Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1778 (2008) 963–973
Biochimica et Biophysica Acta BBA
www.elsevier.com/locate/bbamem

Transmembrane topology of the Acr3 family arsenite transporter from *Bacillus subtilis*

Emil K.J. Aaltonen, Maria Silow *

Department of Cell and Organism Biology, Lund University, Sölvegatan 35, S-223 62 Lund, Sweden

Received 10 August 2007; received in revised form 14 November 2007; accepted 16 November 2007
Available online 26 December 2007

Abstract

The transmembrane topology of the Acr3 family arsenite transporter Acr3 from *Bacillus subtilis* was analysed experimentally using translational fusions with alkaline phosphatase and green fluorescent protein and *in silico* by topology modelling. Initial topology prediction resulted in two models with 9 and 10 TM helices respectively. 32 fusion constructs were made between truncated forms of *acr3* and the reporter genes at 17 different sites throughout the *acr3* sequence to discriminate between these models. Nine strong reporter protein signals provided information about the majority of the locations of the cytoplasmic and extracellular loops of Acr3 and showed that both the N- and the C-termini are located in the cytoplasm. Two ambiguous data points indicated the possibility of an alternative 8 helix topology. This possibility was investigated using another 10 fusion variants, but no experimental support for the 8 TM topology was obtained. We therefore conclude that Acr3 has 10 transmembrane helices. Overall, the loops which connect the membrane spanning segments are short, with cytoplasmic loops being somewhat longer than the extracellular loops. The study provides the first ever experimentally derived structural information on a protein of the Acr3 family which constitutes one of the largest classes of arsenite transporters.

Keywords: Arsenite; Transmembrane topology; Acr3; Translational fusion; Alkaline phosphatase; GFP

1. Introduction

Arsenic is a ubiquitous, non-essential semi metal which is toxic in all forms. Resistance systems towards arsenic therefore evolved very early and exist in all domains of life. In Bacteria, Archaea and Fungi four different types of transmembrane channels or transporters have been identified that constitute the centrepieces of the arsenic resistance machineries [1–4]. In *Escherichia coli* arsenite efflux is mediated by ArsB, in *Sinorhizobium meliloti* arsenite is excreted via the Major Intrinsic Protein AqpS [4] and in *Saccharomyces cerevisiae* detoxification is achieved by arsenite efflux through Acr3p or through compartmentalization of glutathione–arsenic adducts into the vacuole by Ycf1p [1]. The only arsenic transporter which has been extensively studied in molecular detail is the ArsB protein from *E. coli*. Hence, it is known that ArsB has 429 amino acid residues and belongs to the Major Facilitator superfamily of transporters [5]. In the absence of its complex partner, ArsA, the driving force for arsenite pumping in ArsB is the electrochemical gradient across the membrane. When ArsA is present, it hydrolyses ATP and thereby provides the energy to sustain arsenite transport out of the cell [6]. Similar to other Major Facilitator transporters [7], ArsB has a membrane topology with 12 transmembrane (TM) helices [8].

Arsenite efflux in *Bacillus subtilis* is mediated mainly by a membrane spanning transporter (forthwith denoted Acr3) which is more closely related to the Acr3 family of arsenic transporters found in *S. cerevisiae* than to the bacterial ArsB from *E. coli* [3]. *B. subtilis* Acr3 is smaller than ArsB and contains only 346 amino acid residues [3]. On the sequence level ArsB and Acr3 are unrelated. In a recent study, Billard and co-workers determined the prevalence of different types of bacterial arsenite transporters in samples from arsenic contaminated and non-contaminated soil [9]. The study showed that, in contrast to what was previously thought, arsenite transporters of

Abbreviations: PhoA, Alkaline Phosphatase from *E. coli*; GFP, Green Fluorescent Protein; TM, Transmembrane; SCAM, Substituted Cysteine Accessibility Method

* Corresponding author. Tel.: +46 46 2223480; fax: +46 46 2224113.
E-mail address: maria.berggard_silow@cob.lu.se (M. Silow).

0005-2736/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2007.11.011

the Acr3 family appear to be more common than those of the ArsB type. This observation is paralleled also by data from sequencing of bacterial genomes which suggest a more widespread distribution of the Acr3 proteins than of the ArsB proteins. To date however, no mechanistic data is available on any member of the Acr3 family. In order to be able to rationally study the function of a transport protein it is important to have a basic understanding of the structure of the transporter. One way to obtain this information is to determine the structure by X-ray crystallography. Unfortunately, this requires the presence of well diffracting crystals which are notoriously difficult to obtain for membrane proteins. Another way to gain basic structural information of a membrane protein is to analyse the transmembrane topology. This can be done by constructing a series of translational fusions between truncated forms of the membrane protein and reporter proteins which are active only at one specific side of the membrane [10]. In this study we have determined the topology of *B. subtilis* Acr3 by constructing a total of 42 translational fusions between truncated forms of Acr3 and the reporter proteins alkaline phosphatase (PhoA) and Green Fluorescent Protein (GFP) at 22 different points in the amino acid sequence of the protein. The results of the study provide the first experimental information to date on the structure of an Acr3 family arsenite transporter. Based on the experimental data we propose a common topological model for Acr3 proteins from different organisms which will serve as a basis for subsequent mechanistic analysis of arsenite transport by these proteins.

