the recombinantly produced ectoine synthase from *S. alaskensis* is an iron-containing protein.[4] Surprisingly, no iron ion was found in the solved (*Sa*)EctC crystal structures.[4]
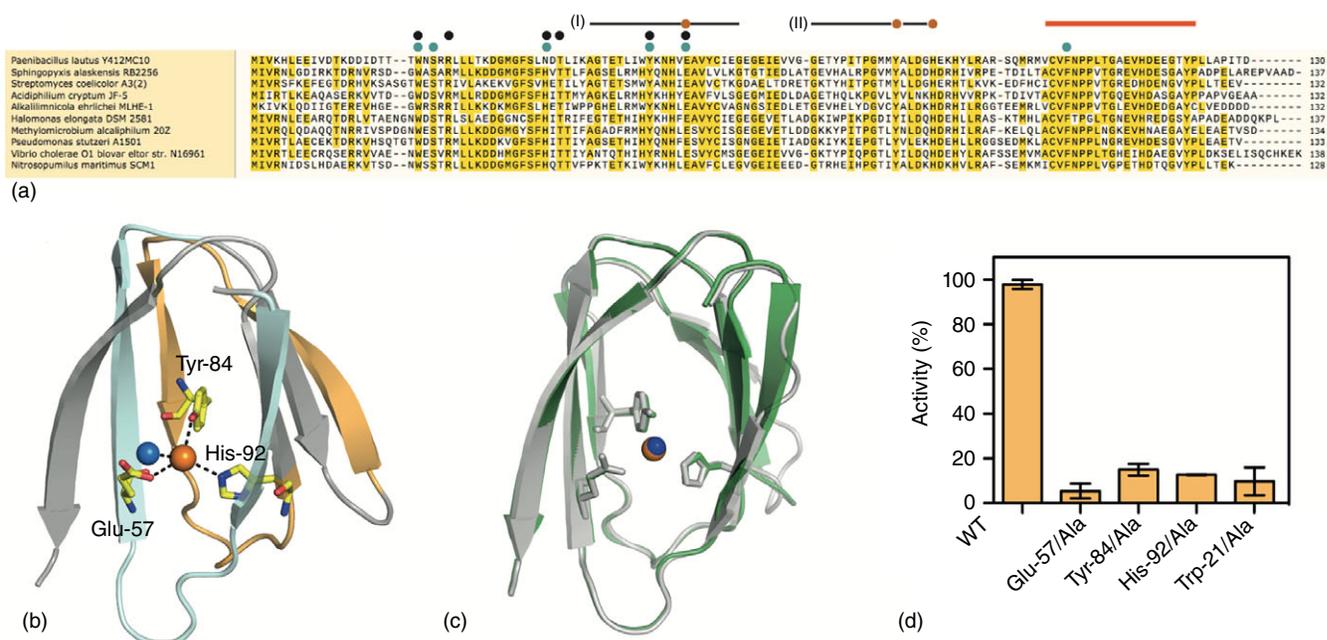
Figure 3 Abbreviated alignment of EctC-type proteins and structural insights into the iron-binding site. (a) Alignment of the amino acid sequences of 10 EctC proteins selected from a dataset of 437 *bona fide* ectoine synthases.[1,11] In selecting the aligned EctC proteins, special emphasis was given to those ectoine synthases that have been functionally characterized at least to some extent[1,4,8,10,17,20,70,71] or are mentioned in the text. Conserved amino acid residues are shaded in yellow; dots shown above the (*Pl*)EctC protein sequence indicate residues involved in the binding of the iron ligand (orange), the N-γ-ADABA substrate (black), or the enzyme reaction product ectoine (blue). The positions of cupin motif-I and cupin motif-II within the EctC amino acid sequence are indicated by a black line. The orange bar highlights the region in the C-terminal segment of EctC that forms a lid over the entry to the cupin barrel. (b, c) A structural view into the Fe^{2+}-binding site. For these figures, the (*Pl*)EctC::Fe crystal structure (PDB code: 5onm) was used (in b) and was overlaid in (c) with the iron-free apoform of the (*Sa*)EctC crystal structure (PDB code: 5bxx). A number of secondary structure elements of the (*Pl*)EctC protein were removed *in silico* in order to highlight the architecture of the iron-binding site and the position of the two cupin motifs. The two cupin motifs are represented in blue (motif-I) and orange (motif-II). The iron ion (represented by an orange sphere) is coordinated by the side chains of Glu57, Tyr84, and His92 of the (*Pl*)EctC protein. The distance between the iron atom and the side chains of these three residues are 2.9, 2.8, and 2.9 Å, respectively. The iron-binding site in the substrate-free (*Pl*)EctC crystal structure contains a localized water molecule (blue sphere). (c) Overlay of the iron-binding site in the (*Pl*)EctC (shown in green) and (*Sa*)EctC (shown in gray). The side chains of the three residues involved in the binding of the Fe^{2+} ion are depicted as sticks. A water molecule (blue sphere) in the (*Sa*)EctC crystal structure occupies the same location as the Fe^{2+} ligand (orange sphere) in the (*Pl*)EctC structure. (d) Single amino acid substitution of the residues involved in iron binding and enzyme catalysis by the (*Pl*)EctC protein were exchanged to Ala residues. These mutants were tested for their ectoine synthase enzyme activity. The enzyme activity of the mutant (*Pl*)EctC proteins is represented relative to that of the wild-type enzyme (set to 100% activity). These data were taken from Czech *et al*. and replotted.[1] Crystal structures of ectoine synthases shown in this figure were prepared using PyMOL (www.pymol.org).[2] Adapted from Czech *et al*.[1] and Hanekop *et al*.[3]

