environmental microbiology

Environmental Microbiology (2016) **18**(4), 1227–1248 doi:10.1111/1462-2920.13156

Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon *Nitrosopumilus maritimus*

Nils Widderich,[1†] Laura Czech,[1†] Felix J. Elling,[2] Martin Könneke,[2] Nadine Stöveken,[1,3] Marco Pittelkow,[1] Ramona Riclea,[4,5] Jeroen S. Dickschat,[4,5] Johann Heider[1,3] and Erhard Bremer[1,3*]

[1]*Laboratory for Molecular Microbiology, Department of Biology, Philipps-University, Karl-von-Frisch Str. 8, D-35043 Marburg, Germany.*

[2]*Organic Geochemistry Group, MARUM – Center for Marine Environmental Sciences, University of Bremen, PO Box 330 440, D-28334 Bremen, Germany.*

[3]*LOEWE-Center for Synthetic Microbiology, Philipps-University Marburg, Hans-Meerwein-Str. 6, D-35043 Marburg, Germany.*

[4]*Kekulé-Institut for Organic Chemistry and Biochemistry, Friedrich-Wilhelms-University Bonn, Gerhard-Domagk Str. 1, D-53121 Bonn, Germany.*

[5]*Institute of Organic Chemistry, TU Braunschweig, Hagenring 30, D-38106 Braunschweig, Germany.*

Summary

Ectoine and hydroxyectoine are compatible solutes widely synthesized by members of the *Bacteria* to cope with high osmolarity surroundings. Inspection of 557 archaeal genomes revealed that only 12 strains affiliated with the *Nitrosopumilus*, *Methanothrix* or *Methanobacterium* genera harbour ectoine/hydroxyectoine gene clusters. Phylogenetic considerations suggest that these *Archaea* have acquired these genes through horizontal gene transfer events. Using the Thaumarchaeon '*Candidatus* Nitrosopumilus maritimus' as an example, we demonstrate that the transcription of its *ectABCD* genes is osmotically induced and functional since it leads to the production of both ectoine and hydroxyectoine. The ectoine synthase and the ectoine hydroxylase were biochemically characterized, and their properties resemble those of their counterparts from *Bacteria*. Transcriptional analysis of osmotically stressed '*Ca*. N. maritimus' cells demonstrated that they possess an ectoine/hydroxyectoine gene cluster (*hyp-ectABCD-mscS*) different from those recognized previously since it contains a gene for an MscS-type mechanosensitive channel. Complementation experiments with an *Escherichia coli* mutant lacking all known mechanosensitive channel proteins demonstrated that the (*Nm*)MscS protein is functional. Hence, '*Ca*. N. maritimus' cells cope with high salinity not only through enhanced synthesis of osmostress-protective ectoines but they already prepare themselves simultaneously for an eventually occurring osmotic down-shock by enhancing the production of a safety-valve (*Nm*MscS).

Received 31 July, 2015; revised 19 November, 2015; accepted 27 November, 2015. *For correspondence. E-mail bremer@staff.uni-marburg.de; Tel: (+49) 6421 2821529; Fax: (+49) 6421 2828979. †These authors contributed equally.

Introduction

Microorganisms are masters of change. Their genetically encoded and physiologically mediated adaptive responses allow them to cope with nutrient limitations and a multitude of cellular and environmentally imposed stresses that otherwise would impair growth or threaten their viability (Storz and Hengge-Aronis, 2000). Increases in the external salinity or osmolarity are such types of stress and are encountered by essentially all free-living microorganisms (Csonka, 1989; Bremer and Krämer, 2000; Roesser and Müller, 2001). They trigger the rapid and passive outflow of water from the cell, thereby causing a drop in vital turgor and an increase in molecular crowding of the cytoplasm that in extreme cases causes growth arrest or even death (Record *et al.*, 1998; Wood, 2011). In the course of evolution, microorganisms have developed two different, but not necessarily mutually exclusive (Deole *et al.*, 2013; Oren, 2013; Becker *et al.*, 2014; Youssef *et al.*, 2014), mechanisms to counteract high osmolarity-instigated water efflux, the so-called *salt in* and the *salt out* strategies (Galinski and Trüper, 1994; Kempf and Bremer, 1998). Both are based on an active raising by the bacterial cell of the osmotic potential of its cytoplasm, thereby indirectly creating an osmotic driving

force to promote water retention and re-entry (Csonka, 1989; Record *et al*., 1998; Bremer and Krämer, 2000; Roesser and Müller, 2001; Wood, 2011).

Microorganisms using the *salt in* strategy accumulate molar concentrations of K^+ and Cl^- ions through transport processes to balance the osmotic gradient across their cytoplasmic membrane (Galinski and Trüper, 1994; Oren, 2013). This is considered as an energetically favourable strategy for microorganisms to adapt to persistent high-salinity environments (Oren, 2011). However, it comes with a penalty since the entire proteome had to be adjusted in the course of evolution to sustained high ionic strength in order to keep proteins soluble and functional (Coquelle *et al*., 2010; Deole *et al*., 2013; Talon *et al*., 2014). As a consequence, only a limited number of microbial species use the *salt in* strategy since it also often restricts the types of habitats that these microorganisms can populate (Galinski and Trüper, 1994; Oren, 2013).

Microorganisms that use the *salt out* strategy amass, either through synthesis or uptake, a restricted set of organic osmolytes, the compatible solutes (Csonka, 1989; da Costa *et al*., 1998; Kempf and Bremer, 1998; Roesser and Müller, 2001; Wood *et al*., 2001). This strategy affords a flexible physiological response to both sustained high-salinity surroundings and to osmotic changes in those ecosystems where the salinity fluctuates more often (Galinski and Trüper, 1994; Bremer and Krämer, 2000). Compatible solutes are operationally defined as small, highly water-soluble organic osmolytes that are fully compliant with cellular biochemistry and physiology (Brown, 1976; Bolen and Baskakov, 2001; Wood, 2011). Their benign nature allows their accumulation to exceedingly high intracellular levels without disturbing cellular physiology and biochemistry. The accumulation of compatible solutes occurs in a fashion that is correlated with the degree of the osmotic stress imposed onto the microbial cell (Kuhlmann and Bremer, 2002; Saum and Müller, 2008; Brill *et al*., 2011; Hoffmann *et al*., 2013). Compatible solutes used by members of the *Bacteria* are typically uncharged or zwitterionic (Csonka, 1989; da Costa *et al*., 1998; Kempf and Bremer, 1998; Klähn and Hagemann, 2011). In contrast, members of the *Archaea* often produce charged derivatives of these types of osmolytes and also synthesize types of compatible solutes that are normally not found in *Bacteria* (da Costa *et al*., 1998; Roesser and Müller, 2001; Roberts, 2004; Müller *et al*., 2005; Empadinhas and da Costa, 2011).

Here, we focus on the tetrahydropyrimidine ectoine and its derivative 5-hydroxyectoine (Galinski *et al*., 1985; Inbar and Lapidot, 1988) (Fig. 1A), widely used compatible solutes in the microbial world (Pastor *et al*., 2010; Widderich *et al*., 2014a). High-osmolarity environments trigger enhanced expression of the ectoine (*ectABC*) and hydroxyectoine (*ectD*) biosynthetic genes (Louis and Galinski, 1997; Prabhu *et al*., 2004; Garcia-Estepa *et al*., 2006; Bursy *et al*., 2007), thereby leading to high-level synthesis of these compounds under osmotically unfavourable conditions (Kuhlmann and Bremer, 2002; Calderon *et al*., 2004; Bursy *et al*., 2007; Salvador *et al*., 2015). Genetic disruption of the ectoine biosynthetic genes causes an osmotic-sensitive growth phenotype, underscoring the physiological role of ectoines as effective microbial osmostress protectants (Canovas *et al*., 1997; Kol *et al*., 2010).

Synthesis of ectoine progresses from the precursor L-aspartate-β-semialdehyde, a central intermediate in microbial amino acid metabolism and cell wall synthesis (Lo *et al*., 2009; Stöveken *et al*., 2011). It comprises the sequential activities of three enzymes: L-2,4-diaminobutyrate transaminase (EctB; EC 2.6.1.76), L-2,4-diaminobutyrate acetyltransferase (EctA; EC 2.3.1.178) and ectoine synthase (EctC; EC 4.2.1.108) (Louis and Galinski, 1997; Ono *et al*., 1999b) (Fig. 1A). A considerable number of ectoine producers also synthesize an ectoine derivative, 5-hydroxyectoine, through a stereospecific reaction that is catalysed by the ectoine hydroxylase (EctD) (EC 1.14.11) (Bursy *et al*., 2007; Höppner *et al*., 2014; Widderich *et al*., 2014a,b) (Fig. 1A). Hydroxyectoine exhibits stress-protective properties that are partially different from and sometimes exceed those of ectoine, in particular with respect to temperature and desiccation stress (Manzanera *et al*., 2002; Garcia-Estepa *et al*., 2006; Bursy *et al*., 2008; Tanne *et al*., 2014).

A recent survey of bacterial and archaeal genome sequences revealed that ectoine and hydroxyectoine biosynthetic genes occur widely in *Bacteria*, are rarely found in *Archaea*, and do not exist in *Eukarya* (Widderich *et al*., 2014a). In a few isolated incidents, the presence of ectoine/hydroxyectoine biosynthetic genes in *Archaea* has been noted, e.g. in the context of the annotation of the genome sequences of the Thaumarchaeon '*Candidatus* Nitrosopumilus maritimus' SCM1 (Walker *et al*., 2010) and of the methanogen *Methanobacterium formicicum* BRM9 (Kelly *et al*., 2014). However, it is unclear whether these genes are functional, whether the properties of the key enzymes for ectoine/hydroxyectoine biosynthesis, the ectoine synthase (EctC) and the ectoine hydroxylase (EctD) (Fig. 1A) are different from those of their bacterial counterparts, and whether the transcription of the archaeal *ect* genes is increased in response to osmotic stress as is observed in members of the *Bacteria* (Kuhlmann and Bremer, 2002; Calderon *et al*., 2004; Bursy *et al*., 2007; Kuhlmann *et al*., 2008; Salvador *et al*., 2015).

Here, we address these questions by interrogating the rapidly expanding number of archaeal genome sequences through bioinformatics to provide a comprehensive overview on the occurrence of ectoine/

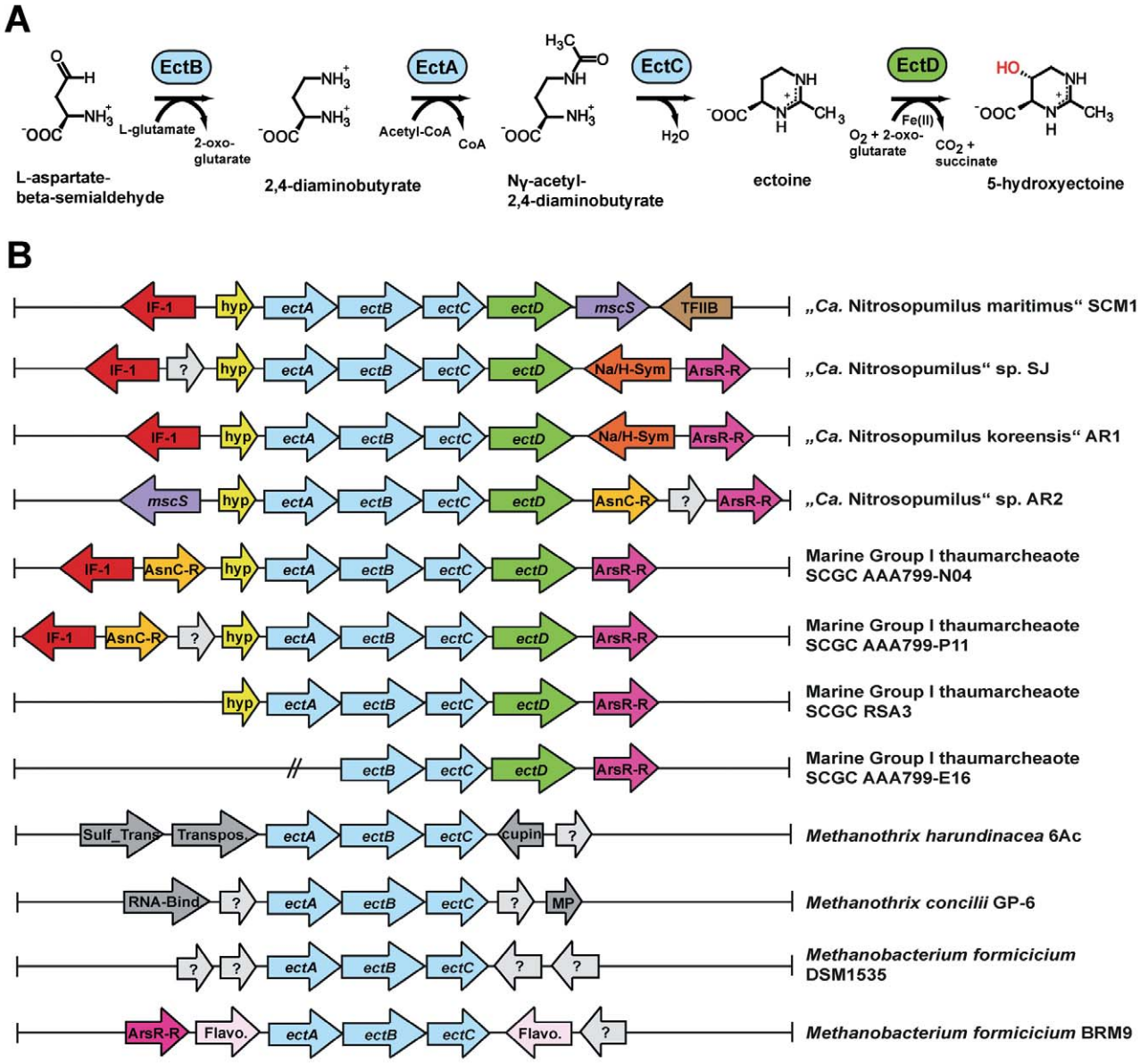

Fig. 1. (A) Biochemical steps required for the biosynthesis of ectoine and 5-hydroxyectoine and (B) genetic organization of *ectABC(D)* gene cluster and their flanking regions in archaeal genomes. In the marine group I thaumarchaeote SCGC AAA799-E16, no *ectA* gene was present; however, this genome sequence was assembled by a metagenomic approach making it likely that this thaumarchaeote also possesses a fully *ectABCD* gene cluster.

hydroxyectoine gene clusters in this domain of life. We used the marine Thaumarchaeon '*Ca.* Nitrosopumilus maritimus' SCM1 (Könneke *et al.*, 2005; Walker *et al.*, 2010), the first cultured representative of the globally abundant ammonia-oxidizing *Archaea* (AOA) (Stahl and de la Torre, 2012; Offre *et al.*, 2013; Bayer *et al.*, 2015; Gubry-Rangin *et al.*, 2015), as a model system to study ectoine/hydroxyectoine production in detail. We demonstrate for the first time in any archaeon that its *ectABCD* gene cluster is functionally expressed in response to high salinity, and that the biochemical properties of key enzymes (EctC, EctD) for the synthesis of ectoine and hydroxyectoine resemble those of their bacterial counterparts. Most interestingly, we found that the *ectABCD* genes of '*Ca.* N. maritimus' SCM1 are not only osmotically inducible but that they are also co-transcribed with a gene that encodes a functional mechanosensitive channel of the MscS family. These safety valves are ubiquitously used by both *Bacteria* and *Archaea* to withstand rapid osmotic downshifts (Kloda and Martinac, 2002; Booth and Blount, 2012; Wilson *et al.*, 2013; Booth, 2014). Hence, the osmotically regulated *ectABCD-mscS* transcriptional

unit is a sophisticated genetic device that allows high-salinity challenged '*Ca*. N. maritimus' SCM1 cells to sequentially cope with increases and decreases in the external osmolarity of their marine and estuarine habitat.