2. Materials and methods

2.1. Bacterial strains/plasmids and primers

Primers used are listed in Table 1. The bacterial strains and plasmids used in this study are listed in Table 2.

2.2. Growth conditions

All liquid cultures were grown at 200 rpm and 37 °C in baffled E-flasks. Colonies on agar plates were grown over night at 37 °C. *E. coli* cells were grown either in LB medium (Luria–Bertani medium) or on LA (Luria–Bertani Agar) plates [11]. Except when fusion gene expression was desired, culture media were supplemented with 0.2% (w/v) glucose. Antibiotics used were kanamycin (Km, 30 μg/ml) ampicillin (Ap, 100 μg/ml) and tetracycline (Tc, 15 μg/ml).

2.3. DNA techniques

DNA restriction and ligation, agarose gel electrophoresis, PCR and electroporation of *E. coli* cells were performed according to standard procedures [11]. Isolation of plasmid DNA from *E. coli* was done by CsCl density gradient centrifugation [12] or with plasmid miniprep kits from BioRad or OMEGA Bio-Tek. DNA sequencing was performed by MWG Biotech, Germany.

2.4. Construction of fusion plasmids

Truncated fragments of *acr3* were amplified by PCR from *B. subtilis* 1A1 chromosomal DNA. The gene encoding Green Fluorescent Protein mut3a (*GFPmut3a*) was amplified by PCR from plasmid pAD123 using the GFP_up/GFP_down primer pair. The modified *phoA* gene encoding *E. coli* alkaline phosphatase without the segment encoding the N-terminal export signal sequence was amplified from pPHOA using the ephoA_up/vphoA_down primer pair. The resulting truncated *acr3* fragments, the *GFPmut3a* fragment and the *phoA* fragment respectively, were inserted into pCR®-Blunt II-TOPO® and transformed into *E. coli* TOP10 selecting for kanamycin resistance. All truncated *acr3* variants were cut from the respective pCR®-Blunt II-TOPO® derivative using NcoI and XhoI and ligated into pBAD His B digested with the same enzymes and transformed into TOP10 cells selecting for ampicillin resistance. The resulting plasmids were named pBAD-X where X indicates the last codon of the truncated *acr3* fragments. *phoA* was cut out from pTOPO-P using XhoI and KpnI and ligated into all pBAD-X plasmids digested with the same enzymes. The resulting plasmids were named pBAD-X-P (Table 2). *GFPmut3a* was cut out from pTOPO-G using XhoI and KpnI and ligated into all pBAD-X plasmids digested with the same enzymes. The resulting plasmids were named pBAD-X-G (Table 2). All the constructed gene fusions in pBAD-X-P and pBAD-X-G were confirmed by DNA sequencing using the v184up and arsBint primers. The pBAD-X-P and pBAD-X-G plasmids and pBAD His B were transformed into *E. coli* LMG194 cells which are deleted for *phoA*.

Table 1
Primers used in this work

Primer	Sequence (5′ to 3′)	Restriction sites
ephoA_up	ATACTCGAGCCGGACACCAGAAATG	XhoI
vphoa_down	ATTGGTACCTTATTTCAGCCCCAG	KpnI
GFP_up	ATACTCGAGCATGAGTAAAGGAGAAG	XhoI
GFP_down	ATTGGTACCCCTACTCAAGCTTGCATG	KpnI
arsBint	ATTGCAGGAATGGTGACACG	
V184up	CGGGCCATGGAACGTTTATCT	NcoI
T38	TTTACCTCGAGTGTAATTTGTTTAATCC	XhoI
T77	ATCGCCTCGAGAACACTTTAATATCTTT	XhoI
T104	ATTATCTCGAGGGTTTATCTGGCAAAAA	XhoI
T139	TATTACTCGAGAATCCTGCCGCGTATTC	XhoI
T149	ATCGCCTCGAGAACATTTGGAAAATTGAGTT	XhoI
T177	ATCTTCTCGAGATCGTAATATTCACTAC	XhoI
T189	ATCGCCTCGAGAAAAGGTAAATGAATACTGA	XhoI
T199	ATCGCCTCGAGAATGTCACCATTCCTGCAAT	XhoI
T204	ATCGCCTCGAGAATACAAAAATATAACGTG	XhoI
T212	ATCGCCTCGAGAAATACCATTCTTTGCCCTT	XhoI
T215	ATCTTCTCGAGAACACTTTTTCATACCA	XhoI
T240	ATCGCCTCGAGAAACCTTTTAAGGAAAACAT	XhoI
T250	ATCTTCTCGAGACAACATCTAACGGCA	XhoI
T252	ATCTTCTCGAGACTCGGACAACATCTAA	XhoI
T273	ATCTCCTCGAGTTCTTCCCTAAAAAGAA	XhoI
T279	ATCGCCTCGAGAAATAGTTTGCCCCGATTTT	XhoI
T284	ATCTCCTCGAGAAAGTAGTAGTTACAGC	XhoI
T296	TAAATCTCGAGGCTAATTCAAAGTTAT	XhoI
T307	ATCGCCTCGAGAAAATTCCAAATACACCAAC	XhoI
T311	TAAATCTCGAGGCTGCACCCGAATGAT	XhoI
T327	ATTTACTCGAGGCTATCATGACTGGCAC	XhoI
T346	ATTTACTCGAGATCGAGTGACTTCC	XhoI