In contrast to (*Sa*)EctC, in all three crystal structures of (*Pl*)EctC, the iron catalyst was clearly visible. Within the (*Pl*)EctC::Fe crystal structure, the iron atom is tetrahedrally coordinated via interactions with the side chains of Glu57, Tyr84, and His92 (Figure 3(b)). The distance between these iron-coordinating side chains and the ion are 2.9, 2.8, and 2.9 Å, respectively. A water molecule completes the tetrahedral arrangement of the (*Pl*)EctC iron-binding site in the substrate-free (*Pl*)EctC::Fe crystal structure; the water molecule has a distance of 2.9 Å to the iron ion (Figure 3(b)).[1]

While in the (*Sa*)EctC crystal structure,[4] no metal atom was visible, a water molecule occupied in (*Sa*)EctC the same position that was observed for the iron ion in the

(*Pl*)EctC::Fe crystal structure (Figure 3(c)). An overlay of the three iron-coordinating amino acid residues in the (*Pl*)EctC and (*Sa*)EctC structures[1,4] revealed that they are perfectly superimposable (Figure 3(c)). This observation indicates that (i) the iron-binding site in the ectoine synthase is already preformed in the absence of the catalytically important cofactor and that (ii) the binding of the iron atom does not seem to trigger substantial structural rearrangements in the overall fold of the EctC protein.

Structure-guided site-directed mutagenesis of both (*Sa*)EctC and (*Pl*)EctC was performed to further probe the iron content and enzyme activity of these proteins. Single amino acid substitutions of the iron-coordinating residues abrogated ectoine synthase enzyme activity either

entirely or to a large extent.[1,4] This is exemplarily shown in Figure 3(d) in which the enzyme activity of (*Pl*)EctC variants carrying Glu57/Ala, Tyr84/Ala, and His92/Ala single amino acid substitutions of the iron-coordinating residues was assessed.[1]

In addition to site-directed mutagenesis experiments targeting the iron-coordinating residues in (*Sa*)EctC and (*Pl*)EctC, iron depletion and metal reconstitution experiments were carried out with the (*Sa*)EctC protein.[4] Both types of experiments showed that the ectoine synthase is critically dependent on the iron cofactor.[1,4] As observed with other metal-containing cupins,[6,77,78,80] the iron ion can be substituted, at least partially, in the (*Sa*)EctC protein by other divalent metals (e.g. Zn^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, and Mn^{2+}) when they are provided at substantial concentrations (10 µM protein and 1 mM metal). Relative to the iron-loaded (*Sa*)EctC protein, these alternative metals allowed a recovery of enzyme activity between 10% and 20%.[4] However, when divalent metals other than Fe^{2+} were provided at low concentration (10 µM protein and 10 µM metal), recovery of enzyme activity was in essence negligable.[4]

Architecture of the *N*-γ-ADABA substrate-binding site

In the (*Pl*)EctC:Fe::*N*-γ-ADABA crystal structure (PDB code: 5onn),[1] the *N*-γ-ADABA substrate for the ectoine synthase is positioned in close proximity to the catalytically important iron atom within the cupin barrel (Figure 4(a)). Furthermore, the iron atom is bound in a fashion similar to that observed in the (*Pl*)EctC:Fe complex (Figure 3(b)). The substrate *N*-γ-ADABA was added in large excess (40 mM) to the crystallization solution of the (*Pl*)EctC protein. However, the obtained crystal structure displayed only a partially bound substrate molecule, with occupancy of approximately 68%. In the active site of the (*Pl*)EctC enzyme, *N*-γ-ADABA is coordinated within through six interactions involving residues Trp21, Arg25, Asn38, Thr40, Tyr52, and Glu57. Interactions of *N*-γ-ADABA with the iron ion further stabilize the substrate within the catalytic core of the ectoine synthase (Figure 4(a)).

Upon closer inspection, one finds that the *N*-γ-ADABA molecule is coordinated in the active site of (*Pl*)EctC via two sets of interactions: (i) direct interactions occurs between the O atom of the acetyl group of *N*-γ-ADABA (acetamide oxygen) and the iron cofactor. The N5 atom of the substrate interacts with Glu57 and with Tyr52; its α-NH_2 moiety interacts with Thr40, and one of the carboxylate O atoms interacts with Asn38. Both of the latter amino acid residues are part of β-sheet βIV. The carboxylate oxygens of *N*-γ-ADABA are also coordinated via interactions with a water molecule, which in turn is held in place via an interaction with the side chain of Arg25. (ii) A second set of interactions is observed for the C3 and C4 atoms of *N*-γ-ADABA, which interact with the side chain of Trp21. Notably, Trp21