Results

Assessing the distribution of ectoine/hydroxyectoine biosynthetic genes in Bacteria and Archaea

To assess the occurrence and taxonomic distribution of ectoine biosynthetic genes, we used the database of the Joint Genome Institute (JGI) of the US Department of Energy (http://jgi.doe.gov/) (Nordberg *et al*., 2013) and the EctC protein from '*Ca*. Nitrosopumilus maritimus' SCM1 (Walker *et al*., 2010) as the query sequence. While the EctA (2,4-diaminobutyrate acetyltransferase) and EctB (L-2,4-diaminobutyrate transaminase) enzymes have counterparts in microbial biosynthetic pathways not related to ectoine biosynthesis, EctC can be regarded as a diagnostic enzyme for ectoine producers (Widderich *et al*., 2014a). Of note is that a restricted number of microorganisms exists that possesses *ectC*-type genes but lack the corresponding *ectAB* genes. The functional relevance of these orphan EctC-type proteins for ectoine biosynthesis is currently not fully understood (Kurz *et al*., 2010; Widderich *et al*., 2014a).

At the time of the BLAST search (08.07.2015), 27.232 completed and partially completed genome sequences of members of the *Bacteria* were represented in the JGI microbial database and 557 genome sequences of *Archaea* had been deposited. Among these 27.789 microbial genome sequences, 1297 hits to EctC-type proteins were found. After removing redundant entries of closely related strains (e.g. there are 181 genomes of strain of *Vibrio cholerae* represented that each possesses an *ect* gene cluster), the curated data set comprised 723 EctC-type proteins; 711 originate from *Bacteria* and only 12 were derived from *Archaea*. We note in this context that a considerable number of EctC-type proteins are misannotated in the database as L-mannose-6-phosphate isomerases (despite the typical localization of the corresponding genes within *ectABC(D)* gene clusters). Often, they are also referred to as RmlC-type cupins, an error that might originate from the bioinformatics assignment of the EctC protein to the RmlC subgroup of the cupin superfamily (Dunwell *et al*., 2001). The RmlC enzyme participates in the L-rhamnose biosynthetic pathway, serves as a carbohydrate epimerase (EC 5.1.3.13) (Dong *et al*., 2007), and therefore catalyses an enzymatic reaction quite different from that of the ectoine synthase EctC (EC 4.2.1.108) (Ono *et al*., 1999b) (Fig. 1A).

We then focused our further analysis on the genome sequences of those 12 *Archaea* that were identified by our database analysis to harbour *ectC*-type genes (Fig. 1B). In each of the corresponding genomes, the *ectC* gene is part of either an *ectABC* operon, or is embedded in an *ect* gene cluster that also comprises *ectD*. Eight members affiliated with the genus *Nitrosopumilus* or with Marine Group I.1a Thaumarchaeota (Pester *et al*., 2011) were represented among the 12 *ect*-containing *Archaea*, and in each of these gene clusters, a copy of the ectoine hydroxylase gene (*ectD*) was also found (Fig. 1B). Two species of the genus *Methanothrix*, *Methanothrix harundinacea* 6Ac and *Methanothrix concilii* GP6 were represented, and two strains of *Methanobacterium* were found as well. In the genome sequence of these four methanogens, a full *ectABC* gene cluster was present, but each of them lacked the ectoine hydroxylase *ectD* gene (Fig. 1B), and such a gene was also not present anywhere else in their genome. The ectoine hydroxylase (EctD) is an oxygen-dependent enzyme (Bursy *et al*., 2007; Höppner *et al*., 2014; Widderich *et al*., 2014b). Since members of the *Methanothrix* and *Methanobacterium* genera are all strict anaerobes, the absence of *ectD* in the genomes of these methanogens is readily understandable. On the other hand, the presence of *ectD* genes in '*Ca*. N. maritimus' SCM1, other members of the *Nitrosopumilus* genus and related Marine Group I Thaumarchaeota (Fig. 1B) is consistent with the lifestyle of these *Archaea* as oxygen-dependent nitrifying microorganisms (Könneke *et al*., 2005; Pester *et al*., 2011; Stieglmeier *et al*., 2014).

'Ca. N. maritimus' SCM1 synthesizes both ectoine and hydroxyectoine in response to osmotic stress

The presence of a particular gene cluster in a microbial genome does not necessarily imply that it is also functional. To study whether any of the detected archaeal *ect* gene clusters were functionally expressed, we chose the Thaumarchaeon '*Ca*. N. maritimus' SCM1 as a model system (Könneke *et al*., 2005; Walker *et al*., 2010). '*Candidatus* Nitrosopumilus maritimus' SCM1 was originally isolated from a tropical fish tank of an aquarium in Seattle (USA) and represents the first cultured ammonium oxidizer within the domain *Archaea* (Könneke *et al*., 2005). '*Ca.* N. maritimus' SCM1 belongs to the phylum Thaumarchaeota (Brochier-Armanet *et al*., 2008; Stieglmeier *et al*., 2014), which are among the most abundant microorganisms on Earth. They are ubiquitous in the Ocean (Stahl and de la Torre, 2012), found in estuarine sediments (Zhang *et al*., 2015) and are also ubiquitous in terrestrial habitats (Gubry-Rangin *et al*., 2015). Being adapted to extreme oligotrophic conditions (Martens-Habbena *et al*., 2009; Könneke *et al*., 2014), ammonia-oxidizing Thaumarchaeota are the predominant nitrifyers in the ocean and contribute significantly to the marine nitrogen cycle (Offre *et al*., 2013). The predicted EctABCD proteins of '*Ca*. N. maritimus' SCM1

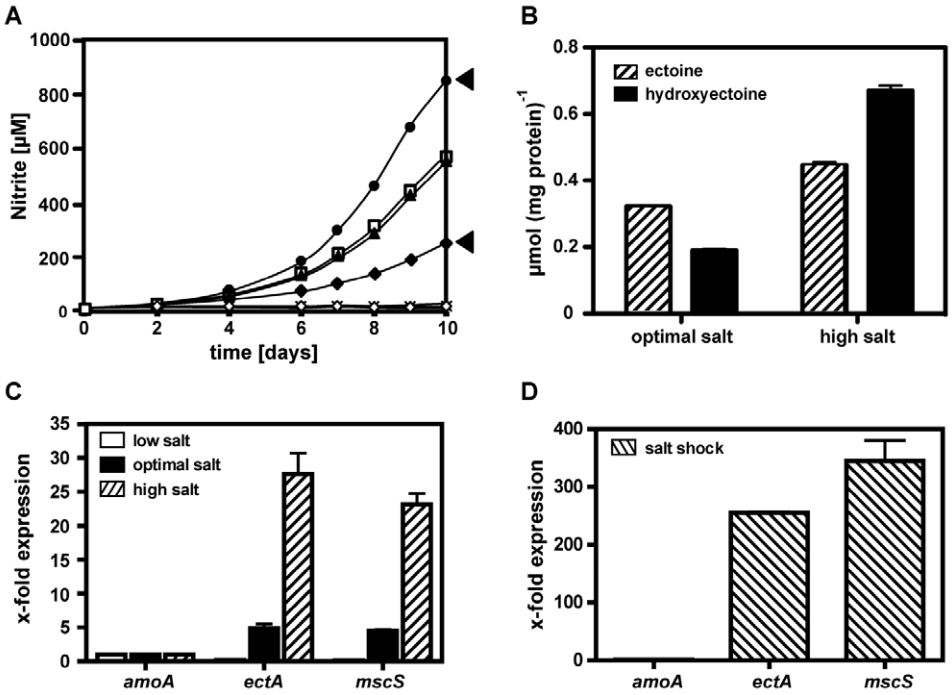

Fig. 2. Growth of '*Ca*. N. maritimus' SCM1 in the presence of different NaCl concentrations, its production of ectoines and the expression of its *ect* operon in response to different NaCl concentrations.
A. Growth of '*Ca*. N. maritimus' SCM1 cells in the presence of 0 M (closed diamonds), 0.15 M (cross), 0.32 M (open triangles), 0.48 M (closed triangles), 0.63 M (closed circles), 0.79 M (open squares), 0.94 M (closed diamonds) and 1.27 M (open diamonds) NaCl was monitored by nitrite production. Optimal sodium chloride concentration (0.63 M; closed circles) and high salt concentration (0.94 M; closed diamonds) in the growth medium of those cells used for the analysis of their ectoine and hydroxyectoine content are indicated by black arrowheads.
B. Intracellular ectoine and hydroxyectoine content per total amount of protein as assessed by HPLC measurements of '*Ca.* N. maritimus' SCM1 cells grown under optimal and high salt concentrations. Changes in expression of the '*Ca.* N. maritimus' SCM1 *ectA* and *mscS* genes under (C) low, optimal and high salt concentrations and (D) after an osmotic shock as assessed by qRT-PCR are depicted. Relative expression was determined using the *amoA* transcript as a reference.

exhibited a degree of amino acid sequence identity of 41%, 53%, 52% and 47% in comparison with the functionally studied ectoine and hydroxyectoine biosynthetic enzymes of *Halomonas elongata* (Ono *et al*., 1999b; Widderich *et al*., 2014a), the natural microbial cell factory for the industrial-scale production of ectoines (Schwibbert *et al*., 2011; Kunte *et al*., 2014).

'*Ca.* N. maritimus' SCM1 can grow over a wide range of salinities (Elling *et al*., 2015). To assess whether it would synthesize both ectoine and hydroxyectoine in response to salt stress, '*Ca*. N. maritimus' SCM1 was grown in a chemically defined medium containing 26 g l^{-1} NaCl (0.63 M NaCl); this medium has a measured osmolarity of 928 mOsm. To impose additional osmotic stress on '*Ca*. N. maritimus' SCM1 cells, we also grew them in the same medium but with increased sodium chloride (NaCl) concentration (48 g l^{-1}) (0.94 M NaCl); this medium had a measured osmolarity of 1.598 mOsm. The increased salinity of the medium had a significant negative impact on the growth of '*Ca*. N. maritimus' SCM1 (Fig. 2A). Under both growth conditions, ectoine and hydroxyectoine were detected by high-performance liquid chromatography (HPLC) analysis, demonstrating that the *ectABCD* gene cluster of '*Ca*. N. maritimus' SCM1 was functional. Since '*Ca*. N. maritimus' SCM 1 is a marine microorganism (Könneke *et al*., 2005), optimal growth conditions occur only at a substantial osmolarity (26 g NaCl l^{-1}) (0.63 M NaCl). Hence, it was not surprising that we already found ectoine and hydroxyectoine in these cells; their total ectoine and hydroxyectoine content were 0.32 µmol mg $protein^{-1}$ and 0.19 µmol mg $protein^{-1}$ respectively. These values increased to 0.44 µmol mg $protein^{-1}$ of ectoine and 0.67 µmol mg $protein^{-1}$ of hydroxyectoine when the cells were grown under heightened osmotic stress in a medium with a sodium chloride concentration to 48 g l^{-1} (0.94 M NaCl). Consequently, the level of synthesis of these compatible solutes by '*Ca*. N. maritimus' SCM1 is responsive to increased osmotic stress (Fig. 2B).

Biochemical properties of the ectoine synthase and ectoine hydroxylase from 'Ca. N. maritimus' SCM1

EctC and EctD are the key enzymes for ectoine and hydroxyectoine synthesis respectively (Fig. 1A) (Ono *et al*., 1999b; Bursy *et al*., 2007; Widderich *et al*., 2014a). We therefore wondered whether the biochemical properties of the EctC and EctD enzymes derived from *Archaea* possess characteristics similar to their bacterial counterparts, or are different. Consequently, we assessed the properties and kinetics of the EctC and EctD enzymes from the ectoine/hydroxyectoine producer '*Ca*. N. maritimus' SCM1. To provide the substantial amounts of proteins required for biochemical studies, we obtained synthetic, codon-optimized versions of the *ectC* and *ectD* genes from '*Ca*. N. maritimus' SCM1 (Walker *et al*., 2010) and inserted them into the pASG-IBA3 expression vector (IBA GmbH, Göttingen, Germany) such that the produced EctC and EctD proteins were attached at their C-termini to a *Strep*-tag II affinity peptide. We refer in the following to these recombinant proteins as (*Nm*)EctC and (*Nm*)EctD.

The (*Nm*)EctC and (*Nm*)EctD proteins were effectively overproduced in *E. coli* BL21 and could be purified to a high degree of homogeneity by affinity chromatography on a Streptactin matrix (Fig. 3). Size exclusion chromatography demonstrated that preparations of both enzymes were essentially free of protein aggregates (Fig. 3). The molecular masses of the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins calculated from their corresponding gene sequences are 15.9 kDa and 35.8 kDa respectively. The two recombinant proteins eluted from the size exclusion columns as species with a calculated molecular mass of 31.9 kDa and 71.4 kDa, respectively, indicating that both the (*Nm*)EctC and (*Nm*)EctD enzymes are dimers in solution (Fig. 3). Such a quaternary assembly has previously also been reported for ectoine synthases (Ono *et al*., 1999a; Kobus *et al*., 2015) and ectoine hydroxylases (Höppner *et al*., 2014; Widderich *et al*., 2014a) from members of the *Bacteria*.

We then determined a set of basic biochemical parameters for the (*Nm*)EctC and (*Nm*)EctD enzymes and subsequently used these data to set up optimal enzyme activity assays. The (*Nm*)EctC and (*Nm*)EctD enzymes had similar temperature profiles with optima at 30°C and 35°C respectively (Fig. 4A). Both also possessed a similar pH dependency with activity optima of 7.0 and 7.5, for the (*Nm*)EctC and (*Nm*)EctD proteins respectively (Fig. 4B). The activity of the ectoine hydroxylase from '*Ca*. N. maritimus' SCM1 was only slightly activated by increases in ionic strength (elicited either through the addition of NaCl or KCl to the enzyme assay buffer), whereas the activity of the ectoine synthase was restricted to a narrow concentration range of salts (Fig. 4C and D).

Using optimized enzyme assay conditions, we determined the kinetic parameters of the ectoine synthase and of the ectoine hydroxylase using their natural substrates N-γ-L-2,4-acetyl-diaminobutyrate (Ono *et al*., 1999b) and ectoine (Bursy *et al*., 2007; Widderich *et al*., 2014a) respectively. Both enzymes showed Michaelis–Menten-type kinetics. The apparent kinetic parameters for the ectoine synthase were: $K_m = 6.4 \pm 0.6$ mM; $v_{max} = 12.8 \pm 0.4$ U mg^{-1}, $k_{cat} = 5.7$ s^{-1}, $k_{cat}/K_m = 0.9$ s^{-1} mM^{-1} (Fig. 5A). Those for the ectoine hydroxylase were: K_m (ectoine) $= 3.8 \pm 0.5$ mM; K_m (2-oxoglutarate) $= 3.1 \pm 0.6$ mM; $v_{max} = 1.8 \pm 0.1$ U mg^{-1}, $k_{cat} = 1.0$ s^{-1}, $k_{cat}/K_m = 0.3$ s^{-1} mM^{-1} (Fig. 5B and C).