2.5. Analysis of alkaline phosphatase activity and GFP fluorescence intensity

E. coli LMG194 containing pBAD His B, pBAD-X-P or pBAD-X-G plasmids were cultured in LB with ampicillin (100 μg/ml) and gene expression was induced at OD_{600} ~ 0.7–0.8 by addition of L-arabinose to the growth medium (optimal final concentration 0.2%). The induced cultures were harvested after 2 h by centrifugation for 10 min at 4500 rpm in a Labor-M bench top centrifuge (Wifug). Cells were resuspended in PhoA-buffer (10 mM Tris/HCl, pH 8.0, 0.1 M NaCl) or GFP-buffer (50 mM Tris/HCl, pH 8.0, 200 mM NaCl, 15 mM EDTA) and the optical density of the suspensions was adjusted to OD_{600} = 1. Two aliquots of 1 ml cell suspension were pelleted by centrifugation and either frozen for Western blot analysis at a later time or resuspended in 500 μl PhoA- or GFP-buffer and used for measurements. PhoA

Table 2
Bacterial strains and plasmids used in this work

Strains and plasmids	Relevant characteristics[a]	Source or reference
Strains		
Escherichia coli		
TOP10	*araD*139 Δ(*araA-leu)7697*	Invitrogen
LMG194	Δ*phoA* Δ*ara714* Tc^r	Invitrogen
Plasmids		
pCR®-Blunt II-TOPO®	Km^r	Invitrogen
pPHOA	Vector encoding *E. coli phoA*	[52]
pAD123	Vector encoding *GFPmut3A*	[53]
pBAD His B	Ap^r	Invitrogen
pTOPO-P	pCR®-Blunt II-TOPO® containing *E. coli phoA*	This work
pTOPO-G	pCR®-Blunt II-TOPO® containing *GFPmut3A*	This work
pBAD-38-P	pBAD His B with *phoA* fused to codon 38 of *Acr3*	This work
pBAD-77-P	pBAD His B with *phoA* fused to codon 77 of *Acr3*	This work
pBAD-104-P	pBAD His B with *phoA* fused to codon 104 of *Acr3*	This work
pBAD-139-P	pBAD His B with *phoA* fused to codon 139 of *Acr3*	This work
pBAD-149-P	pBAD His B with *phoA* fused to codon 149 of *Acr3*	This work
pBAD-177-P	pBAD His B with *phoA* fused to codon 177 of *Acr3*	This work
pBAD-189-P	pBAD His B with *phoA* fused to codon 189 of *Acr3*	This work
pBAD-199-P	pBAD His B with *phoA* fused to codon 199 of *Acr3*	This work
pBAD-204-P	pBAD His B with *phoA* fused to codon 204 of *Acr3*	This work
pBAD-212-P	pBAD His B with *phoA* fused to codon 212 of *Acr3*	This work
pBAD-215-P	pBAD His B with *phoA* fused to codon 215 of *Acr3*	This work
pBAD-250-P	pBAD His B with *phoA* fused to codon 250 of *Acr3*	This work
pBAD-252-P	pBAD His B with *phoA* fused to codon 252 of *Acr3*	This work
pBAD-273-P	pBAD His B with *phoA* fused to codon 273 of *Acr3*	This work
pBAD-279-P	pBAD His B with *phoA* fused to codon 279 of *Acr3*	This work
pBAD-284-P	pBAD His B with *phoA* fused to codon 284 of *Acr3*	This work
pBAD-307-P	pBAD His B with *phoA* fused to codon 307 of *Acr3*	This work
pBAD-311-P	pBAD His B with *phoA* fused to codon 311 of *Acr3*	This work
pBAD-327-P	pBAD His B with *phoA* fused to codon 327 of *Acr3*	This work
pBAD-346-P	pBAD His B with *phoA* fused to codon 346 of *Acr3*	This work
pBAD-38-G	pBAD His B with *Gfpmut3a* fused to codon 38 of *Acr3*	This work
pBAD-77-G	pBAD His B with *Gfpmut3a* fused to codon 77 of *Acr3*	This work
pBAD-104-G	pBAD His B with *Gfpmut3a* fused to codon 104 of *Acr3*	This work
pBAD-139-G	pBAD His B with *Gfpmut3a* fused to codon 139 of *Acr3*	This work
pBAD-149-G	pBAD His B with *Gfpmut3a* fused to codon 149 of *Acr3*	This work
pBAD-177-G	pBAD His B with *Gfpmut3a* fused to codon 177 of *Acr3*	This work
pBAD-189-G	pBAD His B with *Gfpmut3a* fused to codon 189 of *Acr3*	This work
pBAD-199-G	pBAD His B with *Gfpmut3a* fused to codon 199 of *Acr3*	This work
pBAD-204-G	pBAD His B with *Gfpmut3a* fused to codon 204 of *Acr3*	This work
pBAD-212-G	pBAD His B with *Gfpmut3a* fused to codon 212 of *Acr3*	This work
pBAD-215-G	pBAD His B with *Gfpmut3a* fused to codon 215 of *Acr3*	This work
pBAD-240-G	pBAD His B with *Gfpmut3a* fused to codon 240 of *Acr3*	This work
pBAD-250-G	pBAD His B with *Gfpmut3a* fused to codon 250 of *Acr3*	This work
pBAD-252-G	pBAD His B with *Gfpmut3a* fused to codon 252 of *Acr3*	This work
pBAD-273-G	pBAD His B with *Gfpmut3a* fused to codon 273 of *Acr3*	This work
pBAD-279-G	pBAD His B with *Gfpmut3a* fused to codon 279 of *Acr3*	This work
pBAD-284-G	pBAD His B with *Gfpmut3a* fused to codon 284 of *Acr3*	This work
pBAD-296-G	pBAD His B with *Gfpmut3a* fused to codon 296 of *Acr3*	This work
pBAD-307-G	pBAD His B with *Gfpmut3a* fused to codon 307 of *Acr3*	This work
pBAD-311-G	pBAD His B with *Gfpmut3a* fused to codon 311 of *Acr3*	This work
pBAD-327-G	pBAD His B with *Gfpmut3a* fused to codon 327 of *Acr3*	This work
pBAD-346-G	pBAD His B with *Gfpmut3a* fused to codon 346 of *Acr3*	This work