adopts a dual conformation in the crystal structure of the (*Pl*)EctC::Fe::*N*-γ-ADABA complex, in line with the observed partial occupancy of the crystals with the bound substrate. Only in one of these two conformations of Trp21 its side chain is oriented toward the *N*-γ-ADABA substrate (52% occupancy). Comparison of the (*Pl*)EctC::Fe and the (*Pl*)EctC:Fe::*N*-γ-ADABA structures therefore suggests that the presence of the *N*-γ-ADABA molecule induces a rotamer conformational change of the side chain of Trp21 to provide additional stabilizing contacts to the *N*-γ-ADABA molecule. The notion that the side chain of Trp21 is critically involved in the stable positioning of the substrate in the active site is supported by data from a site-directed mutagenesis experiment in which Trp21 of (*Pl*)EctC was replaced with an Ala residue. This single amino acid substitution yielded a (*Pl*)EctC variant with only 9.7% remaining enzyme activity (Figure 3(d)).[1]

Architecture of the ectoine-binding site

In addition to the (*Pl*)EctC::Fe and (*Pl*)EctC::Fe::*N*-γ-ADABA crystal structures, a crystal structure was obtained that contained both the iron ion and ectoine. The ectoine molecule is coordinated within the active site through five direct interactions with the following (*Pl*)EctC residues: Ser23, Asn38, Tyr52, Glu57, and Phe106 (Figure 4(b)). Compared with the coordination of the substrate *N*-γ-ADABA, different interactions and partially different residues are involved in binding of the reaction product ectoine. In particular, Tyr52 and Glu57 are no longer involved in any H-bonding interactions with N1 of ectoine (derived from the amidic N5 of *N*-γ-ADABA), but now both residues form H-bonds to N3 of ectoine, which is derived from the α-amino group of *N*-γ-ADABA. These new H-bonds toward N3 of ectoine also appear to be stronger than those formed originally with *N*-γ-ADABA, as suggested from their shorter distances, and are further stabilized by an interaction with Phe106 (Figure 4(b)). In addition, the side chain of Asn38 makes a direct contact to the carboxylate of ectoine, and this side chain is in turn held in place through stabilizing interactions with the side chain of Ser23. Direct interaction between the iron and the methyl group of the ectoine molecule might also be formed (Figure 4(b)). Notably, the side chain of Trp21 adopts a single conformation in the (*Pl*)EctC::Fe::ectoine structure and thereby provides additional stabilizing contacts to the ectoine ligand (Figure 4(b)).

MECHANISTIC ASPECTS

A superimposition of the (*Pl*)EctC::Fe::*N*-γ-ADABA and (*Pl*)EctC::Fe::*N*-γ-ectoine crystal structures revealed that the substrate and the reaction product occupy almost the same position within the active site of the (*Pl*)EctC
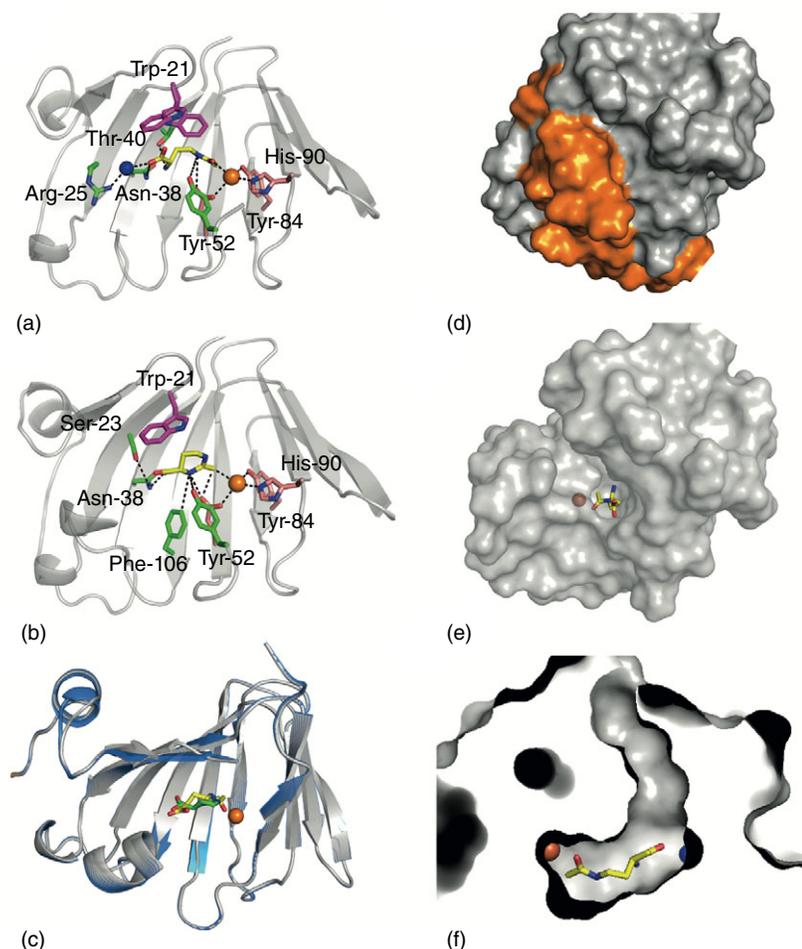
Figure 4 Crystallographic views into the catalytic core of the *P. lautus* ectoine synthase. (a) The iron- and substrate-binding network within the catalytic core of the (*Pl*)EctC protein. The side chain of Trp21 that adopts two different conformations in the (*Pl*)EctC::Fe::N-γ-ADABA crystal structure (PDB code: 5onn)[1] is emphasized in pink. The N-γ-ADABA molecule is shown as yellow sticks. The iron is represented as an orange sphere, and a water molecule (blue sphere) mediating indirect contacts between the N-γ-ADABA molecule and the side chain of Arg25 is highlighted. (b) Details of the iron- and ectoine-bound state of the (*Pl*)EctC catalytic core are shown (PDB code: 5ono).[1] In this crystal structure, the side chain of Trp21 adopts only a single conformation and provides stabilizing contacts to the ectoine ligand via cation-π interactions. The ectoine molecule is depicted in yellow sticks. (c) Cartoon representation of a structural overlay of the (*Pl*)EctC::Fe::N-γ-ADABA (gray) and (*Pl*)EctC::Fe::ectoine (blue) crystal structures.[1] The position of the substrate and reaction product within the lumen of the cupin barrel is highlighted, and the iron cofactor is shown as an orange sphere. In panels (a)–(c), parts of the secondary structure elements of the (*Pl*)EctC protein were removed in order to provide an unobstructed view into the catalytic core of the ectoine synthase. (d) Surface representation of the (*Pl*)EctC::Fe::N-γ-ADABA crystal structure[1] in which the lid region is highlighted in orange. (e) View into the catalytic core of the (*Pl*)EctC::Fe::N-γ-ADABA crystal structure[1] after the *in silico* removal of the lid region (amino acids 103–130). (f) Cross section through the catalytic core of the (*Pl*)EctC::Fe::N-γ-ADABA crystal structure[1] revealing a plausible entry tunnel for the N-γ-ADABA substrate (shown in yellow sticks) and its spatial position relative to the catalytically critical iron catalyst (orange sphere) and a water molecule (blue sphere) that interacts with the N-γ-ADABA molecule. Crystal structures of ectoine synthases shown in this figure were prepared using PyMOL (www.pymol.org).[2] Adapted from Czech *et al*.[1] and Hanekop *et al*.[3]