The basic biochemical features and kinetic parameters of the '*Ca*. N. maritimus' SCM1 ectoine hydroxylase are very similar to those of a considerable number of previously characterized EctD proteins (eight enzymes in total) from members of the *Bacteria* (Bursy *et al*., 2007; 2008; Widderich *et al*., 2014a). However, this situation is somewhat different for the ectoine synthase, an enzyme for which only two representatives have been kinetically characterized in some detail so far, namely the enzymes from *H. elongata* and *Acidiphilium cryptum* (Ono *et al*., 1999a; Moritz *et al*., 2015). The ectoine synthase from *H. elongata* is a highly salt tolerant enzyme (Ono *et al*., 1999a), whereas both the EctC proteins from *A. cryptum* JF-1 (Moritz *et al*., 2015) and *Ca.* N. maritimus' SCM1 (Fig. 4C and D) are sensitive to high salt concentrations. Moritz and colleagues (2015) suggested that these different sensitivities of the EctC enzymes from *H. elongata* and *A. cryptum* JF-1 against high salt concentrations result from their different overall net-negative charge (calculated isoelectric points: 4.87 and 6.03 for the EctC proteins from *H. elongata* and *A. cryptum* respectively). The calculated pI of the '*Ca*. N. maritimus' SCM1 EctC protein is 5.65, and hence closer to that of *A. cryptum* JF-1; its salt-sensitive properties (Fig. 4C and D) is thus consistent with the hypothesis put forward by Moritz and colleagues (2015).

The amino acid sequences of key enzymes for ectoine/hydroxyectoine biosynthesis in Archaea cluster with their counterparts from Bacteria

The amino acid sequences of the 723 EctC-type proteins retrieved from the above-described database analysis were aligned and they exhibited, relative to the (*Nm*)EctC protein, a degree of sequence identity ranging between 79% (EctC from the marine Groupe_I Thaumarchaeota SCGC_AAA799-E16) and 33% (EctC from *Micrococcus luteus*). We then used this amino acid sequence alignment to construct a rooted tree of EctC-type proteins using tools provided via the iTOL web server (Letunic and Bork, 2011), in order to assess the taxonomic distribution

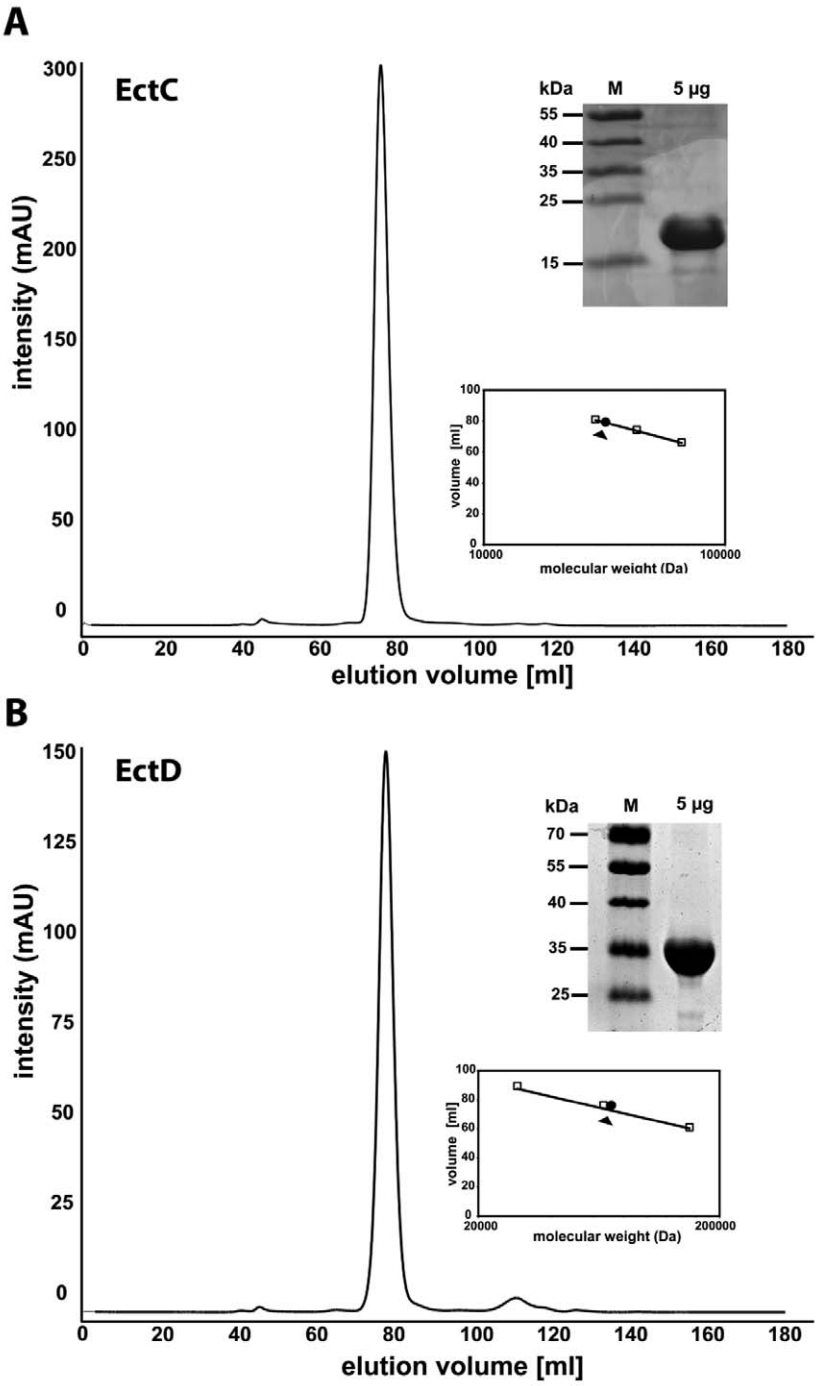

Fig. 3. Assessment of the purity of the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins and their quaternary assembly. (A) the (*Nm*)EctC-*Strep*-tag II and (B) (*Nm*)EctD-*Strep*-tag II recombinant proteins were isolated by affinity chromatography on a Streptactin matrix and analysed by SDS-polyacrylamide gel electrophoresis and size exclusion chromatography. (A) (*Nm*)EctC-*Strep*-tag II and (B) (*Nm*)EctD-*Strep*-tag II. The black arrowhead in the inserts of (A) and (B) indicate the calculated molecular mass of the recombinant proteins relative to that of marker proteins. For the (*Nm*)EctC-*Strep*-tag II protein the following marker proteins were used: albumin (66 kDa), ovalbumin (43 kDa), carboanhydrase (29 kDa). For the (*Nm*)EctD-*Strep*-tag II protein, the following marker proteins were used: alcohol-dehydrogenase (150 kDa), albumin (66 kDa), carboanhydrase (29 kDa).

of the 12 archaeal proteins and their affiliation with the nearest bacterial orthologues (Fig. 6). Nineteen microbial phyla were represented among those 723 microorganisms that possess EctC-type proteins; 16 of these phyla were derived from *Bacteria*, and three were derived from *Archaea*. The few EctC orthologues from archaeal strains cluster at three locations in the EctC-based phylogenetic tree (Fig. 6) (a phylogenetic tree listing the names of the EctC-possessing microorganisms is provided in Fig. S1), suggesting that the corresponding *ect* genes were most likely obtained by lateral gene transfer events from bacterial donor strains harbouring closely related orthologues.

In the eight members of the Thaumarchaeota possessing *ect* gene clusters, all EctC orthologues form a monophyletic clade by themselves. EctC proteins of mostly *Halomonas* species and related genera of the *Halomonadaceae*, which are affiliated with the

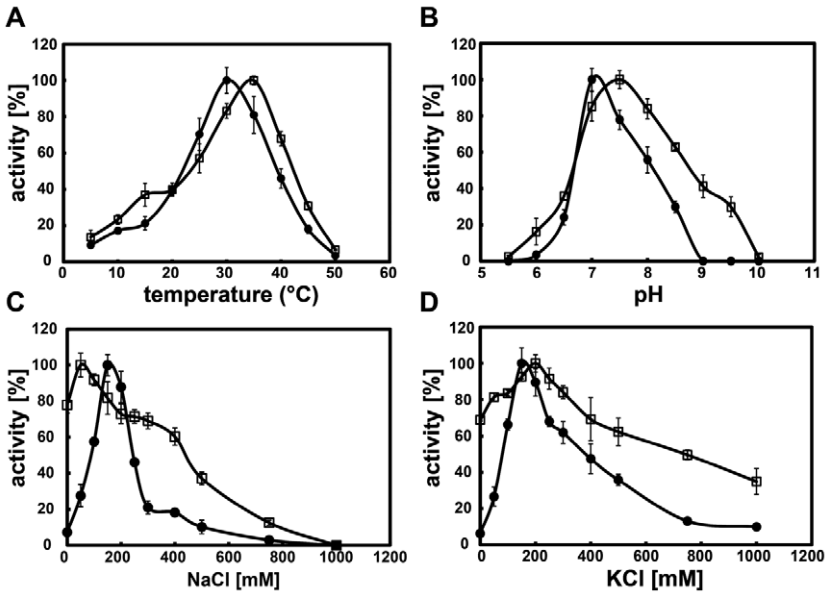

Fig. 4. Biochemical properties of the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins. Black dots: (*Nm*)EctC-*Strep*-tag II; open squares: (*Nm*)EctD-*Strep*-tag II.

Gammaproteobacteria, represent the most similar adjacent clade. Because all microorganisms from these two clades populate marine habitats, lateral transfer of the *ect* genes from a member of the *Halomonadaceae* to an ancestral strain of the genus *Nitrosopumilus* is plausible. The presence of some EctC sequences from marine *Alphaproteobacteria* (e.g. genera *Nesiotobacter*, *Zhangella* or *Marinitalea*) within the *Halomonadaceae* clade suggests that lateral gene transfer events have still been ongoing after the *ect* genes were transferred to an ancestor of *Nitrosopumilus*. The acquisition of ectoine/hydroxyectoine biosynthetic genes in the indicated strains by a lateral gene transfer event(s) is further supported by the fact that none of the AOA [*Nitrosopumilus sediminis* AR2, *Nitrosopumilus salaria* BD31, *Nitrosopumilus koreensis* MY1, *Nitrosoarchaeum limnia* (strains SFB1 and BG20), *Nitrosotenuis uzonensis* N4, *Nitrosophera gargensis* and *Nitrosocaldus* sp.] contain ectoine/hydroxyectoine biosynthetic genes.

The other two clades of archaeal EctC orthologues are present in some members of the methanogenic genera *Methanothrix* and *Methanobacterium*, respectively, but they are not found in all sequenced strains of these genera. In case of *Methanothrix*, the closest relatives are from the strictly anaerobic sulfate-reducing deltaproteobacterial genus *Desulfarculus* and the aerobic alphaproteobacterial genus *Sneathiella*, and they form a common clade. Although very diverse in their taxonomy, all these strains were isolated from freshwater sediments or wastewater treatment plants, creating ample opportunity for lateral gene transfer. From the large phylogenetic distance between *Nitrosopumilus* and *Methanothrix*, it is clear that the respective gene transfer events must have happened independently. Among the seven genome-sequenced species of the genus *Methanobacterium*, *ect* genes appear only to be present in the species *M. formicicum*. The respective EctC sequences form a common cluster with EctC orthologues from terrestrial and marine *Bacillaceae*, but also some *Betaproteobacteria* and *Gammaproteobacteria*. Therefore, it may be inferred that the *ect* genes in *M. formicicum* have been acquired by a very recent lateral transfer from one of the bacterial species with closely related *ectABC* sequences.

Since it is highly likely that the *ect* gene clusters present in members of the *Archaea* (Fig. 1B) were acquired via lateral gene transfer events, we attempted to identify possible bacterial donors by using the corresponding amino acid sequences of the ectoine/hydroxyectoine biosynthetic enzymes as query sequences in a BLASTP search of microbial genomes (Nordberg *et al*., 2013) and then recording the top 10 hits. As documented in Table S1 as the results for this type of analysis for the Thaumarchaeota '*Ca*. N. maritimus' SCM1, '*Ca*. *Nitrosopumilus*' sp. SJ, '*Ca*. Nitrosopumilus koreensis' AR1 and '*Ca*. Nitrosolumilus' sp. AR2, no clear pattern emerged that would allow to precisely pinpoint possible bacterial donor species for the archaeal *ect* gene clusters.

Taken together, our phylogenetic assessment of EctC-type proteins is consistent with our biochemical and kinetic studies of the '*Ca*. N. maritimus' SCM1 ectoine synthase (Figs 4 and 5) that ascribe bacterial-like properties to this archaeal protein. In analysing the rooted phylogenetic tree of EctC-type proteins, we noted that all orphan EctC-like proteins (Kurz *et al*., 2010; Widderich *et al*., 2014a) cluster together and form a distinct branch close to the root of the tree (Fig. 6). Among the 723 EctC-type proteins analysed, 24 belong to this group, and none was from an archaeon. The taxonomic affiliation

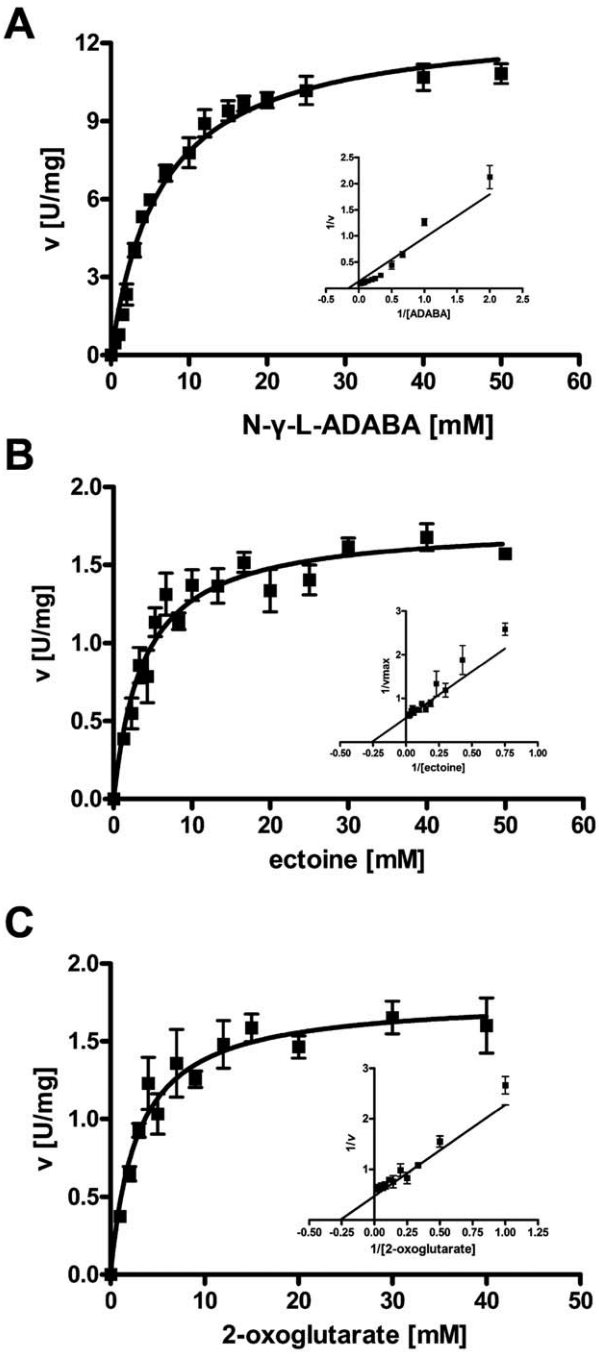

Fig. 5. Kinetic parameters of the (A) EctC for its substrate Nγ-ADABA and of the ectoine hydroxylase EctD for its substrate ectoine (B) and its co-substrate 2-oxoglutarate (C).

harbouring these orphan *ectC* genes is rather diverse (Fig. 6).