[a] Ap^r, Km^r and Tc^r indicate resistance to ampicillin, kanamycin and tetracycline, respectively.

activity was measured in permeabilized cells using a colorimetric assay with *p*-nitrophenyl phosphate as described previously [13]. The linear increase in optical density at 420 nm as a function of time was recorded and PhoA activity calculated as μmol substrate hydrolysed per minute. GFP fluorescence was measured essentially as described [14] in a FLUOstar microtiter plate reader (Corell 96 well plate, exitation filter 485 nm, emission 520 nm, 28 °C). Fluorescence intensities are given in arbitrary units (A.U.).

2.6. Isolation of membranes

E. coli LMG194 containing pBAD His B, pBAD-X-P or pBAD-X-G plasmids were cultured and gene expression induced as above. Before harvest, cells were cooled in an ice bath for 15 min. With exception of the incubation at 30 °C, isolation of membranes was then done as described in [15]. Protein concentrations were determined using the BCA protein assay kit (Pierce) with

bovine serum albumin as standard. The membranes were diluted to a protein concentration of 175 ng/μl or diluted further as indicated. After mixing 1:1 with 2× SDS-loading buffer, 15 μl of each sample was analyzed by SDS-PAGE and Western blot as described below. Positive controls were prepared as described below.

2.7. Detection of fusion proteins

Frozen cell pellets were thawed on ice and resuspended in 2.5 ml PhoA- or 500 μl GFP-buffer and mixed 1:1 with 2× SDS-loading buffer and incubated at 40 °C for 30 min. As positive controls, *E. coli* alkaline phosphatase (GE Healthcare) and the soluble fraction of a cell extract from *E. coli* TOP10 cells producing GFPuv from pGFPuv (Clontech laboratories) were incubated with SDS-sample buffer at 95 °C for 5 min. 15 μl of all samples were subjected to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using the Schägger and von Jagow system [16]. For Western blot, proteins were transferred from the gel to a polyvinylidene fluoride blotting membrane (Millipore) by wet blot as described before [17]. For detection of PhoA antigen, rabbit antiserum against *E. coli* alkaline phosphatase was used as the primary antibody [13]. GFP was detected using a primary rabbit polyclonal antibody raised against green fluorescent protein (Invitrogen). The secondary antibody was horseradish-peroxidase-conjugated donkey-anti-rabbit antibody from the ECL Western blotting analysis kit (GE Healthcare). Super Signal reagent (Pierce) and a Kodak Image Station was used for detection of bound secondary antibody.

2.8. Prediction methods

The amino acid sequence for *B. subtilis* Acr3 (GI:2635024) was downloaded from GenBank (http://ncbi.nlm.nih.gov). The most current web-versions of ten different topology prediction methods were used to model the topology of Acr3: DAS (http://www.sbc.su.se/~miklos/DAS/) [18], HMMTOP (http://www.enzim.hu/hmmtop/) [19], MEMSAT (http://bioinf.cs.ucl.ac.uk/psipred/) [20], PHD (http://cubic.bioc.columbia.edu/predictprotein/predictprotein.html) [21], SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiframe0.html) [22], TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) [23], TMMod (http://liao.cis.udel.edu/website/servers/TMMOD/scripts/frame.php?p=description) [24], TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) [25], Toppred (http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html) [26]. All methods were used in single sequence mode and all user adjustable parameters were left at their default values. Modelling with constraints was done using HMMTOP [19] or TMHMMFix (http://www.sbc.su.se/~melen/TMHMMfix/) [27]. Modelling of aligned sequences was done using HMMTOP [19]. Modelling of the distance of sequence segments to the membrane centre was done with Z-pred (http://www.cbr.su.se/zpred/) using default settings [28]. The amino acid sequences of the five additional documented and putative Acr3 family transporters were downloaded from GenBank; *Saccharomyces cerevisiae* (GI:2498103), *Mycobacterium tuberculosis* H37Rv (GI:15609780), *Streptomyces coelicolor* A3(2) (GI:6900906), *Haemophilus influenzae* R2846 (GI:53732906) and *Nostoc punctiforme* PCC 73102 (GI:23126620). Alignment was done with the MEGA 3.1 software using the Blosum protein weight matrix with default settings [29].