enzyme (Figure 4(c)).[1] This finding indicates that the ectoine synthase does not undergo substantial structural rearrangements during enzyme catalysis.

The carboxy-terminal segment of the ectoine synthase from the psychrophilic bacterium *S. alaskensis*[81] could only partially be resolved in the (*Sa*)EctC crystal structure.[4] In contrast, it is fully visible in the structure[1] derived from the thermotolerant bacterium *P. lautus*.[76] As assessed by

an alignment of 437 EctC-type proteins,[1] this segment of EctC contains a particularly high number of strictly conserved amino acid residues (Figure 3(a)).[1] When viewed in the framework of the overall fold of (*Pl*)EctC, the carboxy terminus of ectoine synthases seems to form a lid covering the catalytic core of the enzyme (Figure 4(d) and (e)). This reaction chamber is buried deep inside the (*Pl*)EctC protein and becomes visible when one removes *in silico*

the lid region from the (*Pl*)EctC::Fe::*N*-γ-ADABA crystal structure (Figure 4(e)). At the bottom of the enzyme reaction chamber one finds the *N*-γ-ADABA substrate juxtaposition to the iron catalyst (Figure 4(f)).[1] Keeping in mind that crystal structures only represent a snapshot of a 'frozen state' of otherwise dynamic enzymes, we note that the *N*-γ-ADABA substrate is trapped in the (*Pl*)EctC::Fe::*N*-γ-ADABA crystal structure in an elongated configuration instead of a prebent form that would facilitate the EctC-catalyzed water-elimination reaction (Figure 1). Overall, an inspection of the (*Pl*)EctC::Fe, (*Pl*)EctC::Fe::*N*-γ-ADABA, and (*Pl*)EctC::Fe::*N*-γ-ectoine crystal structures allowed a proposal for the enzyme reaction mechanism carried out by the ectoine synthase.[1]

As revealed by studies with the ectoine synthase from *H. elongata*, the EctC-catalyzed reaction is in essence irreversible.[70] However, the *H. elongata* EctC enzyme was able to hydrolyze synthetic ectoine derivatives with either reduced [4,5-dihydro-2-methylimidazole-4-carboxylate (DHMICA)] or expanded (homoectoine) ring sizes.[70] Furthermore, the (*He*)EctC protein can form in a side reaction the synthetic compatible solute 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (ADPC) by cyclic condensation of glutamine.[70]

CONCLUDING REMARKS

Orphan EctC-type proteins: enzymes in search of a function?