A novel transcriptional organization of the ect gene cluster in 'Ca. N. maritimus' SCM1

Inspection of the organization of the *ectABCD* gene cluster of '*Ca*. N. maritimus' SCM1 and those of other members of this genus (Fig. 1B) revealed a conserved genetic arrangement that is also found in many members of the *Bacteria* (Widderich *et al*., 2014a). The distance between the *ectA* and *ectB* genes is three bp, three bp between *ectB* and *ectC*, and the *ectC* and *ectD* coding regions overlap by two bp. These observations suggest that the *ectABCD* gene cluster of '*Ca*. N. maritimus' SCM1 is transcribed as part of an operon. While the genetic organization of the 3′ region of the *ect* genes is variable (Fig. 1B), they are all preceded in the *Nitrosopumilus* species and other marine Thaumarchaeota by an open reading frame (hyp) which codes for a hypothetical protein (Hyp; 125 amino acids) of unknown function. The coding region of the *hyp* gene and *ectA* in '*Ca*. N. maritimus' SCM1 overlaps by three bp, suggesting that the *hyp* gene is co-transcribed with the *ectABCD* gene cluster (Fig. 1B). As a specialty among members the genus *Nitrosopumilus* and other marine *Thaumarchaea* containing *ect* genes, the *ectABCD* genes of '*Ca*. N. maritimus' SCM1 are followed by an open reading frame (Fig. 1B) that was originally annotated to code for a protein of unknown function (Walker *et al*., 2010). Our renewed database searches identified this protein as a member of the MscS family (see below), mechanosensitive channels that serve as safety valves against the cell-disrupting consequences of severe osmotic downshifts (Levina *et al*., 1999). They are ubiquitously found in both *Bacteria* and *Archaea* (Kloda and Martinac, 2002; Booth and Blount, 2012; Naismith and Booth, 2012; Wilson *et al*., 2013; Booth, 2014; Booth *et al*., 2015). We refer to this protein in the following as the (*Nm*)MscS channel. The distance between the end of the *ectD* gene and the start of the *mscS* gene is 25 bp (Fig. 1B), suggesting that the *mscS* gene might be co-transcribed with the *ectABCD* gene cluster as well. Overall, the tight physical organization of the *hyp-ectABCD-mscS* gene cluster (Gene IDs: Nmar_1347, Nmar_1346, Nmar_1345, Nmar_1344, Nmar_1343, Nmar_1342) suggests that these genes are transcribed as an operon.

To test for a possible co-transcription of the *hyp-ectABCD-mscS* genes (Fig. 7A), we carried out an reverse transcription polymerase chain reaction (RT-PCR) analysis using RNA samples that were isolated from cells of '*Ca*. N. maritimus' SCM1 that had been grown under increased osmotic stress conditions (48 g NaCl l^{-1} which corresponds to 0.94 M NaCl) (Fig. 2A). The data presented in Fig. 7B demonstrate that the *hyp-ectABCD-mscS* gene cluster is indeed expressed as a transcriptional unit. Co-transcription of ectoine/hydroxyectoine biosynthetic genes (Widderich *et al*., 2014a) has been routinely observed in members of the *Bacteria* (Kuhlmann and Bremer, 2002; Bursy *et al*., 2007; Kuhlmann *et al*., 2008). However, there is no report where they are co-transcribed with a gene for a mechanosensitive channel.

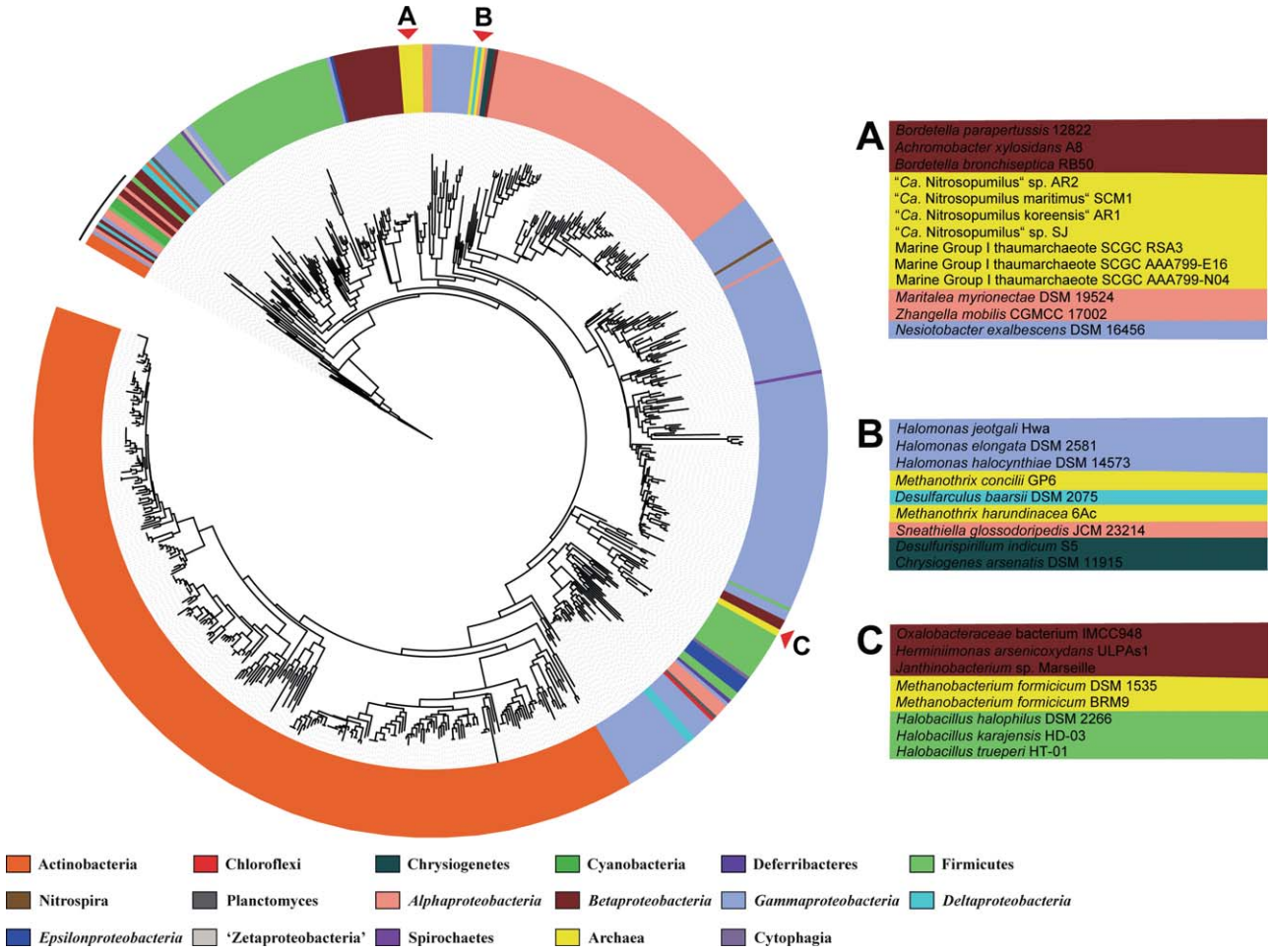

Fig. 6. Phylogenetic analysis of EctC-type proteins. Based on an amino acid sequence alignment of 723 EctC-type proteins, a rooted phylogenetic tree was constructed with the ITOL program (Letunic and Bork, 2011). The different phyla are marked by different colours and, for simplicity, the names of the *ectC*-possessing microorganisms were left out. An expanded version of the tree containing the names of the microorganisms predicted to produce the EctC protein is given in Fig. S1. The three regions in the phylogenetic tree populated by archaeal EctC proteins are highlighted by red arrowheads. An enlarged section of the phylogenetic tree representing the archaeal EctC proteins and their nearest bacterial orthologues are shown in A to C. The position of those 24 EctC-type proteins that originate from bacteria that lack identifiable *ectAB* genes are marked above the phylogenetic tree by a black bar.

The eight *ect* gene clusters present in Thaumarchaeota are all preceded by a gene (*hyp*) (Fig. 1B) that encodes a protein (125 amino acids) of unknown function. These Hyp proteins possess a degree of amino acid sequence identity ranging between 82% to 31%, and such proteins were found through database searches also in the Thaumarchaeota MY2 and N4, both of which do not possess *ect* genes. The Hyp proteins show no significant sequence similarity to annotated proteins in databases, and we can thus neither speculate about their potential function, nor can we comment on the obvious question why the *hyp* gene is co-transcribed with the *ect* gene cluster in various Thaumarchaeota (Fig. 7).

Expression of the *ect* genes in *Bacteria* is typically induced when the cells are exposed either to sustained or suddenly imposed osmotic stress (Kuhlmann and Bremer, 2002; Calderon *et al*., 2004; Bursy *et al*., 2007; Kuhlmann *et al*., 2008; Salvador *et al*., 2015). We therefore wondered whether this was also the case in the archaeon '*Ca*. N. maritimus' SCM 1. We therefore isolated total RNA from cells that had been grown under low, optimal and high salt conditions (Fig. 2A) (13, 25 and 39 g NaCl l^{-1}, respectively) and assed the transcript levels of a gene positioned either early (*ectA*) or late (*mscS*) in the *hyp-ectABCD-mscS* poly-cistronic messenger (m)RNA via q-RT-PCR. The cells of the cultures grown under different salt-stress conditions were all harvested when they were in the same growth phase to exclude possible effects of this parameter on the quantitative RT-PCR data. We used the transcript of the *amoA* gene (Gene ID: Nmar_1500) of '*Ca*. N. maritimus' SCM 1, as a reference (Nakagawa and Stahl, 2013) to benchmark the *ectA* and *mscS* transcript

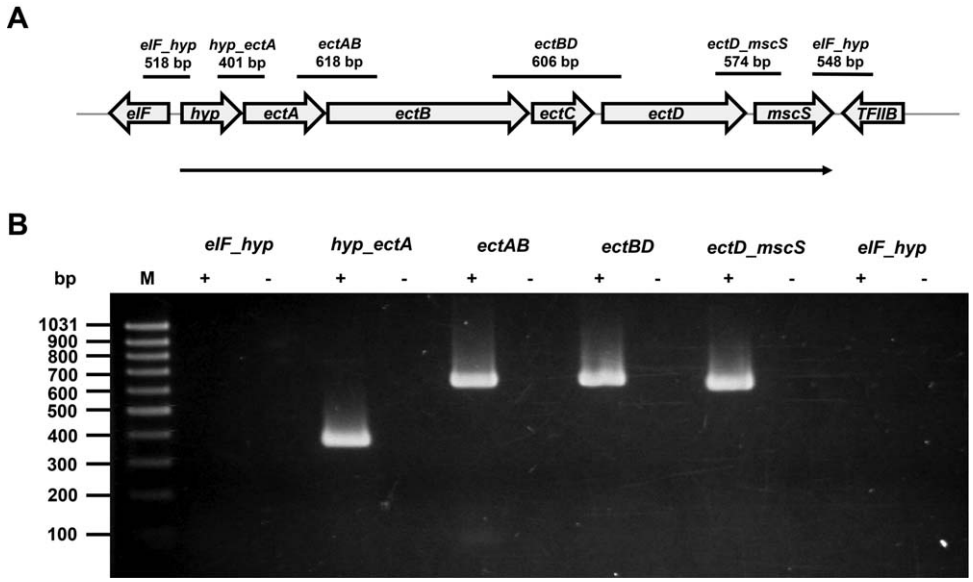

Fig. 7. Analysis of the co-transcription of the *hyp-ectABCD-mscS* gene cluster from '*Ca.* N. maritimus' SCM1 by RT-PCR.
A. Genetic organization of the *hyp-ectABCD-mscS* gene cluster and predictions of the size of the PCR fragments made under the assumption that a given set of genes is co-transcribed.
B. Analysis of the sizes of the DNA fragments generated in the RT-PCR experiment by agarose gel electrophoresis. The symbol '+' indicates those samples reacted with reverse transcriptase, and the symbol '-' denotes samples prepared in its absence to ensure that the observed PCR products did not result from DNA contaminations of the RNA samples.

levels. Indeed, the *amoA* transcript turned out to be a suitable reference marker for salt-stressed and non-stressed cells since the level of its mRNA did not vary significantly under the above-described growth conditions (Fig. 2C). In contrast, the level of the *ectA* and of the *mscS* transcripts was strongly upregulated from an already substantial level in response to salt stress; by 5.7-fold and 5.1-fold respectively (Fig. 2C). An even greater enhanced induction of the *ectA* and *mscS* transcript levels was observed when '*Ca*. N. maritimus' SCM1 cells that had been grown under optimal salt conditions (Fig. 2A) (26 g NaCl l^{-1}) were subjected to a severe osmotic upshift (by adding 23 g NaCl l^{-1} to the culture). This severe salt shock triggered a 52-fold and 76-fold increase in the *ectA* and *mscS* transcripts from their levels in cells grown under optimal salt conditions (Fig. 2D). Hence, both chronic and acute salt stresses are important environmental cues leading to strongly enhanced transcription of the *hyp-ectABCD-mscS* operon.

(Nm)MscS is a functional mechanosensitive channel

Since there is no report in the literature describing the co-transcription of an *ect* gene cluster with a gene encoding a mechanosensitive channel, we wondered whether the (*Nm*)MscS channel was functional. To study the functionality of the (*Nm*)MscS channel, we used a genetic complementation experiment (Levina *et al*., 1999) with an *E. coli* mutant strain (MJF641) (also referred to as the Δ7 strain) that lacks all currently know MscM-, MscS- and MscL-type mechanosensitive channels (Edwards *et al*., 2012). We obtained a synthetic version of the '*Ca*. N. maritimus' SCM 1 *mscS* gene that had been optimized for its expression in *E. coli* and placed it under the transcriptional control of the *lac* promoter present on plasmid pTrc99a (Amann *et al*., 1988), thereby yielding plasmid pLC18. We then used a test developed by Levina and colleagues (1999) to assess the functionality of mechanosensitive channels *in vivo*. It relies on a rapid down-shock of high osmolarity grown cells lacking or expressing channel-forming proteins and then monitoring cell viability (Levina *et al*., 1999). The parent (strain Frag1) of the Δ7 mutant-strain carrying the vector plasmid pTrc99a survived such an osmotic downshift essentially unscathed, whereas only 12.9% (±5%) of the cells of the Δ7 mutant-strain MJF641 (pTrc99a) survived such a treatment. In contrast, when MJF641 (pLC18; *mscS*$^{+}$) was subjected to such an osmotic down-shock, the (*Nm*)MscS protein rescued cellular survival to a large extent (62% ± 5% surviving cells) (Fig. 8A). This experiment therefore unambiguously demonstrates that the gene following the *ectABCD* gene cluster (Figs 1B and 7A) in the '*Ca*. N. maritimus' SCM1 genome (Walker *et al*., 2010) encodes a functional mechanosensitive channel.