3. Results

3.1. Topology modelling of B. subtilis Acr3

The topology of Acr3 was modelled using nine different prediction methods. All modelling programs predicted the amino terminus to have a cytoplasmic location (Table 3) and displayed only very slight variations for the first 300 amino acid residues. This segment was predicted by all methods to contain eight TM helices. For the last 46 residues, however, the modelling programs gave different results. Six modelling programs predicted 2 TM segments and an overall topology with 10 TM helices (Fig. 1A). In contrast, three programs predicted one TM segment and a large cytoplasmic loop in the last 46 residues, resulting in a 9 TM helix topology (Fig. 1B). In topology modelling, different prediction algorithms give differences in starting- and end points for predicted TM segments depending on e.g. the hydrophobicity scales, structural states etc. that are used in the calculations [18–26]. Due to these differences, if several prediction programs are used, the modelling results must somehow be compiled to serve as a basis for experimental design. Compilation can be done for each helix by taking the average predicted starting point and average predicted end point from prediction models which give overlapping results. The resulting "consensus helices" are then taken to represent the trans-membrane segments. It should be noted that the actual helices spanning the membrane may deviate significantly from the consensus helices if overlap in the predictions is low.

Table 3
Results from topology predictions of *B. subtilis* Acr3 using different prediction programs

Prediction server	Orientation		Predicted number of helices
	N-terminus	C-terminus	
DAS	Nd[a]	Nd[a]	10
HMMTOP	Cytoplasmic	Cytoplasmic	10
MEMSAT	Cytoplasmic	Cytoplasmic	10
PHD	Cytoplasmic	Cytoplasmic	10
SOSUI	Cytoplasmic	Extracellular	9
TMHMM	Cytoplasmic	Cytoplasmic	10
TMMod	Cytoplasmic	Cytoplasmic	10
TMpred	Cytoplasmic	Extracellular	9
Top pred	Cytoplasmic	Extracellular	9

[a]Not determined.

By comparing the results from the topology predictions of Acr3 we obtained 8 consensus helices in the first 300 amino acid residues and one or two consensus helices in the last 46 amino acid residues assuming a 9 or 10 helix topology respectively (Fig. 1A and B). The consensus helices were between 20 and 24 amino acids in length (Table 4). On the basis of these results we selected points in Acr3 to be analysed by translational fusions both with PhoA and GFP. PhoA contains two disulfide bonds which are essential for the activity of the enzyme. The rationale behind using PhoA as a probe of membrane protein topology is that the disulfide bond only forms if the PhoA part of the fusion protein is exported into the periplasm [30,31]. Hence, PhoA fusion points in loops located on the extracellular side of the membrane will display alkaline phosphatase activity whereas cytoplasmic fusion points will not. GFP has recently been demonstrated to function as an efficient reporter for cytoplasmic location in membrane protein topology studies in *E. coli* as it is only fluorescent when present on the cytoplasmic side of the membrane [10,32–34]. The use of both reporters in our experiment would thus provide complementary data for each fusion point. The fusion sites were chosen towards the ends of predicted cytoplasmic and extracellular loops and were all