The genes encoding *bona fide* ectoine synthases typically colocalize with other ectoine biosynthetic genes.[7,11–13,45,46,71] However, during the genome-driven investigation of compatible solute biosynthesis routes in the plant pathogen *Pseudomonas syringae* pv. syringae, Kurz *et al*.[93] identified an orphan *ectC* gene. Accordingly, his bacterium did not produce ectoine; however, ectoine production was detected when 'surface-sterilized' leaves of its host plant *Syringa vulgaris* were provided to cultures of *P. syringae* pv. syringae. These findings were interpreted by Kurz *et al*.[93] as the plant providing the *N*-γ-ADABA substrate for the orphan EctC protein, implying that *P. syringae* pv. syringae possesses import systems for this compound. Indeed, as previously observed for *H. elongata*[75] and *Salmonella enterica* serovar typhimurium, osmotically inducible transport systems for compatible solutes seem to import *N*-γ-ADABA to provide osmostress protection.[94] However, this could not be demonstrated for the studied *P. syringae* pv. syringae bacterium. However, the heterologous expression of the orphan *ectC* gene in an *ectC* mutant for *H. elongata* that produces enhanced levels of *N*-γ-ADABA showed that the orphan EctC-type protein was at least partially functional.[93] Database searches[1,11,93] showed that such orphan *ectC*-type genes are present in a phylogenetically quite heterogeneous groups of bacteria. Notably, they are present in microorganisms that harbor either complete ectoine biosynthetic gene clusters or lack them.[11] The evolutionary history of these EctC-type proteins and their true enzymatic and physiological function(s) remains to be elucidated. When the orphan EctC-type proteins are viewed in the context of the amino acid sequences and structures of *bona fide* ectoine synthase, one finds that the iron-coordinating residues present in (*Pl*)EctC are strictly conserved (145 EctC-type proteins were inspected),[1,11] implying that these enzymes are in all likelihood metal dependent as well.

Ectoine synthase has an ecophysiological and metabolically important counterpart

In severely stressed microorganisms, compatible solutes (e.g. ectoine and 5-hydroxyectoine) are accumulated to exceedingly high intracellular concentrations via *de novo* synthesis.[34,35,95,96] These organic osmolytes are released into the environment upon severe osmotic down-shocks via the transient opening of mechanosensitive channels,[97] or when bacterial cells disintegrate subsequent to attack by phages or toxins.[98] This enables microbial cells to acquire ectoines and other types of compatible solutes via transport systems.[40,41] The released and reacquired compatible solutes thus provide new opportunities for microbial communities by allowing their use either as osmostress protectants or as nutrients.[98] Indeed, different types of ectoine transport systems are known,[48,99–103] and a considerable number of microorganisms can metabolize the nitrogen-rich ectoine molecule for use as sole energy, carbon, or nitrogen sources.[46,99,104,105]

A key step in the catabolism of ectoine is the enzymatic opening of the tetrahydropyrimidine ring. This is carried out by the ectoine hydrolase EutD/DeoA (EC 3.5.4.44),[46,104] an enzyme that belongs to the M24 peptidase family. Members of this enzyme family are often metal dependent,[106,107] but it is unknown if this is true for the ectoine hydrolase as well. The ectoine synthase (EctC) and the ectoine hydrolase (EutD/DeoA) not only belong to different protein superfamilies, but their amino acid sequences are also not related to each other.[1,46,104]

The enzyme reaction catalyzed by EutD yields primarily *N*-α-ADABA (and perhaps also some *N*-γ-ADABA),[46] which is then further catabolized in several enzymatic steps to L-aspartate.[46,104] While the transcription of the ectoine/hydroxyectoine biosynthetic genes is induced in response to high environmental osmolarity/salinity,[11–13] that of the catabolic genes is instead upregulated in response to the presence of ectoines in the growth medium.[46,99,104] However, ectoines are not the true inducers of enhanced transcription of the catabolic genes; instead, *N*-α-ADABA (but not the major EctC substrate *N*-γ-ADABA) serves as an internal inducer and acts through its interaction with the PLP-containing MocR/GabR-type repressor EnuR.[108,109]

RELATED ARTICLES

The ectoine hydroxylase: a nonheme-containing iron(II) and 2-oxoglutarate-dependent dioxygenase; Tracking Metal Ions in Biology Using X-Ray Methods; Iron Proteins with Mononuclear Active Sites; The Organomercurial Lyase MerB; 6-Pyruvoyl-Tetrahydropterin Synthase; Peptidyl-α-Hydroxyglycine α-Amidating Lyase (PAL); N-Formimino-L-Glutamate Iminohydrolase from *Pseudomonas Aeruginosa*

REFERENCES

1   L Czech, A Höppner, S Kobus, A Seubert, R Riclea, JS Dickschat, J Heider, SHJ Smits and E Bremer, *Sci Rep*, **9**, 364 (2019).

2   WL Delano, The PyMol Molecular Graphics System, Delano Scientific, San Carlos, CA (2002).

3   N Hanekop, M Höing, L Sohn-Bösser, M Jebbar, L Schmitt and E Bremer, *J Mol Biol*, **374**, 1237–1250 (2007).

4   N Widderich, S Kobus, A Höppner, R Ricela, A Seubert, JS Dickschat, J Heider, SHJ Smits and E Bremer, *PLoS One*, **11**, e0151285 (2016).

5   MY Galperin and EV Koonin, *J Biol Chem*, **287**, 21–28 (2012).

6   JA Hangasky, CY Taabazuing, MA Valliere and MJ Knapp, *Metallomics*, **5**, 287–301 (2013).

7   P Louis and EA Galinski, *Microbiology*, **143**, 1141–1149 (1997).

8   H Ono, K Sawada, N Khunajakr, T Tao, M Yamamoto, M Hiramoto, A Shinmyo, M Takano and Y Murooka, *J Bacteriol*, **181**, 91–99 (1999).