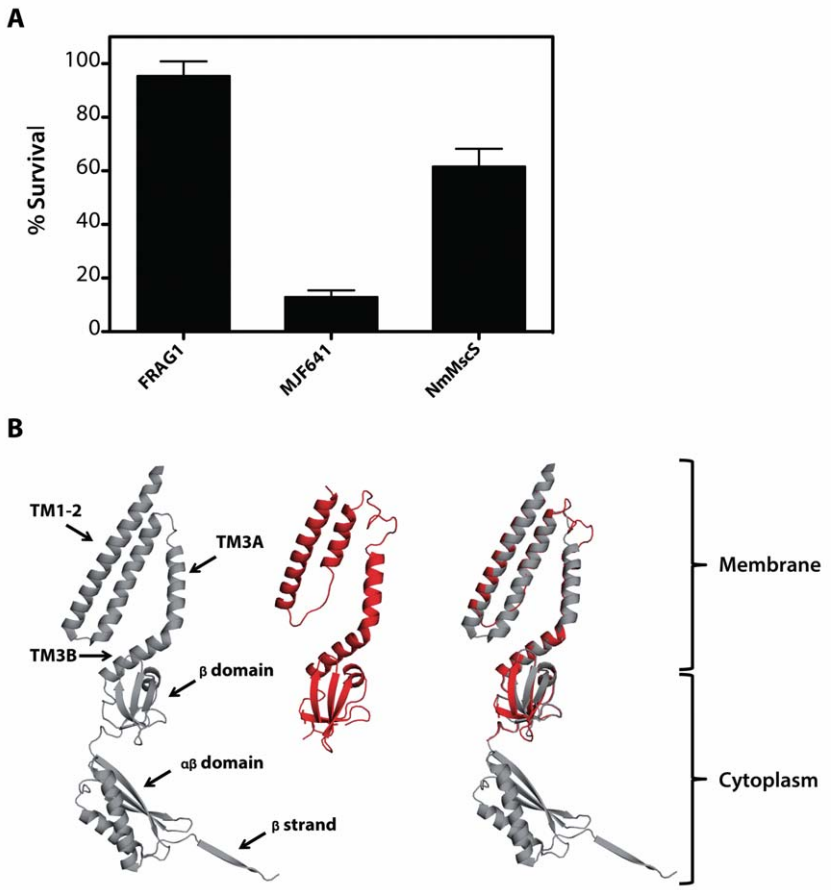

Fig. 8. Functional assessment of the (*Nm*)MscS protein and *in silico* prediction of its structure.
A. Cells of the *E. coli* strain MJF641 which is defective in all currently known MscL-, MscS- and MscM-type channels (also referred to as Δ7 mutant) (Edwards *et al*., 2012) and its parent strain Frag1 carrying the empty vector plasmid pTrc99a (Amann *et al*., 1988) were grown in high-salinity minimal medium (with 0.3 M NaCl) and were then subjected to a rapid osmotic downshift (minimal medium without additional NaCl) (Levina *et al*., 1999). The same procedure was used with strain MJF641 harbouring a plasmid (pLC18) that contains the '*Ca*. N. maritimus' SCM1 *mscS*$^{+}$ gene expressed under the control of the *lac* promoter present on pTrc99a. The number of cells surviving such rapid osmotic downshifts was determined by spotting them on LB agar plates. The data shown represent the mean and standard deviation of four independently conducted experiments.
B. Crystal structure of a monomer of the *E. coli* MscS protein (left) from the open *E. coli* MscS heptameric channel assembly (protein database accession code: 2VV5) (Wang *et al*., 2008) was automatically chosen by the SWISS-MODEL web server (Biasini *et al*., 2014) as the template to generate an *in silico* model (middle) of the (*Nm*)MscS protein. The PyMOL suite (https://www.pymol.org/) was used to visualize these structures and overlay them with each other (right).

In silico analysis of the (Nm)MscS mechanosensitive channel

Mechanosensitive channels acting as safety valves against osmotic downshifts can be grouped into three types (MscM, MscS, MscL) and provide the microbial cell with a graded stress response to an osmotic challenge through their different threshold levels of opening and their different diameters of the fully open channels (Haswell *et al*., 2011; Booth and Blount, 2012; Edwards *et al*., 2012; Booth, 2014). A database search identified the '*Ca*. N. maritimus' SCM1 protein as a member of the MscS family with an overall level of amino acid sequence identity to the *E. coli* MscS protein (Levina *et al*., 1999) of about 12%. However, in comparison to the *E. coli* protein, (*Nm*)MscS lacks about 100 amino acids at its carboxy-terminus (Fig. S2). When this is taken into consideration, and a stretch of badly aligning 20 amino acids from the N-terminus are removed as well from the comparison of the two proteins, the degree of amino acid sequence identity increases to about 16%. Consistent with this amino acid sequence alignment, the SWISS MODEL server (http://swissmodel. Expacy.org/) (Biasini *et al*., 2014) automatically chose a monomer of the homo-heptamertic *E. coli* MscS channel in its 'open' form (PDB entry 2VV5) (Wang *et al*., 2008) as its modelling template (Fig. 8B). Like its *E. coli* MscS counterpart (Bass *et al*., 2002; Wang *et al*., 2008), the (*Nm*)MscS protein is predicted to contain three membrane-spanning segments and a carboxy-terminus that protrudes into the cytoplasm (Fig. 8B). Inspection of the (*Nm*)MscS *in silico* model revealed an important difference with respect to the *E. coli* MscS protein, as the cytoplasmic αβ-domain and an adjacent β-sheet is missing (Fig. 8B).

Interestingly, a protein similar to that of (*Nm*)MscS (Fig. S2) is encoded in the genome of '*Ca.* Nitrosopumilus' sp. AR2. However, in contrast to the situation in '*Ca*. N. maritimus' SCM1, the putative *mscS* gene is transcribed divergently from the *hyp-ectABCD* gene cluster (Fig. 1B). Of note is also that the genome sequence of '*Ca*. N. maritimus' SCM1 (Walker *et al*., 2010) encodes in addition to the (*Nm*)MscS protein (gene ID: Nmar_1342), four other MscS-encoding genes (gene ID: Nmar_1337. Nmar_1773, Nmar_1335, Nmar_0971). These MscS-type proteins vary considerably with respect to the length of their carboxy-termini. '*Ca.* Nitrosopumilus maritimus' SCM1 does not possess a gene for an MscL-type channel.

We also inspected the genome sequences of all other archaeal strains that possess *ect* gene clusters for the presence of *mscS* and *mscL*-type genes. Each of these microorganisms possesses multiple numbers (between one and four) of *mscS*-type genes but only *M. harundinacea* 6Ac and *Methanotrix concilii* GP-6 possess *mscL*-type genes as well (Table S2). Hence, MscS-type mechanosensitive channels (Levina *et al*., 1999; Wang *et al*., 2008) seem to be the preferred device to physiologically cope with rapid osmotic downshifts by those *Archaea* whose genome sequences we have inspected.

Discussion

Ectoine and its derivative 5-hydroxyectoine are well-recognized compatible solutes that are synthesized, or taken up, in response to osmotic stress by many microorganisms (Pastor *et al*., 2010; Widderich *et al*., 2014a). Their biophysical properties allow them not only to serve as effective stress protectants but also to function as chemical chaperones as they are able to preserve the functionality of proteins, macromolecular complexes, membranes and even entire cells (Lippert and Galinski, 1992; Manzanera *et al*., 2002; Harishchandra *et al*., 2010; Pastor *et al*., 2010; Kunte *et al*., 2014; Tanne *et al*., 2014). Our updated database search identified the signature enzyme for ectoine production, the ectoine synthase (Ono *et al*., 1999b; Widderich *et al*., 2014a), in about 4.7% of 27.789 microbial genome sequences deposited in the JGI database at the time of the search. Predicted ectoine producers are found in 19 microbial phyla. Only three of these represent *Archaea*, and only 12 archaeal strains belonging to the genera *Nitrosopumilus*, *Methanothrix* and *Methanobacterium* are predicted to engage in ectoine synthesis among the 557 *Archaea* with deposited genome sequences. This data set therefore reinforces our previous conclusion that ectoine is a compatible solute primarily synthesized by members of the *Bacteria* (Widderich *et al*., 2014a).

It is notable that among the 12 archaeal strains possessing ectoine biosynthetic genes, eight are affiliated with the genus *Nitrosopumilus*. We therefore used '*Ca*. N. maritimus' SCM1 (Könneke *et al*., 2005; Walker *et al*., 2010) as a model system to study ectoine/hydroxyectoine synthesis and the genetic control of its *ect* gene cluster in this physiologically well-characterized and globally abundant member of the domain *Archaea* (Stahl and de la Torre, 2012; Offre *et al*., 2013). Both the transcription of the ectoine/hydroxyectoine biosynthetic genes and the production of these compatible solutes in '*Ca*. N. maritimus' SCM1 are responsive to osmotic stress (Fig. 2), thereby strongly suggesting that ectoines serve an osmostress-protective function in this Thaumarchaeon, as firmly established in many members of the *Bacteria* (Pastor *et al*., 2010; Kunte *et al*., 2014). We surmise that the data presented here on these issues for '*Ca*. N. maritimus' SCM1 are relevant for an understanding of the physiology and genetics of the production of these compatible solutes by other strains and species of the Thaumarchaeota, and we suggest that is also true for those species of the methanogens *Methanothrix* and *Methanobacterium* carrying ectoine biosynthetic genes (Fig. 1B).

A phylogenetic tree derived from an amino acid sequence alignment of 723 unique EctC-type proteins revealed that the 12 archaeal EctC orthologues cluster in three separate locations among their bacterial counterparts (Figs 6 and S1). These data strongly suggest that the *Archaea* harbouring *ectABC(D)* gene clusters have acquired them via horizontal gene transfer events from bacterial donor strains living in the same habitats. This suggestion is also consistent with our finding that the key enzymes for ectoine biosynthesis, the ectoine synthase (EctC) and the ectoine hydroxylase (EctD) from '*Ca*. N. maritimus' SCM1 possessed biochemical and kinetic properties (Figs 4 and 5) resembling those of their bacterial counterparts (Ono *et al*., 1999b; Bursy *et al*., 2007; 2008; Widderich *et al*., 2014a; Moritz *et al*., 2015). Since the 12 EctC proteins derived from *Archaea* are found in three different locations in the ectoine synthase-based phylogenetic tree (Figs 6 and S1), it seems likely that the *ectABC(D)* genes were introduced into members of the genera *Nitrosopumilus*, *Methanothrix* and *Methanobacterium* by separate events. However, by comparing the amino acid sequences of the archaeal ectoine/hydroxyectoine biosynthetic enzymes to their nearest orthologues from *Bacteria* (Table S1), we were unable to derive at a consistent picture of the potential donor *Bacteria*.

It is well known that horizontal gene transfer is an important driver of microbial evolution helping microorganisms to rapidly obtain new metabolic capabilities and develop new stress management skills so that they can better cope with the situation in their original habitat, or explore new ecosystems (Ochman *et al*., 2000; Soucy *et al*., 2015). This process shaped the gene content of archaeal genomes in a major way (Nelson-Sathi *et al*., 2015). For instance, it was concluded from a metagenomic analysis of deep Mediterranean waters that about 24% of the genes derived from planktonic Thaumarchaeota were of bacterial origin (Deschamps *et al*., 2014).

While this paper was under review, the genome sequences of two novel *Nitrosopumilus* strains originating from coastal surface waters of the Northern Adriatic Sea became available. We found that neither '*Ca*. Nitrosopumilus piranensis' D3C, nor '*Ca*. Nitrosopumilus adriaticus' NF 5 (Bayer *et al*., 2015) possess ectoine

biosynthetic genes. Remarkably, based upon a 16 S-23S rRNA tree derived from cultivated Thaumarchaeota with sequenced genomes, '*Ca*. Nitrosopumilus piranensis' D3C, is the closest relative of '*Ca*. N. maritimus' SCM1 and '*Ca*. Nitrosopumilus koreensis' AR1 (Bayer *et al*., 2015), both of which possess ectoine/hydroxyectoine biosynthetic gene clusters (Fig. 1B). Hence, it seems likely that the possession, or lack, of ectoine/hydroxyectoine biosynthetic genes among the AOA reflects specific ecological niche adaptive evolutionary processes (Blainey *et al*., 2011; Gubry-Rangin *et al*., 2015).

No genetic system is currently available for '*Ca*. N. maritimus' SCM1 (Walker *et al*., 2010) that would allow the construction of a mutant derivative lacking the *ectABCD* gene cluster. However, the isolation of a taxonomically very close relative, '*Ca*. Nitrosopumilus piranensis' D3C, that lacks these genes naturally (Bayer *et al*., 2015) provides the opportunity to compare the physiology and growth properties of these two Thaumarchaeota to reveal possible benefits of ectoine/hydroxyectoine production.

The function-preserving features of ectoines have triggered considerable biotechnological interest and led to the development of an industrial scale production process that relies on a severe osmotic down-shock to release newly synthesized ectoines from microbial producer cells (Schwibbert *et al*., 2011; Kunte *et al*., 2014). This so-called bacterial milking procedure builds on an evolutionarily conserved cellular emergency stress reaction to sudden osmotic down-shocks that relies on the transient opening of MscM-, MscS- and MscL-type mechanosensitive channels to prevent cell rupture (Booth, 2014). These types of safety valves have a deep evolutionary origin and are found both in *Archaea* and *Bacteria* (Kloda and Martinac, 2002; Pivetti *et al*., 2003).

Most interestingly, we discovered that the *ectABCD* gene cluster of '*Ca*. N. maritimus' SCM1 is co-transcribed with the gene for an MscS-type mechanosensitive channel (*hyp-ectABCD-mscS*) (Fig. 7). While nothing can currently be concluded about the role of the *hyp* gene product, the *mscS*-encoded mechanosensitive channel is functional (Fig. 8A). A genetic configuration that entails the co-transcription of a *mscS* gene along with the *ectABC(D)* operon has not been described before in any ectoine/hydroxyectoine-producing microorganism. The *hyp-ectABCD-mscS* operon seems to be a rather sophisticated genetic device to counteract the detrimental effects of osmotic fluctuations on cell physiology. The osmotic induction of its transcription allows '*Ca*. N. maritimus' SCM1 to cope with increased salinity through enhanced production of the osmostress protectants ectoine and hydroxyectoine. At the same time, it prepares the high osmolarity stressed cells for an osmotic down-shift that eventually will occur in narrow transition zones from marine to estuarine ecosystems through enhanced preemptive synthesis of the (*Nm*)MscS (Santoro *et al*., 2008) mechanosensitive channel. We are aware that in the hyperthermophilic Creanarchaeon *Thermoproteus tenax*, an *mscS*-type gene is co-transcribed with enzymes required for the synthesis of the compatible solute trehalose (Zaparty *et al*., 2013). However, in contrast to the biological data we provide here for the (*Nm*)MscS protein (Fig. 8A), the functionality of the Msc protein from *T. tenax* has not yet been experimentally assessed.

Both the MscS and the MscL mechanosensitive channels are widely considered as non-specific pores when they are fully opened and they have calculated pore diameters of about 13 Å and 30 Å respectively (Cruickshank *et al*., 1997; Wang *et al*., 2008). However, the fact that both in '*Ca*. N. maritimus' SCM1 and in *T. tenax*, *mscS*-type genes are co-expressed with genes encoding enzymes for the synthesis of the compatible solutes ectoine/hydroxyectoine and trehalose, respectively, raises the question if the (*Nm*)MscS and (*Tt*)MscS channels are not endowed with a certain degree of substrate specificity. It is of interest to consider in this context the properties of the MscCG mechanosensitive channel of *Corynebacterium glutamicum*. Like (*Nm*)MscS, MscCG is a MscS-type channel and participates in the osmotic stress response of *C. glutamicum* (Börngen *et al*., 2010). However, it also serves as an export channel for glutamate *in vivo* (Nakamura *et al*., 2007; Becker and Krämer, 2015). Despite these latter observations, electrophysiological studies did not suggest that the MscCG channel possesses substrate specificity for glutamate (Becker *et al*., 2013; Nakayama *et al*., 2013).

Most members of the diverse MscS family are structurally built on a common core architecture in which the monomer possesses three trans-membrane-spanning segments. The membrane-embedded helix one and two form the sensor for membrane tension and part of the broken helix 3 lines the pore of the heptameric channel assembly (Bass *et al*., 2002; Wang *et al*., 2008; Naismith and Booth, 2012) (Fig. 8B). This latter helix provides a structural link between the pore domain and the carboxy-terminal part of the MscS monomer that forms an extended funnel-like vestibule inside the cytoplasm in the fully assembled heptameric MscS channel. Solutes and solvents must pass on their way out of the cell through this structurally complex vestibule when the channel opens (Bass *et al*., 2002; Wang *et al*., 2008; Naismith and Booth, 2012; Wilson *et al*., 2013; Booth *et al*., 2015). Through conformational changes, the cytoplasmic cage probably plays an important role for the opening and closing reactions of the MscS channel as a whole. It might also serve as a sensor for increases in macromolecular crowding of the cytoplasm (Rowe *et al*., 2014), a process that is

strongly influenced by osmotically instigated water fluxes in or out of the cell (Wood, 2011).