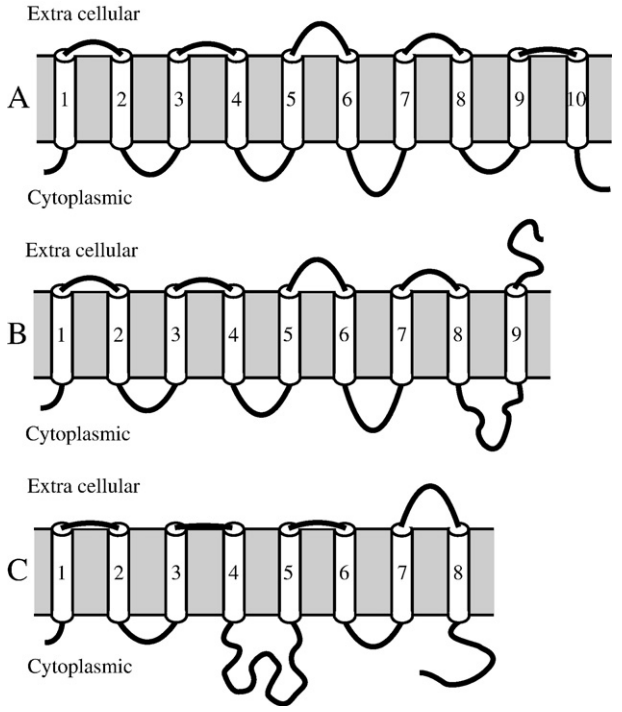

Fig. 1. Predicted alternative topologies for *B. subtilis* Acr3. Initial modelling of the transmembrane topology of Acr3 resulted in two different models with 10 and 9 TM helices respectively. (A) In the 10 TM model all loops are relatively short. The longest loop, between helices 6 and 7, is 26 amino acid residues long. (B) The topology of the nine helix model is identical to the 10 TM model for the first 300 amino acid residues. In the last 46 amino acid residues, however, it has only one TM helix and a 34 residue long intracellular loop between helices 8 and 9. (C) Constrained modelling resulted in an 8 helix topology when fusion points 177 and 240 were set to cytoplasmic locations. A striking difference in the 8 TM model is the presence of a cytoplasmic loop of 21 to 91 amino acid residues (depending on which constraints were used). Note that the figure is not drawn to scale.

constructed to minimize disturbance of the K+R bias [35]. For simplicity we will refer in the following text to the produced fusion proteins as e.g. "variant 38-P" or "variant 38-G", where 38 is the number of the C-terminal amino acid residue in Acr3 which is fused to the reporter in the produced fusion protein. P and G correspond to PhoA and GFP respectively.

Table 4
Transmembrane consensus helices of *B. subtilis* Acr3 predicted by different methods

Helix no.	Length (a.a.)	Average starting point	Average end point	Sequence
H1	21	11 (±2)	31 (±2)	LTIWIFLAMALGIGLGFIFPS
H2	21	41 (±3)	61 (±2)	VGTTSIPLAIGLVLMMYPPLA
H3	24	78 (±3)	100 (±2)	LILSLVQNWIIGPTLMFILAIIFL
H4	20	107 (±2)	126 (±2)	YMIGLIMIGLARCIAMVIVW
H5	22	142 (±5)	163 (±2)	AFNSIFQMLFFSVYAYIFVTVT
H6	21	180 (±5)	200 (±4)	EVAKSVFIYLGVPFIAGMVTR
H7	20	220 (±3)	239 (±3)	ISPITLIALLFTIIVMFSLK
H8	22	251 (±3)	272 (±2)	VRVAIPLLIYFVLMFFVSFFLG
H9[a]	21[a]	300 (±4)[a]	320 (±5)[a]	VAVGVFGIHSGAAFAAVIGPL[a]
H9[b]	22[b]	285 (±2)[b]	306 (±1)[b]	LAFTAGSNNFELAIAVAVGVFG[b]
H10[b]	24[b]	311 (±1)[b]	334 (±3)[b]	AAFAAVIGPLVEVPVMIALVKVAL[b]

The standard deviations of the starting points and end points are given in parenthesis.
[a] Consensus helix for a 9 helix topology.
[b] Consensus helix for a 10 helix topology.

3.2. Analysis of fusion points in the predicted first eight TM segments

All prediction programs suggested the first 300 amino acid residues of Acr3 to form eight TM segments with a cytoplasmic location of the N-terminus. To test this experimentally we made 23 complementary fusion constructs with PhoA and GFP and one fusion construct with only GFP (the PhoA variant for fusion point 240 was not obtained) at 12 points in the sequence of Acr3. The fusion points were at amino acid residues 38, 77, 104, 139, 177, 215, 240, 250, 252, 273, 279 and 284. The analysis of the reporter protein signals showed that variants 38-P, 104-P, 250-P and 252-P displayed high PhoA activities with concomitant low signals from 38-G, 104-G, 250-G and 252-G (Fig. 2). This suggests that all these four fusion points are located on the extracellular side of the plasma membrane. Variants 77-G, 139-G, 177-G, and 215-G displayed higher GFP emission intensity relative to the corresponding PhoA variants (Fig. 2). This suggests that these fusion points are located in the cytoplasm. For variant 240-G a strong GFP signal was observed, suggesting a cytoplasmic location also for this fusion point. With the exception of the signals from 240-G and 177-G, the results of the experimental fusion analysis support the prediction of eight TM helices in the first 300 amino acids of Acr3. In the experiments, the location of the N-terminus could not be unequivocally determined as a fusion protein with a fusion point before the first predicted TM-segment would inevitably

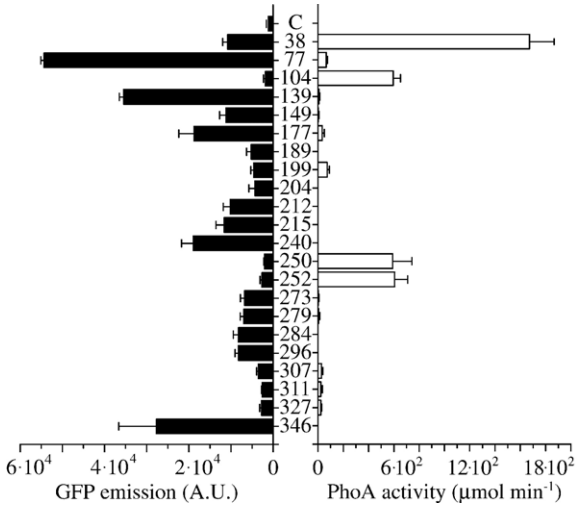

Fig. 2. Results from activity measurements for fusions with PhoA (right panel) and fluorescence emission measurements for fusions with GFP (left panel). Sample "C" corresponds to cells containing empty pBAD vector. Activity data for PhoA are averages from six experiments and emission data for GFP are averages of nine experiments. Error bars show the standard deviations of the experimental data. Fusion constructs 240 and 296 were not obtained with *phoA* and could therefore not be measured. The fusion proteins are denoted by their fusion points.