9   P Peters, EA Galinski and HG Trüper, *FEMS Microbiol Lett*, **71**, 157–162 (1990).

10 AS Reshetnikov, VN Khmelenina, II Mustakhimov and YA Trotsenko, *Methods Enzymol*, **495**, 15–30 (2011).

11 L Czech, L Hermann, N Stöveken, AA Richter, A Höppner, SHJ Smits, J Heider and E Bremer, *Genes (Basel)*, **9**, 177 (2018).

12 JM Pastor, M Salvador, M Argandona, V Bernal, M Reina-Bueno, LN Csonka, JL Iborra, C Vargas, JJ Nieto and M Canovas, *Biotechnol Adv*, **28**, 782–801 (2010).

13 HJ Kunte, G Lentzen and E Galinski, *Curr Biotechnol*, **3**, 10–25 (2014).

14 EA Galinski, HP Pfeiffer and HG Trüper, *Eur J Biochem*, **149**, 135–139 (1985).

15 CC Lo, CA Bonner, G Xie, M D'Souza and RA Jensen, *Microbiol Mol Biol Rev*, **73**, 594–651 (2009).

16 N Stöveken, M Pittelkow, T Sinner, RA Jensen, J Heider and E Bremer, *J Bacteriol*, **193**, 4456–4468 (2011).

17 AS Reshetnikov, VN Khmelenina and YA Trotsenko, *Arch Microbiol*, **184**, 286–297 (2006).

18 AA Richter, C-N Mais, L Czech, K Geyer, A Hoeppner, AHJ Smits, TJ Erb and E Bremer, *Front Microbiol*, **10**, 2811 (2019).

19 AA Richter, S Kobus, L Czech, A Hoeppner, J Zarzycki, TJ Erb, L Lauterbach, JS Dickschat, E Bremer and SHJ Smits, *J Biol Chem*, **295**, 2822–2838 (2020).

20 N Widderich, L Czech, FJ Elling, M Könneke, N Stöveken, M Pittelkow, R Riclea, JS Dickschat, J Heider and E Bremer, *Env Microbiol*, **18**, 1227–1248 (2016).

21 M Ren, X Feng, Y Huang, H Wang, Z Hu, S Clingenpeel, BK Swan, MM Fonseca, D Posada, R Stepanauskas, JT Hollibaugh, PG Foster, T Woyke and H Luo, *ISME J*, **13**, 2150–2161 (2019).

22 L Inbar and A Lapidot, *Eur J Biochem*, **162**, 621–633 (1987).

23 J Bursy, AJ Pierik, N Pica and E Bremer, *J Biol Chem*, **282**, 31147–31155 (2007).

24 J Prabhu, F Schauwecker, N Grammel, U Keller and M Bernhard, *Appl Environ Microbiol*, **70**, 3130–3132 (2004).

25 R Garcia-Estepa, M Argandona, M Reina-Bueno, N Capote, F Iglesias-Guerra, JJ Nieto and C Vargas, *J Bacteriol*, **188**, 3774–3784 (2006).

26 N Widderich, A Höppner, M Pittelkow, J Heider, SH Smits and E Bremer, *PLoS One*, **9**, e93809 (2014).

27 A Höppner, N Widderich, M Lenders, E Bremer and SHJ Smits, *J Biol Chem*, **289**, 29570–29583 (2014).

28 L Weinisch, I Kirchner, M Grimm, S Kuhner, AJ Pierik, R Rossello-Mora and S Filker, *Microb Ecol*, **77**, 317–331 (2019).

29 L Weinisch, S Kuhner, R Roth, M Grimm, T Roth, DJA Netz, AJ Pierik and S Filker, *PLoS Biol*, **16**, e2003892 (2018).