We therefore note with considerable interest that (*Nm*)MscS protein lacks, in comparison with the *E. coli* MscS protein (Bass *et al*., 2002; Wang *et al*., 2008), a large part of the carboxy-terminus (Fig. S2). This should lead to a structurally rather different make-up of the cytoplasmic portion of the (*Nm*)MscS channel (Fig. 8B). Mechanosensitive channels open in response to increased tension in the plain of the lipid bilayer of the cytoplasmic membrane that results from an increase in turgor (Naismith and Booth, 2012; Booth, 2014). These types of proteins therefore have to closely interact with lipids to perform their physiological function. The types of lipids that are present in the Thaumarchaeon '*Ca*. N. maritimus' SCM1 are very different from those of *E. coli* (Elling *et al*., 2014), yet the (*Nm*)MscS protein is physiological active in the surrogate *E. coli* host strain (Fig. 8A). Based on the discussed properties of the (*Nm*)MscS protein, it might be well worthwhile to study its electrophysiological properties, gating behaviour and potential substrate specificity in its own right.

Experimental procedures

Chemicals and synthesis of the substrate of the ectoine synthase

Ectoine and hydroxyectoine were kind gifts from Dr Irina Bagyan (bitop AG, Witten, Germany). Anhydrotetracycline, desthiobiotine and the strepavidin affinity matrix for the purification of *Strep*-tag II-labelled protein were purchased from IBA GmbH (Göttingen, Germany). Alkaline hydrolysis of ectoine was performed as reported previously (Kunte *et al*., 1993) to obtain the substrate, *N*-γ-acetyl-L-2,4-diaminobutyrate (N-γ-ADABA), for the EctC (Peters *et al*., 1990; Ono *et al*., 1999b). All chemicals required for this synthetic process were purchased from Sigma Aldrich (Steinheim, Germany) or Acros (Geel, Belgium). Briefly, hydrolysis of ectoine (284 mg, 2.0 mmol) was accomplished in aqueous potassium hydroxide (KOH) (50 ml, 0.1 M) for 20 h at 50°C (Kunte *et al*., 1993). The reaction mixture was subsequently neutralized with perchloric acid (60% in water, 4 ml), and the precipitated potassium perchlorate was filtered off. The filtrate was concentrated under reduced pressure, and the residue was purified by repeated chromatography on a silica gel column (Merck silica gel 60). Chromatography was performed by using a gradient of ethanol/25% ammonia/water 50:1:2 – 10:1:2 as the eluent to yield N-γ-ADABA. We recovered 192 mg (1.20 mmol; 60%) of N-γ-ADABA from the starting material (284 mg ectoine, 2.0 mmol). The identity and purity of the isolated N-γ-ADABA was unequivocally established both by thin-layer chromatography (TLC) (silica gel 60 F254 TLC plates; Merck) and nuclear magnetic resonance (^{1}H-NMR and ^{13}C-NMR) spectroscopy (Peters *et al*., 1990; Ono *et al*., 1998) on a Bruker AVIII-400 or DRX-500 NMR spectrometer. (i) TLC: R_f of Nγ-ADABA = 0.55 (ethanol/25% ammonia/water 7:1:2); (ii) ^{1}H-NMR (400 MHz, D_2O): δ = 3.71 (dd, ^3J(H,H) = 7.6 Hz, ^3J(H,H) = 5.6 Hz, 1H, CH), 3.41 – 3.24 (m, 2H, CH_2), 2.15 – 2.01 (m, 2H, CH_2), 1.99 (s, 3H, CH_3) ppm; (iii) ^{13}C-NMR (100 MHz, D_2O): δ = 177.5 (CO), 177.0 (COOH), 55.3 (CH), 38.3 (CH_2), 33.0 (CH_2), 24.6 (CH_3) ppm.

Recombinant DNA procedures and construction of plasmids

The DNA sequences of the *ectC*, *ectD* and *mscS* genes were retrieved from the genome sequence (accession number: NC_010085) (Walker *et al*., 2010) of '*Ca*. N. maritimus' SCM1 and were used as templates for the synthesis of codon-optimized versions of these genes for their expression in *E. coli*. These synthetic genes were constructed either by GenScript (Piscataway, USA), or by LifeTechnologies (Darmstadt, Germany). Their DNA sequences were deposited in the NCBI database under accession numbers KR002039 (*ectC*), JN019033 (*ectD*) and KT313590 (*mscS*). To allow the overproduction and purification of the '*Ca*. N. maritimus' SCM1 EctC and EctD proteins in *E. coli*, we constructed recombinant versions of the corresponding genes. The *ectC* and *ectD* sequences were retrieved from the plasmids provided by the suppliers of the synthetic constructs, and inserted into the expression vector pASG-IBA3 (IBA GmbH, Göttingen, Germany). The resulting expression vectors [pWN48 (*ectC*$^+$) and pMP45 (*ectD*$^+$)] permit the effective synthesis of the '*Ca*. N. maritimus' SCM1 recombinant EctC and EctD proteins in *E. coli*. Both proteins are fused at their carboxy-termini with a *Strep*-tag II affinity peptide (NWSHPQFEK). This allows the purification of the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins by affinity chromatography on a streptavidin matrix.

For the heterologous expression of the *mscS* from '*Ca*. N. maritimus' SCM1 in *E. coli*, the codon-optimized synthetic *mscS* gene was amplified from the plasmid obtained from the supplier (LifeTechnologies) by PCR using custom synthesized DNA primers. Short DNA fragments carrying NcoI and HindIII restriction sites, respectively, were attached either to the 5′ or to the 3′ prime ends of the PCR product, thereby enabling its directional insertion into the expression vector pTrc99a (Amann *et al*., 1988). This positioned the transcription of the *mscS* gene under the control of the *lac* promoter carried by the pTrc99a vector and resulted in the isolation of plasmid pLC18. The correct nucleotide sequence of the *mscS* gene carried by pLC18 was ascertained by DNA sequence analysis, which was performed by Eurofins MWG Operon (Ebersberg, Germany).

Bacterial strains

The *E. coli* strain MJF641 which is defective in all currently known MscL-, MscS- and MscM-type channels (*mscL mscS mscK ybdG ybiO yjeP ynal*) (also referred to as Δ7) and its parent Frag1 (*thi rha lac gal*) have been described (Edwards *et al*., 2012) and were a kind gift of Dr Ian Booth and Dr Samanta Miller (University of Aberdeen, Aberdeen, Scotland; UK). The *E. coli* B strain BL21 (Dubendorff and Studier, 1991) served as the host strain for plasmids that were used for the overexpression of recombinant versions of the '*Ca*. N. maritimus' SCM1 *ectC* and *ectD* genes. The '*Ca*. N.

maritimus' strain SCM1 (Könneke *et al*., 2005) was from stocks of the laboratory of Dr M. Könneke (MARUM, University of Bremen; Germany).

Media and growth conditions

'*Candidatus* Nitrosopumilus maritimus' SCM1 was cultivated at 28°C in 15-l batch cultures using a 4-(2-hydroxyethyl)-1-piperazineethanesulfonate (HEPES)-buffered medium as described previously (Könneke *et al*., 2005; Martens-Habbena *et al*., 2009). Ammonia (1.5 mM NH_4Cl) served as the energy source, and 1 mM sodium bicarbonate was added as the carbon source. For optimal growth conditions, cells of *Ca.* N. maritimus from two batches (a total of 30 l) were grown in media with 26 g l^{-1} NaCl. This medium has an osmolarity of 928 mOsm. For growth conditions with increased salinity, cells were grown in three batches (a total of 45 l) of media containing 48 g l^{-1} NaCl. This medium has a measured osmolarity of 1598 mOsm. During incubation, cultures were slightly shaken by hand once a day to provide oxygen. A spectrophotometric assay was used to follow the growth of '*Ca*. N. maritimus' SCM1 by measuring the formation of nitrite (Strickland and Parson, 1972). Cells were harvested in the late growth phase (production of 1.1-1.3 mM nitrite) with a cross-flow filtration system equipped with a 0.1 μm pore size filter cassette (Sartocon-Slice Microsart, Sartorius, Göttingen, Germany). For ectoine and hydroxyectoine analysis, concentrated cell suspensions were fixed with 4% formaldehyde (16% formaldehyde solution (w/v) Thermo Scientific, Rockford, USA), and stored refrigerated.

For studying the expression of the *ectA* and *mscS* in '*Ca*. N. maritimus' SCM1 at different salt conditions of the growth medium, cells were cultured in 5 L batch cultures at 28°C in HEPES buffered medium (pH 7.6) containing 1.5 mM ammonia as sole energy source as described previously (Könneke *et al*., 2014). Medium for low, optimal and high salt conditions contained 13, 26 and 39 g NaCl L^{-1} respectively. Cultures grown at different salinities were harvested in the same growth phase (when about two thirds of the ammonia was converted to nitrite) with a cross-flow filtration system (0.1 μm Sartocon slice Microsart filter, Sartorius Stedim, Göttingen, Germany). Concentrated cells were centrifuged (45 min, 4415 x *g*), and pellets were stored with RNAlater (SIGMA Life Science, Taufkirchen, Germany) at 4°C following the instructions of the manufacturer. Salt shock conditions were created by addition of 23 g NaCl L^{-1} into a 15 L batch culture of '*Ca*. N. maritimus' SCM1 in mid-growth phase grown at optimal salt conditions (26 g NaCl L^{-1}). After an additional incubation period of 24 h at high-salinity growth conditions, the cells were harvested as described above. Growth in all cultivation experiments of '*Ca*. N. maritimus' SCM1 was monitored by following the formation of nitrite. Purity of the cultures was routinely checked by phase contrast microscopy.

Escherichia coli strains were routinely maintained on LB agar plates (Miller, 1972). When strains contained plasmids, ampicillin (100 μg ml^{-1}) was added to the growth medium to select for the presence of the plasmids. For the analysis of cell viability of strains Frag1(pTrc99a), MJF641(pTrc99a) and MJF641(pLC18) subsequent to an osmotic down-shock, we used the growth medium and the procedure described by Levina and colleagues (1999). Osmotically unstressed cells were grown at 37°C in a 100-ml Erlenmeyer flask filled with 20 ml of medium in a shaking water bath (set to 220 rpm) using a citrate-phosphate buffered chemically defined medium (pH 7.0). It contained per litre: 8.58 g Na_2HPO_4, 0.87 g K_2HPO_4, 1.34 g citric acid, 1.0 g (NH_4SO_4), 0.001 g thiamine, 0.1 g $MgSO_4{\cdot}7H_2O$ and 0.002 g $(NH_4)_2SO_4{\cdot}FeSO_4{\cdot}6H_2O$, and was supplemented with 0.2% (w/v) glucose as the carbon source (Levina *et al*., 1999). This medium possesses a measured osmolarity of 235 mOsm. For cells that were grown at high salinity, 0.3 M NaCl were added to the basal medium; it had a measured osmolarity of 730 mOsm. The osmolarity of growth media was determined with an osmometer (Vapor Pressure Osmometer 5500, Wesco, USA).

Functional complementation of the mechano-channel-defective E. coli *mutant MJF641 by the '*Ca. *N. maritimus' SCM1* MscS *protein*

The *E. coli* strains Frag1 and MJF641 (Edwards *et al*., 2012) harbouring different plasmids were inoculated in 5 ml LB medium containing ampicillin (100 μg ml^{-1}) and were grown for 5 h. Cells were subsequently transferred into the above-described citrate-phosphate medium and incubated overnight at 37°C. Cells were then diluted to an OD_{578} of 0.05 into 20 ml of the above-described minimal medium, or into 20 ml medium that contained 0.3 M NaCl, and the cultures were subsequently grown to an OD_{578} of 0.15. At this point, IPTG was added to the cultures (final concentration 1 mM) to induce the activity of the *lac* promoter present on the backbone of the expression plasmid pTrc99a (Amann *et al*., 1988), thereby triggering the transcription of the codon-optimized '*Ca.* N. maritimus' SCM1 *mscS* gene. Growth of the cells was subsequently continued until they reached an OD_{578} of about 0.35. These cultures were then diluted 20-fold into pre-warmed medium (37°C) containing 1 mM IPTG with (no osmotic down-shock) or without (osmotic down-shock) 0.3 M NaCl; the cells were subsequently incubated at 37°C in a shaking water bath for 30 min. To determine the number of the cells that survived the osmotic down-shock, 50-μl samples were taken after 30 min of incubation from these cultures, serial diluted in four independent sets in media with corresponding osmolarities and four 5 μl samples of the osmotically downshifted cells were then spotted onto Luria–Bertani (LB) agar plates (Miller, 1972). Those from the high osmolarity grown cells were spotted onto LB agar plates with a total NaCl content of 0.3 M NaCl. Colony-forming units were determined after overnight incubation of the LB plates at 37°C.

Overproduction, purification and determination of the quaternary assembly of the ectoine synthase and of the ectoine hydroxylase

Cells of strain BL21, harbouring either the *ectC* expression plasmid pWN48 or the *ectD* expression plasmid pMP45 were grown in a minimal medium, and the overproduction of the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins was initiated by adding the inducer (anhydrotetracycline) of

the TetR repressor controlled *tet* promoter to the growth medium. Cleared cell extracts of the protein overproducing cultures were prepared and used to purify the (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins by affinity chromatography on streptactin affinity resin as detailed elsewhere (Hoeppner *et al*., 2014; Kobus *et al*., 2015). The purity of the recombinant (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins was assessed by SDS-polyacrylamide gel electrophoresis (12.5% and 15% respectively); the electrophoretically separated proteins were stained with Coomassie Brilliant Blue. To analyse the quaternary assembly of the ectoine synthase and ectoine hydroxylase, we performed size-exclusion chromatography. For the analysis of the ectoine synthase, a HiLoad 16/600 Superdex 75 pg column (GE Healthcare, München, Germany) was used, and the column was run in a buffer containing 20 mM TES (pH 7.0) and 150 mM NaCl. For the analysis of the ectoine hydroxylase, a HiLoad 16/600 Superdex 200 pg column (GE Healthcare, München, Germany) was used, and the column was run in a buffer containing 20 mM TES (pH 7.5) and 200 mM KCl.

Enzyme assays of the 'Ca. N. maritimus' SCM1 ectoine synthase and of the ectoine hydroxylase

High-performance liquid chromatography-based enzyme assays (Bursy *et al*., 2007; Widderich *et al*., 2014a) were used to assess the biochemical and kinetic properties of the affinity-purified (*Nm*)EctC-*Strep*-tag II and (*Nm*)EctD-*Strep*-tag II proteins. Single parameters (e.g. the salt concentrations) were changed to determine optimal conditions, and variations of the substrate concentration were used to assess the kinetic parameters of the EctC. The optimized assay buffer contained 20 mM TES (pH 7.0), 150 mM NaCl, 1 mM $FeSO_4$ and 10 mM Nγ-ADABA and was run for 15 min at 30°C. Similarly, the optimized assay buffer for the EctD contained 20 mM TES (pH 7.5), 200 mM KCl, 1 mM $FeSO_4$, 10 mM 2-oxoglutarate, and 6 mM ectoine and was run for 15 min at 35°C under vigorous shaking to assure aeration, since EctD is an oxygen-dependent enzyme. To determine the kinetic parameters of the (*Nm*)EctC and (*Nm*)EctD enzymes, the substrate concentration was varied for the EctC enzyme between 0 and 50 mM; for EctD, the substrate concentration of ectoine varied between 0 mM and 50 mM while keeping the concentration of the co-substrate 2-oxoglutarate constant at 10 mM. To determine the kinetic parameters for the co-substrate of EctD, the concentration was varied between 0 mM and 40 mM, while the concentration of the ectoine substrate was kept constant at 6 mM. Ectoine synthase and EctD enzyme assays were terminated by adding acetonitrile to a final concentration of 50% to the total enzyme assay solution. Denatured proteins were removed by centrifugation (5 min) in a table-top Eppendorf centrifuge, and portions of the supernatant were then used for HPLC analysis of the formed ectoine or 5-hydroxyectoine respectively.