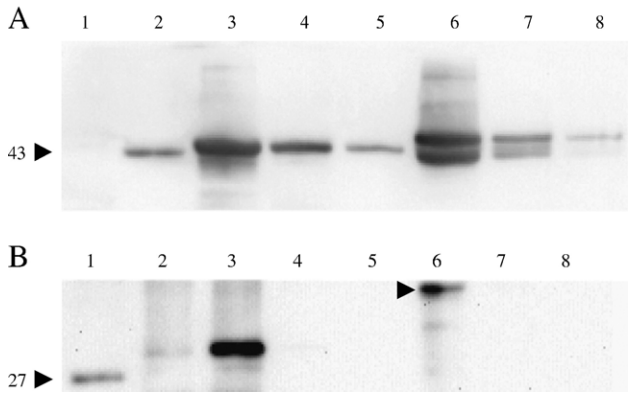

Fig. 3. (A) Western blot of membranes isolated from cells expressing fusion variants 38-P and 104-P. Lane 1) negative control with membranes from cells containing empty pBAD vector, Lane 2) positive control containing purified PhoA, Lanes 3–5) 1:1, 1:10 and 1:100 dilutions of membranes from cells producing fusion variant 38-P. Lanes 6–8) 1:1, 1:10 and 1:100 dilutions of membranes from cells producing fusion variant 104-P. (B) Western blot of membranes isolated from cells expressing fusion variants 77-G and 346-G. Lane 1) positive control containing GFPuv, Lane 2) negative control with membranes from cells containing empty pBAD vector, Lanes 3–5) 1:1, 1:10 and 1:100 dilutions of membranes from cells producing fusion variant 77-G. Lanes 6–8) 1:1, 1:10 and 1:100 dilutions of membranes from cells producing fusion variant 346-G. The size of the full length fusion protein of variant 346-G in lanes 6–8 is indicated by the arrow.

become trapped in the cytosol. However, the strong PhoA signals for fusion points 38 and 104 provide indirect evidence that the N-terminus must be located in the cytoplasm since the polypeptide chain has to traverse the membrane to reach the extracellular side. In support of this conclusion a Western blot analysis of membranes from cells expressing variants 38-P, 104-P and 77-G shows that these variant are indeed anchored in the membrane (Fig. 3A and B).

3.3. Analysis of fusion points aimed at discriminating between a 9 and a 10 helix topology

To clarify the ambiguous signal from variants 177-G and 240-G and to experimentally distinguish between a nine helix topology and a ten helix topology, a set of 9 additional fusion constructs at 5 different points in the sequence were made (fusion variant 296-P was not obtained). Two fusion points (307 and 311) were designed to be positive reporters for a 10 helix topology, two fusion points (296 and 327) were designed to be positive reporters for a nine helix topology and one fusion point (346) was designed to reveal the location of the C-terminus. In the 10 helix prediction model, fusion points 307 and 311 are located in the loop between helices 9 and 10 (Fig. 1A). In the 9 helix prediction model fusion point 296 is located in the predicted cytoplasmic loop between helices 8 and 9 and fusion point 327 is located at the end of the predicted helix 9 (Fig. 1B). For the 9 fusion constructs all signals except one were too weak to be useful for model building. Fusion variant 346-G, which is at the C-terminus of Acr3, displayed a very strong GFP signal whereas the signal from variant 346-P was negligible (Fig. 2). This strongly suggests that the C-terminus of Acr3 is located in the cytoplasm.

3.4. Analysis of fusion points aimed at discriminating between an 8 and a 10 helix topology

The experimental data from variant 346-G as well as the data for all the other conclusive data points (38-P, 77-G, 104-P, 139-G, 215-G, 250-P and 252-P) fit very well with a structural model with 10 TM segments and cytoplasmic locations for the N and C-termini as suggested by the majority of the prediction programs. However, the signals from variants 177-G and 240-G were not consistent with such a 10 TM topology.