30 T Harding, MW Brown, AG Simpson and AJ Roger, *Genome Biol Evol*, **8**, 2241–2258 (2016).

31 T Harding, AJ Roger and AGB Simpson, *Front Microbiol*, **8**, 944 (2017).

32 L Czech and E Bremer, *PLoS Biol*, **16**, e2005163 (2018).

33 S Fenizia, K Thume, M Wirgenings and G Pohnert, *Mar Drugs*, **18**, 42 (2020).

34 EA Galinski and HG Trüper, *FEMS Microbiol Rev*, **15**, 95–108 (1994).

35 N Gunde-Cimerman, A Plemenitas and A Oren, *FEMS Microbiol Rev*, **42**, 353–375 (2018).

36 AU Kuhlmann and E Bremer, *Appl Environ Microbiol*, **68**, 772–783 (2002).

37 PH Yancey, *Sci Progr*, **87**, 1–24 (2004).

38 PH Yancey, *J Experi Biol*, **208**, 2819–2830 (2005).

39 MS da Costa, H Santos and EA Galinski, *Adv Biochem Eng Biotechnol*, **61**, 117–153 (1998).

40 E Bremer and R Krämer, *Annu Rev Microbiol*, **73**, 313–314 (2019).

41 JM Wood, *Annu Rev Microbiol*, **65**, 215–238 (2011).

42 AU Kuhlmann, J Bursy, S Gimpel, T Hoffmann and E Bremer, *Appl Environ Microbiol*, **74**, 4560–4563 (2008).

43 L Czech, S Poehl, P Hub, N Stoeveken and E Bremer, *Appl Environ Microbiol*, **84**, e01772–17 (2018).

44 J van den Berg, AJ Boersma and B Poolman, *Nat Rev Microbiol*, **15**, 309–318 (2017).

45 MJ Leon, T Hoffmann, C Sanchez-Porro, J Heider, A Ventosa and E Bremer, *Front Microbiol*, **9**, 108 (2018).

46 K Schwibbert, A Marin-Sanguino, I Bagyan, G Heidrich, G Lentzen, H Seitz, M Rampp, SC Schuster, HP Klenk, F Pfeiffer, D Oesterhelt and HJ Kunte, *Env Microbiol*, **13**, 1973–1994 (2011).

47 LM Stiller, EA Galinski and E Witt, *Genes (Basel)*, **9**, 184 (2018).

48 AU Kuhlmann, T Hoffmann, J Bursy, M Jebbar and E Bremer, *J Bacteriol*, **193**, 4699–4708 (2011).

49 J Bursy, AU Kuhlmann, M Pittelkow, H Hartmann, M Jebbar, AJ Pierik and E Bremer, *Appl Environ Microbiol*, **74**, 7286–7296 (2008).

50 Z Ignatova and LM Gierasch, *Proc Natl Acad Sci USA*, **103**, 13357–13361 (2006).

51 TO Street, DW Bolen and GD Rose, *Proc Natl Acad Sci USA*, **103**, 13997–14002 (2006).

52 SS Stadmiller, AH Gorensek-Benitez, AJ Guseman and GJ Pielak, *J Mol Biol*, **429**, 1155–1161 (2017).

53 K Lippert and EA Galinski, *Appl Microbo Biotechnol*, **37**, 61–65 (1992).

54 C Tanne, EA Golovina, FA Hoekstra, A Meffert and EA Galinski, *Front Microbiol*, **5**, 150 (2014).

55 MK Chattopadhyay, R Kern, MY Mistou, AM Dandekar, SL Uratsu and G Richarme, *J Bacteriol*, **186**, 8149–8152 (2004).

56 S Diamant, N Eliahu, D Rosenthal and P Goloubinoff, *J Biol Chem*, **276**, 39586–39591 (2001).

57 M Manzanera, S Vilchez and A Tunnacliffe, *FEMS Microbiol Lett*, **233**, 347–352 (2004).

58 S Barth, M Huhn, B Matthey, A Klimka, EA Galinski and A Engert, *Appl Environ Microbiol*, **66**, 1572–1579 (2000).

59 S Knapp, R Ladenstein and EA Galinski, *Extremophiles*, **3**, 191–198 (1999).

60 S Kolp, M Pietsch, EA Galinski and M Gutschow, *Biochim Biophys Acta*, **1764**, 1234–1242 (2006).

61 DV Yanykin, M Malferrari, S Rapino, G Venturoli, AY Semenov and MD Mamedov, *Photosynth Res*, **141**, 165–179 (2019).

62 RK Harishchandra, S Wulff, G Lentzen, T Neuhaus and HJ Galla, *Biophys Chem*, **150**, 37–46 (2010).

63 MA Schröter, S Meyer, MB Hahn, T Solomun, H Sturm and HJ Kunte, *Sci Rep*, **7**, 15272 (2017).

64 S Brands, P Schein, KF Castro-Ochoa and EA Galinski, *Arch Biochem Biophys*, **674**, 108097 (2019).

65 H Abdel-Aziz, W Wadie, O Scherner, T Efferth and MT Khayyal, *J Nat Prod*, **78**, 1309–1315 (2015).

66 R Graf, S Anzali, J Buenger, F Pfluecker and H Driller, *Clin Dermatol*, **26**, 326–333 (2008).

67 G Lentzen and T Schwarz, *Appl Microbiol Biotechnol*, **72**, 623–634 (2006).

68 M Kurz, *Saline Systems*, **4**, 6 (2008).

69 SF Altschul, W Gish, W Miller, EW Myers and DJ Lipman, *J Mol Biol*, **215**, 403–410 (1990).

70 EM Witt, NW Davies and EA Galinski, *Appl Microbiol Biotechnol*, **91**, 113–122 (2011).

71 KD Moritz, B Amendt, EMHJ Witt and EA Galinski, *Extremophiles*, **19**, 87–99 (2015).

72 JH Miller, Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1972).

73 S Kobus, N Widderich, A Hoeppner, E Bremer and SHJ Smits, *Acta Cryst*, **F71**, 1027–1032 (2015).

74 HJ Kunte, EA Galinski and HG Trüper, *J Microbiol Meth*, **17**, 129–136 (1993).

75 D Canovas, C Vargas, F Iglesias-Guerra, LN Csonka, D Rhodes, A Ventosa and JJ Nieto, *J Biol Chem*, **272**, 25794–25801 (1997).