HPLC analysis of ectoine and hydroxyectoine content

Ectoine and hydroxyectoine were detected by HPLC analysis using an Agilent 1260 Infinity LC system (Agilent, Waldbronn, Germany), a GROM-SIL Amino 1PR column (GROM, Rottenburg-Hailfingen, Germany) essentially as described (Kuhlmann and Bremer, 2002), with the exception that a 1260 Infinity Diode Array Detector (DAD) (Agilent) was employed, instead of the previously used UV/Vis detection system. The ectoine and hydroxyectoine content of samples was quantified using the OpenLAB software suite (Agilent). When samples of EctC or EctD enzyme activity assays were measured, 5 µl to 10 µl samples were injected into the system. To determine the ectoine and hydroxyectoine content of '*Ca*. N. maritimus' SCM1 cells, samples were lysed by re-suspending them in 20% ethanol for 1 h. Cellular debris was then removed by centrifugation (13 000 r.p.m. in a Sorval centrifuge at 4°C) for 20 min, and the supernatant was subsequently lyophylized to complete dryness. The resin was re-suspended in a mixture of water : acetonitrile (50:50 v/v), and samples were then injected into the HPLC system. Since cells of '*Ca*. N. maritimus' SCM1 grown at different salinities exhibit different cell sizes, the total protein content of the originally ethanolic extract were used for a protein assay (Pierce Protein Assay Kit; ThermoScientific, Schwerte, Germany), and these values were used to standardize the ectoine/hydroxyectoine content of the cell samples.

Mapping of the transcriptional organization of the ectoine/hydroxyectoine gene cluster

To assess the transcriptional organization of the *ect* operon, cells of '*Ca*. N. maritimus' SCM1 grown under enhanced osmotic stress conditions (the medium had an osmolarity of 1598 mOsm) were used to isolate total RNA. Cell lysis was achieved by re-suspension in 20% ethanol, and total RNA was isolated from these cell extracts by using the High Pure RNA Isolation Kit (Roche, Mannheim, Germany) according to the instructions of the user manual. Samples of RNA were further purified using the RNeasy Kit (Qiagen, Hilden, Germany) as described in the user manual and used for one-step RT-PCR assays. To analyse whether the *hypectABCD-mscS* genes are transcribed in a unit, four intergenic regions of the putative operon were amplified from isolated RNA using the Qiagen One Step RT-PCR Kit and custom synthesized DNA primers (MWG, Ebersberg, Germany). As controls, we also amplified DNA regions between genes that were divergently transcribed (Fig. 7A). To ensure that the formed PCR products did not result from DNA contaminations of RNA sample used for the RT-PCR reaction, an assay was performed in which total RNA was added after the reverse transcription step.

Assessment of the transcript levels of the 'Ca. N. maritimus' SCM1 ectA and mscS genes in response to varying osmolarities

For studying the expression of the ectoine biosynthesis gene cluster in *'Ca.* N. maritimus' SCM1 under different osmotic growth conditions, cells that had been cultured either at low (13 g NaCl L^{-1}; 220 mM), optimal (26 g NaCl L^{-1}; 440 mM) or high salt (39 g NaCl L^{-1}; 660 mM) were used. Total RNA was extracted from these cells using the peqGOLD TriFast Kit (VWR International GmbH, Erlangen, Germany) according to

the manufacturer's instructions. The extracted RNA solutions were treated with RNAse-free DNAse I (Life Technologies GmbH, Darmstadt, Germany) to remove residual chromosomal DNA, again following the manufacturer's instructions. The absence of DNA contamination was ascertained by PCR analysis. The relative abundance of the *hyp-ectABCD-mscS* mRNA to the mRNA of *amoA* (Nmar_1500) was determined by real-time PCR in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories GmbH, München, Germany) with the LightCycler RNA Master SYBR green I kit (Roche Diagnostics, Mannheim, Germany). Each reaction of the one-step RT-PCR was conducted in a 20-µl volume containing 125 ng template RNA, 0.5 µM of each primer, 3.25 mM $Mn(OAc)_2$ and 7.5 µl of LightCycler RNA Master SYBR green I. The following PCR primer sets were used: ectA_fwd (5′-TTAGAGAGCCTCGAGTTGATGACGC-3′) and ectA_rev (5′-GTCAAGAGGCTTGTTGTTTTGCACC-3′), mscS_fwd (5′-CGCAAAGGAACTATTCTCAAGCTGG-3′) and mscS_rev (5′-GCGAGAAATTGAAACAAGAACCTCG-3′), and amoA_fwd (5′-CCAAGTAGGTAAGTTCTATAA-3′) and amoA_rev (5′-AAGCGGCCATCCATCTGTA-3′) as described previously (Nakagawa and Stahl, 2013). The PCR cycling conditions were used as described in the manufacturer's instructions with denaturation at 95°C for 5 s, annealing at 60°C for 10 s and extension at 72°C for 5 s. The relative expression of each gene under the tested conditions was determined by using the *amoA* transcript level in *'Ca. N. maritimus' SCM1* cells grown under optimal salt concentrations as the standard. The level of the *amoA* transcript was set to one and those of *ectA* and *mscS* are expressed in relation to this reference transcript.

Database searches and phylogenetic analysis of EctC and EctD-type proteins

The amino acid sequence of the '*Ca*. N. maritimus' SCM1 EctC protein (Walker *et al*., 2010) was used as the template for BLAST searches of the microbial database of the JGI of the US Department of Energy (http://jgi.doe.gov/) (Nordberg *et al*., 2013). EctC-type amino acid sequences of closely related strains of the same species were removed, and the remaining 723 retrieved amino acid sequences were aligned using the MAFFT multiple amino acid sequence alignment server (http://mafft.cbrc.jp/alignment/server/) (Katoh and Standley, 2013). This curated data set was then used to construct a rooted phylogenetic tree of EctC-type sequences by employing the iTOL software suit (http://itol.embl.de/) (Letunic and Bork, 2011).

Modelling of three-dimensional protein structures and preparation of figures of crystal structures

The amino acid sequence of the '*Ca*. N. maritimus' SCM1 MscS protein was submitted via the website of the SWISS-Model server (http://swissmodel.expasy.org/) (Biasini *et al*., 2014) to generate an *in silico* model of the three-dimensional structure of this protein. The program automatically chose the monomer of the heptameric *E. coli* MscS open crystal structure (Wang *et al*., 2008) as the modelling template (PDB accession code: 2VV5). The model of the monomer of the (*Nm*)MscS protein generated via the SWISS-Model web server was visualized and compared with the *E. coli* MscS protein using resources provided by the PyMol suit (https://www.pymol.org/).

Acknowledgements

We thank Irina Bagyan for the kind gift of ectoines and Ian Booth and Samata Miller for generously providing bacterial strains. We greatly appreciate the expert technical assistance of Jochen Sohn and Jutta Gade in the course of this project and highly value the kind help of Vickie Koogle in the language editing of our manuscript. We thank Tamara Hoffmann for helpful discussions and critical reading of the manuscript. Funding for this study was provided by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 987 (to E.B. and J.H.) (Marburg), the Transregio SFB 51 (to J.S.D.) (Braunschweig/Oldenburg), The G.W. Leibniz Prize (to K.-U. Hinrichs; MARUM, Bremen) and through the LOEWE Program of the State of Hessen (via the Centre for Synthetic Microbiology; Synmicro; Marburg) (to J.H. and E.B.). N.W. and L.C. are recipients of PhD fellowships from the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-Mic, Marburg) and gratefully acknowledge its generous support.

References

Amann, E., Ochs, B., and Abel, K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69:** 301–315.

Bass, R.B., Strop, P., Barclay, M., and Rees, D.C. (2002) Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **298:** 1582–1587.

Bayer, B., Vojvoda, J., Offre, P., Alves, R.J., Elisabeth, N.H., Garcia, J.A., *et al*. (2015) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* in press. doi: 10.1038/ismej.2015.200

Becker, E.A., Seitzer, P.M., Tritt, A., Larsen, D., Krusor, M., Yao, A.I., *et al*. (2014) Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet* **10:** e1004784.

Becker, M., and Krämer, R. (2015) MscCG from *Corynebacterium glutamicum*: functional significance of the C-terminal domain. *Eur Biophys J* **44:** 577–588.

Becker, M., Borngen, K., Nomura, T., Battle, A.R., Marin, K., Martinac, B., and Krämer, R. (2013) Glutamate efflux mediated by *Corynebacterium glutamicum* MscCG, *Escherichia coli* MscS, and their derivatives. *Biochim Biophys Acta* **1828:** 1230–1240.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., *et al*. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42:** W252–W258.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6:** e16626.

Bolen, D.W., and Baskakov, I.V. (2001) The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol* **310:** 955–963.

Booth, I.R. (2014) Bacterial mechanosensitive channels: progress towards an understanding of their roles in cell physiology. *Curr Opin Microbiol* **18:** 16–22.

Booth, I.R., and Blount, P. (2012) The MscS and MscL families of mechanosensitive channels act as microbial emergency release valves. *J Bacteriol* **194:** 4802–4809.

Booth, I.R., Miller, S., Müller, A., and Lehtovirta-Morley, L. (2015) The evolution of bacterial mechanosensitive channels. *Cell Calcium* **57:** 140–150.

Börngen, K., Battle, A.R., Möker, N., Morbach, S., Marin, K., Martinac, B., and Krämer, R. (2010) The properties and contribution of the *Corynebacterium glutamicum* MscS variant to fine-tuning of osmotic adaptation. *Biochim Biophys Acta* **1798:** 2141–2149.

Bremer, E., and Krämer, R. (2000) Coping with osmotic challenges: osmoregulation through accumulation and release of compatible solutes. In *Bacterial Stress Responses*. Storz, G., and Hengge-Aronis, R. (eds). Washington DC, USA: ASM Press, pp. 79–97.

Brill, J., Hoffmann, T., Bleisteiner, M., and Bremer, E. (2011) Osmotically controlled synthesis of the compatible solute proline is critical for cellular defense of *Bacillus subtilis* against high osmolarity. *J Bacteriol* **193:** 5335–5346.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6:** 245–252.

Brown, A.D. (1976) Microbial water stress. *Bacteriol Rev* **40:** 803–846.

Bursy, J., Pierik, A.J., Pica, N., and Bremer, E. (2007) Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. *J Biol Chem* **282:** 31147–31155.

Bursy, J., Kuhlmann, A.U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A.J., and Bremer, E. (2008) Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. *Appl Environ Microbiol* **74:** 7286–7296.

Calderon, M.I., Vargas, C., Rojo, F., Iglesias-Guerra, F., Csonka, L.N., Ventosa, A., and Nieto, J.J. (2004) Complex regulation of the synthesis of the compatible solute ectoine in the halophilic bacterium *Chromohalobacter salexigens* DSM 3043^{T}. *Microbiology* **150:** 3051–3063.

Canovas, D., Vargas, C., Iglesias-Guerra, F., Csonka, L.N., Rhodes, D., Ventosa, A., and Nieto, J.J. (1997) Isolation and characterization of salt-sensitive mutants of the moderate halophile *Halomonas elongata* and cloning of the ectoine synthesis genes. *J Biol Chem* **272:** 25794–25801.

Coquelle, N., Talon, R., Juers, D.H., Girard, E., Kahn, R., and Madern, D. (2010) Gradual adaptive changes of a protein facing high salt concentrations. *J Mol Biol* **404:** 493–505.

da Costa, M.S., Santos, H., and Galinski, E.A. (1998) An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv Biochem Eng Biotechnol* **61:** 117–153.

Cruickshank, C.C., Minchin, R.F., Le Dain, A.C., and Martinac, B. (1997) Estimation of the pore size of the large-conductance mechanosensitive ion channel of *Escherichia coli*. *Biophys J* **73:** 1925–1931.

Csonka, L.N. (1989) Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53:** 121–147.

Deole, R., Challacombe, J., Raiford, D.W., and Hoff, W.D. (2013) An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. *J Biol Chem* **288:** 581–588.

Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F., and Lopez-Garcia, P. (2014) Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaeota and euryarchaeota. *Genome Biol Evol* **6:** 1549–1563.

Dong, C., Major, L.L., Srikannathasan, V., Errey, J.C., Giraud, M.F., Lam, J.S., *et al*. (2007) RmlC, a C3′ and C5′ carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation. *J Mol Biol* **365:** 146–159.

Dubendorff, J.W., and Studier, F.W. (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219:** 45–59.

Dunwell, J.M., Culham, A., Carter, C.E., Sosa-Aguirre, C.R., and Goodenough, P.W. (2001) Evolution of functional diversity in the cupin superfamily. *Trends Biochem Sci* **26:** 740–746.

Edwards, M.D., Black, S., Rasmussen, T., Rasmussen, A., Stokes, N.R., Stephen, T.L., *et al*. (2012) Characterization of three novel mechanosensitive channel activities in *Escherichia coli*. *Channels* **6:** 272–281.

Elling, F.J., Könneke, M., Lipp, J.S., Becker, K.W., Gagen, E.J., and Hinrichs, K.-U. (2014) Effect of growth phase on the membrane lipid composition of the thaumarchaeon *Nitrosopumilus martitimus* and their implication for archaeal lipid distribution in the marine environment. *Geochim Cosmochim Acta* **141:** 579–597.

Elling, F.J., Könneke, M., Mußamnn, M., Greve, A., and Hinrichs, K.-U. (2015) Influence of temperature, pH, and salinity on membrane lipid composition and Tex86 of marine planktonic thaumarchaealisolates. *Geochim Cosmochim Acta* in press. doi:10.1016/j.gca.2015.09.004

Empadinhas, N., and da Costa, M.S. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ Microbiol* **13:** 2056–2077.

Galinski, E.A., and Trüper, H.G. (1994) Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol Rev* **15:** 95–108.

Galinski, E.A., Pfeiffer, H.P., and Trüper, H.G. (1985) 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus *Ectothiorhodospira*. *Eur J Biochem* **149:** 135–139.

Garcia-Estepa, R., Argandona, M., Reina-Bueno, M., Capote, N., Iglesias-Guerra, F., Nieto, J.J., and Vargas, C. (2006) The *ectD* gene, which is involved in the synthesis of the compatible solute hydroxyectoine, is essential for thermoprotection of the halophilic bacterium *Chromohalobacter salexigens*. *J Bacteriol* **188:** 3774–3784.

Gubry-Rangin, C., Kratsch, C., Williams, T.A., McHardy, A.C., Embley, T.M., Prosser, J.I., and Macqueen, D.J. (2015)

Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proc Natl Acad Sci USA* **112:** 9370–9375.

Harishchandra, R.K., Wulff, S., Lentzen, G., Neuhaus, T., and Galla, H.J. (2010) The effect of compatible solute ectoines on the structural organization of lipid monolayer and bilayer membranes. *Biophys Chem* **150:** 37–46.

Haswell, E.S., Phillips, R., and Rees, D.C. (2011) Mechanosensitive channels: What can they do and how do they do it? *Structure* **19:** 1356–1369.

Hoeppner, A., Widderich, N., Bremer, E., and Smits, S.H.J. (2014) Overexpression, crystallization and preliminary X-ray crystallographic analysis of the ectoine hydroxylase from *Sphingopyxis alaskensis*. *Acta Cryst* **F70:** 493–496.

Hoffmann, T., Wensing, A., Brosius, M., Steil, L., Völker, U., and Bremer, E. (2013) Osmotic control of *opuA* expression in *Bacillus subtilis* and its modulation in response to intracellular glycine betaine and proline pools. *J Bacteriol* **195:** 510–522.

Höppner, A., Widderich, N., Lenders, M., Bremer, E., and Smits, S.H.J. (2014) Crystal structure of the ectoine hydroxylase, a snapshot of the active site. *J Biol Chem* **289:** 29570–29583.

Inbar, L., and Lapidot, A. (1988) The structure and biosynthesis of new tetrahydropyrimidine derivatives in actinomycin D producer *Streptomyces parvulus*. Use of ^{13}C- and ^{15}N-labeled L-glutamate and ^{13}C and ^{15}N NMR spectroscopy. *J Biol Chem* **263:** 16014–16022.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780.

Kelly, W.J., Leahy, S.C., Li, D., Perry, R., Lambie, S.C., Attwood, G.T., and Altermann, E. (2014) The complete genome sequence of the rumen methanogen *Methanobacterium formicicum* BRM9. *Stand Genomic Sci* **9:** 15.

Kempf, B., and Bremer, E. (1998) Uptake and synthesis of compatible solutes as microbial stress responses to high osmolarity environments. *Arch Microbiol* **170:** 319–330.

Klähn, S., and Hagemann, M. (2011) Compatible solute biosynthesis in cyanobacteria. *Environ Microbiol* **13:** 551–562.

Kloda, A., and Martinac, B. (2002) Mechanosensitive channels of bacteria and archaea share a common ancestral origin. *Eur Biophys J* **31:** 14–25.

Kobus, S., Widderich, N., Hoeppner, A., Bremer, E., and Smits, S.H.J. (2015) Overproduction, crystallization and X-ray diffraction data analysis of ectoine synthase from the cold-adapted marine bacterium *Sphingopyxis alaskensis*. *Acta Cryst* **F71:** 1027–1032.

Kol, S., Merlo, M.E., Scheltema, R.A., de Vries, M., Vonk, R.J., Kikkert, N.A., *et al*. (2010) Metabolomic characterization of the salt stress response in *Streptomyces coelicolor*. *Appl Environ Microbiol* **76:** 2574–2581.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543–546.

Könneke, M., Schubert, D.M., Brown, P.C., Hugler, M., Standfest, S., Schwander, T., *et al*. (2014) Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc Natl Acad Sci USA* **111:** 8239–8244.

Kuhlmann, A.U., and Bremer, E. (2002) Osmotically regulated synthesis of the compatible solute ectoine in *Bacillus pasteurii* and related Bacillus spp. *Appl Environ Microbiol* **68:** 772–783.

Kuhlmann, A.U., Bursy, J., Gimpel, S., Hoffmann, T., and Bremer, E. (2008) Synthesis of the compatible solute ectoine in *Virgibacillus pantothenticus* is triggered by high salinity and low growth temperature. *Appl Environ Microbiol* **74:** 4560–4563.

Kunte, H.J., Galinski, E.A., and Trüper, G.H. (1993) A modified FMOC-method for the detection of amino acid-type osmolytes and tetrahydropyrimidines (ectoines). *J Microbiol Meth* **17:** 129–136.

Kunte, H.J., Lentzen, G., and Galinski, E. (2014) Industrial production of the cell protectant ectoine: protection, mechanisms, processes, and products. *Curr Biotechnol* **3:** 10–25.

Kurz, M., Burch, A.Y., Seip, B., Lindow, S.E., and Gross, H. (2010) Genome-driven investigation of compatible solute biosynthesis pathways of *Pseudomonas syringae* pv. syringae and their contribution to water stress tolerance. *Appl Environ Microbiol* **76:** 5452–5462.

Letunic, I., and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39:** W475–W478.

Levina, N., Totemeyer, S., Stokes, N.R., Louis, P., Jones, M.A., and Booth, I.R. (1999) Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J* **18:** 1730–1737.

Lippert, K., and Galinski, E.A. (1992) Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl Microbiol Biotechnol* **37:** 61–65.

Lo, C.C., Bonner, C.A., Xie, G., D'Souza, M., and Jensen, R.A. (2009) Cohesion group approach for evolutionary analysis of aspartokinase, an enzyme that feeds a branched network of many biochemical pathways. *Microbiol Mol Biol Rev* **73:** 594–651.

Louis, P., and Galinski, E.A. (1997) Characterization of genes for the biosynthesis of the compatible solute ectoine from *Marinococcus halophilus* and osmoregulated expression in *Escherichia coli*. *Microbiology* **143:** 1141–1149.

Manzanera, M., Garcia de Castro, A., Tondervik, A., Rayner-Brandes, M., Strom, A.R., and Tunnacliffe, A. (2002) Hydroxyectoine is superior to trehalose for anhydrobiotic engineering of *Pseudomonas putida* KT2440. *Appl Environ Microbiol* **68:** 4328–4333.

Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461:** 976–979.

Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor New York: Cold Spring Harbor Laboratory.

Moritz, K.D., Amendt, B., Witt, E.M.H.J., and Galinski, E.A. (2015) The hydroxyectoine gene cluster of the non-halophilic acidophile *Acidiphilium cryptum*. *Extremophiles* **19:** 87–99.

Müller, V., Spanheimer, R., and Santos, H. (2005) Stress response by solute accumulation in archaea. *Curr Opin Microbiol* **8:** 729–736.

Naismith, J.H., and Booth, I.R. (2012) Bacterial mechanosensitive channels – MscS: Evolution's solution to creating sensitivity in function. *Annu Rev Biophys* **41:** 157–177.

Nakagawa, T., and Stahl, D.A. (2013) Transcriptional response of the archaeal ammonia oxidizer *Nitrosopumilus maritimus* to low and environmentally relevant ammonia concentrations. *Appl Environ Microbiol* **79:** 6911–6916.

Nakamura, J., Hirano, S., Ito, H., and Wachi, M. (2007) Mutations of the *Corynebacterium glutamicum* NCgl1221 gene, encoding a mechanosensitive channel homolog, induce L-glutamic acid production. *Appl Environ Microbiol* **73:** 4491–4498.

Nakayama, Y., Yoshimura, K., and Iida, H. (2013) Electrophysiological characterization of the mechanosensitive channel MscCG in *Corynebacterium glutamicum*. *Biophys J* **105:** 1366–1375.

Nelson-Sathi, S., Sousa, F.L., Roettger, M., Lozada-Chavez, N., Thiergart, T., Janssen, A., *et al*. (2015) Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517:** 77–80.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., *et al*. (2013) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* **42:** D26–D31.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299–304.

Offre, P., Spang, A., and Schleper, C. (2013) Archaea in biogeochemical cycles. *Annu Rev Microbiol* **67:** 437–457.

Ono, H., Okuda, M., Tongpim, S., Imai, K., Shinmyo, A., Sakuda, S., *et al*. (1998) Accumulation of compatible solutes, ectoine and hydroxyectoine, in a moderate halophile, *Halomonas elongata* KS3, isolated from dry salty land in Thailand. *J Ferment Bioeng* **85:** 362–368.

Ono, H., Sawada, K., Khunajakr, N., Tao, T., Yamamoto, M., Hiramoto, M., *et al*. (1999a) Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, *Halomonas elongata*. *J Bacteriol* **181:** 91–99.

Ono, H., Sawada, K., Khunajakr, N., Tao, T., Yamamoto, M., Hiramoto, M., *et al*. (1999b) Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, *Halomonas elongata*. *J Bacteriol* **181:** 91–99.

Oren, A. (2011) Thermodynamic limits to microbial life at high salt concentrations. *Environ Microbiol* **13:** 1908–1923.

Oren, A. (2013) Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front Microbiol* **4:** 315.

Pastor, J.M., Salvador, M., Argandona, M., Bernal, V., Reina-Bueno, M., Csonka, L.N., *et al*. (2010) Ectoines in cell stress protection: uses and biotechnological production. *Biotechnol Adv* **28:** 782–801.

Pester, M., Schleper, C., and Wagner, M. (2011) The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14:** 300–306.

Peters, P., Galinski, E.A., and Trüper, H.G. (1990) The biosyntheis of ectoine. *FEMS Microbiol Lett* **71:** 157–162.

Pivetti, C.D., Yen, M.R., Miller, S., Busch, W., Tseng, Y.H., Booth, I.R., and Saier, M.H., Jr (2003) Two families of mechanosensitive channel proteins. *Microbiol Mol Biol Rev* **67:** 66–85.

Prabhu, J., Schauwecker, F., Grammel, N., Keller, U., and Bernhard, M. (2004) Functional expression of the ectoine hydroxylase gene (*thpD*) from *Streptomyces chrysomallus* in *Halomonas elongata*. *Appl Environ Microbiol* **70:** 3130–3132.

Record, M.T., Jr, Courtenay, E.S., Cayley, D.S., and Guttman, H.J. (1998) Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem Sci* **23:** 143–148.

Roberts, M.F. (2004) Osmoadaptation and osmoregulation in archaea: update 2004. *Front Biosci* **9:** 1999–2019.

Roesser, M., and Müller, V. (2001) Osmoadaptation in bacteria and archaea: common principles and differences. *Environ Microbiol* **3:** 743–754.

Rowe, I., Anishkin, A., Kamaraju, K., Yoshimura, K., and Sukharev, S. (2014) The cytoplasmic cage domain of the mechanosensitive channel MscS is a sensor of macromolecular crowding. *J Gen Physiol* **143:** 543–557.

Salvador, M., Argandona, M., Pastor, J.M., Bernal, V., Canovas, M., Csonka, L.N., *et al*. (2015) Contribution of RpoS to metabolic efficiency and ectoines synthesis during the osmo- and heat-stress response in the halophilic bacterium *Chromohalobacter salexigens*. *Environ Microbiol Rep* **7:** 301–311.

Santoro, A.E., Francis, C.A., de Sieyes, N.R., and Boehm, A.B. (2008) Shifts in the relative abundance of ammonia-oxidizing bacteria and archaea across physicochemical gradients in a subterranean estuary. *Environ Microbiol* **10:** 1068–1079.

Saum, S.H., and Müller, V. (2008) Growth phase-dependent switch in osmolyte strategy in a moderate halophile: ectoine is a minor osmolyte but major stationary phase solute in *Halobacillus halophilus*. *Environ Microbiol* **10:** 716–726.

Schwibbert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., *et al*. (2011) A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581^{T}. *Environ Microbiol* **13:** 1973–1994.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015) Horizontal gene transfer: building the web of life. *Nat Rev Genet* **16:** 472–482.

Stahl, D.A., and de la Torre, J.R. (2012) Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol* **66:** 83–101.

Stieglmeier, M., Alves, R.J.E., and Schleper, C. (2014) The phylum Thaumarchaeota. In *The Prokaryotes – Other Major Lineages of Bacteria and the Archaea*. Rosenberg, E.I. (ed.). Berlin, Germany: Springer, pp. 347–362.

Storz, G., and Hengge-Aronis, R. (2000) *Bacterial Stress Responses*. Washington, DC, USA: ASM Press.

Stöveken, N., Pittelkow, M., Sinner, T., Jensen, R.A., Heider, J., and Bremer, E. (2011) A specialized aspartokinase enhances the biosynthesis of the osmoprotectants ectoine

and hydroxyectoine in *Pseudomonas stutzeri* A1501. *J Bacteriol* **193:** 4456–4468.
Strickland, J.D.H., and Parson, T.R. (1972) *A Practical Handbook of Seawater Analysis*. Ottawa, Canada: Fisheries Research Board of Canada.
Talon, R., Coquelle, N., Madern, D., and Girard, E. (2014) An experimental point of view on hydration/solvation in halophilic proteins. *Front Microbiol* **5:** 66.
Tanne, C., Golovina, E.A., Hoekstra, F.A., Meffert, A., and Galinski, E.A. (2014) Glass-forming property of hydroxyectoine is the cause of its superior function as a dessication protectant. *Front Microbiol* **5:** 150.
Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., *et al*. (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107:** 8818–8823.
Wang, W., Black, S.S., Edwards, M.D., Miller, S., Morrison, E.L., Bartlett, W., *et al*. (2008) The structure of an open form of an *E. coli* mechanosensitive channel at 3.45 A resolution. *Science* **321:** 1179–1183.
Widderich, N., Höppner, A., Pittelkow, M., Heider, J., Smits, S.H.J., and Bremer, E. (2014a) Biochemical properties of ectoine hydroxylases from extremophiles and their wider taxonomic distribution among microorganisms. *PLoS ONE* **9:** e93809.
Widderich, N., Pittelkow, M., Höppner, A., Mulnaes, D., Buckel, W., Gohlke, H., *et al*. (2014b) Molecular dynamics simulations and structure-guided mutagenesis provide insight into the architecture of the catalytic core of the ectoine hydroxylase. *J Mol Biol* **426:** 586–600.
Wilson, M.E., Maksaev, G., and Haswell, E.S. (2013) MscS-like mechanosensitive channels in plants and microbes. *Biochemistry* **52:** 5708–5722.
Wood, J.M. (2011) Bacterial osmoregulation: a paradigm for the study of cellular homeostasis. *Annu Rev Microbiol* **65:** 215–238.
Wood, J.M., Bremer, E., Csonka, L.N., Krämer, R., Poolman, B., van der Heide, T., and Smith, L.T. (2001) Osmosensing and osmoregulatory compatible solute accumulation by bacteria. *Comp Biochem Physiol A Mol Integr Physiol* **130:** 437–460.
Youssef, N.H., Savage-Ashlock, K.N., McCully, A.L., Luedtke, B., Shaw, E.I., Hoff, W.D., and Elshahed, M.S. (2014) Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the *Halobacteriales*. *ISME J* **8:** 636–649.
Zaparty, M., Hagemann, A., Brasen, C., Hensel, R., Lupas, A.N., Brinkmann, H., and Siebers, B. (2013) The first prokaryotic trehalose synthase complex identified in the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *PLoS ONE* **8:** e61354.
Zhang, Y., Chen, L., Dai, T., Tian, J., and Wen, D. (2015) The influence of salinity on the abundance, transcriptional activity, and diversity of AOA and AOB in an estuarine sediment: a microcosm study. *Appl Microbiol Biotechnol* **99:** 9825–9833.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig. S1. Phylogenetic tree of EctC-type proteins. Based on an amino acid sequence alignment of 723 EctC-type proteins, a rooted phylogenetic tree was constructed with the ITOL program (Letunic and Bork, 2011). The three regions in the phylogenetic tree populated by archaeal EctC proteins are highlighted by red arrowheads. The position of those 24 EctC-type proteins that originate from bacteria that lack identifiable *ectAB* genes are marked above the phylogenetic tree by a black bar.
Fig. S2. Alignment of the amino acid sequence of the *E. coli* MscS protein with those MscS-type proteins whose corresponding genes are located next to ectoine/hydroxyectoine biosynthetic gene clusters in the genomes of '*Ca.* Nitrosopumilus maritimus' SCM 1 and '*Ca.* Nistrosopumilus' sp. AR2 (see Fig. 1B in the main text). Red bars indicate the membrane-spanning segments of the *E. coli* MscS protein.
Table S1. Amino acid sequence relatedness of the ectoine biosynthetic proteins from '*Ca*. Nitrosopumilus' strains to other ectoine biosynthetic enzymes.
Table S2. Distribution of mechanosensitive channels in potential ectoine-producing *Archaea*.