76 DA Mead, S Lucas, A Copeland, A Lapidus, JF Cheng, DC Bruce, LA Goodwin, S Pitluck, O Chertkov, X Zhang, JC Detter, CS Han, R Tapia, M Land, LJ Hauser, YJ Chang, NC Kyrpides, NN Ivanova, G Ovchinnikova, T Woyke, C Brumm,

77 R Hochstein, T Schoenfeld and P Brumm, *Stand Genomic Sci*, **6**, 381–400 (2012).

77 G Agarwal, M Rajavel, B Gopal and N Srinivasan, *PLoS One*, **4**, e5736 (2009).

78 R Uberto and EW Moomaw, *PLoS One*, **8**, e74477 (2013).

79 JM Dunwell, A Culham, CE Carter, CR Sosa-Aguirre and PW Goodenough, *Trends Biochem Sci*, **26**, 740–746 (2001).

80 JM Dunwell, A Purvis and S Khuri, *Phytochemistry*, **65**, 7–17 (2004).

81 L Ting, TJ Williams, MJ Cowley, FM Lauro, M Guilhaus, MJ Raftery and R Cavicchioli, *Env Microbiol*, **12**, 2658–2676 (2010).

82 L Holm, *Protein Sci*, **29**, 130–140 (2020).

83 L Holm and LM Laakso, *Nucleic Acids Res*, **44**, W351–W355 (2016).

84 JK Hobbs, SM Lee, M Robb, F Hof, C Barr, KT Abe, JH Hehemann, R McLean, DW Abbott and AB Boraston, *Proc Natl Acad Sci USA*, **113**, 6188–6193 (2016).

85 NJ Schnicker, SM De Silva, JD Todd and M Dey, *Biochemistry*, **56**, 2873–2885 (2017).

86 M Peng, XL Chen, D Zhang, XJ Wang, N Wang, P Wang, JD Todd, YZ Zhang and CY Li, *Appl Environ Microbiol*, **85**, e03127–18 (2019).

87 HA Bullock, H Luo and WB Whitman, *Front Microbiol*, **8**, 637 (2017).

88 AW Johnston, RT Green and JD Todd, *Curr Opin Chem Biol*, **31**, 58–65 (2016).

89 S Broy, C Chen, T Hoffmann, NL Brock, G Nau-Wagner, M Jebbar, SH Smits, JS Dickschat and E Bremer, *Environ Microbiol*, **7**, 2362–2378 (2015).

90 E Sahin and CJ Roberts, *Methods Mol Biol*, **899**, 403–423 (2012).

91 E Krissinel and K Henrick, *J Mol Biol*, **372**, 774–797 (2007).

92 W Lovenberg, BB Buchanan and JC Rabinowitz, *J Biol Chem*, **238**, 3899–3913 (1963).

93 M Kurz, AY Burch, B Seip, SE Lindow and H Gross, *Appl Environ Microbiol*, **76**, 5452–5462 (2010).

94 R Garcia-Estepa, D Canovas, F Iglesias-Guerra, A Ventosa, LN Csonka, JJ Nieto and C Vargas, *Syst Appl Microbiol*, **29**, 626–633 (2006).

95 JM Wood, E Bremer, LN Csonka, R Krämer, B Poolman, T van der Heide and LT Smith, *Comp Biochem Physiol A Mol Integr Physiol*, **130**, 437–460 (2001).

96 B Kempf and E Bremer, *Arch Microbiol*, **170**, 319–330 (1998).

97 IR Booth, *Curr Opin Microbiol*, **18**, 16–22 (2014).

98 DT Welsh, *FEMS Microbiol Rev*, **24**, 263–290 (2000).

99 M Jebbar, L Sohn-Bösser, E Bremer, T Bernard and C Blanco, *J Bacteriol*, **187**, 1293–1304 (2005).

100 V Vermeulen and HJ Kunte, *Extremophiles*, **8**, 175–184 (2004).

101 K Grammann, A Volke and HJ Kunte, *J Bacteriol*, **184**, 3078–3085 (2002).

102 SI Kuhlmann, AC Terwisscha van Scheltinga, R Bienert, HJ Kunte and C Ziegler, *Biochemistry*, **47**, 9475–9485 (2008).

103 H Peter, B Weil, A Burkovski, R Krämer and S Morbach, *J Bacteriol*, **180**, 6005–6012 (1998).

104 A Schulz, N Stöveken, IM Binzen, T Hoffmann, J Heider and E Bremer, *Environ Microbiol*, **19**, 926–946 (2017).

105 C Vargas, M Jebbar, R Carrasco, C Blanco, MI Calderon, F Iglesias-Guerra and JJ Nieto, *J Appl Microbiol*, **100**, 98–107 (2006).

106 SL Roderick and BW Matthews, *Biochemistry*, **32**, 3907–3912 (1993).

107 JF Bazan, LH Weaver, SL Roderick, R Huber and BW Matthews, *Proc Natl Acad Sci USA*, **91**, 2473–2477 (1994).

108 A Schulz, L Hermann, S-A Freibert, T Bönig, T Hoffmann, R Riclea, JS Dickschat, J Heider and E Bremer, *Env Microbiol*, **19**, 4599–4619 (2017).

109 A Tramonti, C Nardella, ML di Salvo, S Pascarella and R Contestabile, *FEBS J*, **285**, 3925–3944 (2018).