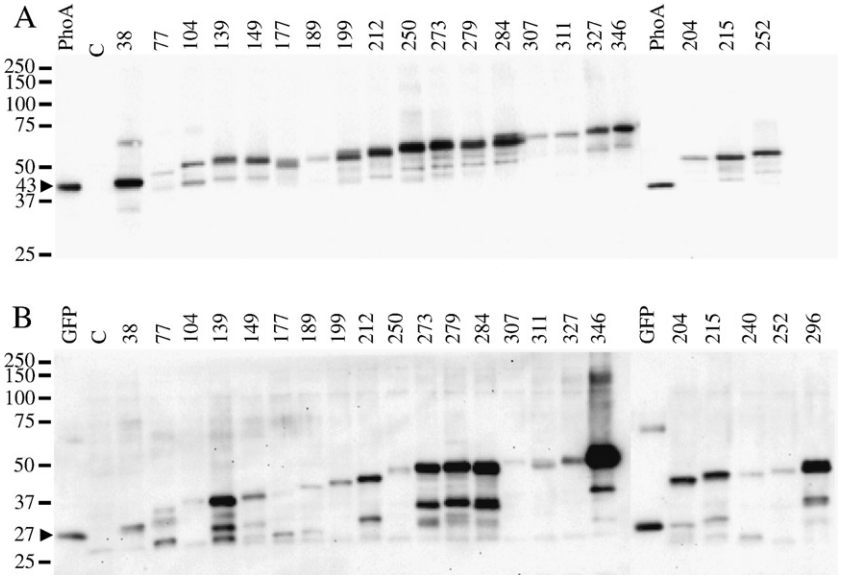

Fig. 4. (A) Western blot of cell samples used for measurements of alkaline phosphatase activity. Purified PhoA was included as positive controls in the lanes denoted PhoA. (B) Western blot of cell samples used for measurements of GFP emission. A cell extract containing GFP was included as positive controls in lanes denoted GFP. The lanes denoted C correspond to cells containing empty pBAD vector. The fusion proteins are denoted by their fusion points.

To create a new model which could account for all the observed results, including that from variants 177-G and 240-G, we modelled the topology of Acr3 again, now using the advanced mode of HMMTOP or the prediction server TMHMMfix [27]. These prediction programs can build refined topological models by constraining the locations of segments of the amino acid sequence to that determined by experiments (cytosolic/extracellular/intra membranous). Several combinations of constraints (data not shown) resulted in topology models with 8 TM helices with a cytoplasmic loop between helices 4 and 5 with a length varying between 21 and 91 amino acids (Fig. 1C). To test the 8 TM model experimentally we designed another fourteen translational fusion constructs (at positions 149, 159, 169, 189, 199, 204 and 212) which were evenly spaced throughout the predicted intracellular loop. Complementary fusion constructs were obtained for all points except fusion points 159 and 169. For these fusion points neither the PhoA variant nor the GFP variant was obtained. The reporter protein signals measured for the 10 fusion variants in the putative cytoplasmic loop were, with two exceptions, too weak to be used for model building. The exceptions, variants 149-G and 212-G, showed moderately strong GFP signals complemented by very low signals for variants 149-P and 212-P. This suggests that fusion points 149 and 212 are located in the cytoplasm (Fig. 2).

3.5. Analysis of reporter protein antigen for all fusion proteins

To verify the presence of full length fusion protein we analysed all cell extracts that had been used for measurement of reporter protein signal by Western blot using antibodies directed against PhoA or GFP. The Western blot analysis confirmed the presence of PhoA and GFP antigen in all samples. In all samples which had displayed strong reporter protein signals bands were found that corresponded well with the expected sizes of the full length fusion proteins (Fig. 4A and B). However, in addition to the full length fusion proteins bands with other sizes were also found indicating the presence of truncated variants of the fusion proteins. This behaviour has been seen previously in translational fusion analyses [36,37] and is thought to result from partial degradation of the fusion protein.

4. Discussion

4.1. The translational fusion method as a means to analyse membrane protein topology

A prerequisite for the rational study of membrane protein function is a basic understanding of the protein structure. This type of information can be obtained by several different techniques. The techniques that give the highest resolution data, NMR and X-ray crystallography, are presently limited to proteins of a relatively small size or that form well diffracting crystals. Hence 3D structures for membrane proteins are still scarce. To circumvent this problem, alternative techniques have been developed which give structural information with lower resolution. One subset of those techniques is methods that report on membrane protein topology. Several alternative topology analysis techniques are available, for example N-linked glycosylation site insertion scanning mutagenesis, protease cleavage protection, epitope insertion scanning mutagenesis, substituted cysteine accessibility method (SCAM) and translational fusion scanning. Of the listed techniques, N-linked glycosylation site analysis is limited to eukaryotic systems, protease cleavage protection is best suited for proteins with relatively large extra membranous domains and SCAM analysis requires the presence of a cysteine-free protein to avoid disturbing background signal. Similar to the other techniques, epitope insertion scanning mutagenesis and translational fusion scanning have in common that the TM segments are silent and hence that structural information is inferred from signals from the extra membranous domains or loops. Following this, poor signals are obtained when probe sites are located within the TM segments. In the present study we have chosen to work with translational fusion scanning. The reason for this is that topology prediction suggests that the extra membranous regions of Acr3 are very short and hence that the epitope insertion scanning approach would likely produce a greater number of ambiguous results. Moreover, our attempts to substitute the single cysteine residue of Acr3 by site directed mutagenesis have been fruitless due to the presence of an extremely stable hairpin structure in the DNA upstream of the Cys codon ($Tm > 85$ °C). The main criticism of the translational fusion scanning method is that it is an invasive technique which may disrupt the native protein folding [38]. Despite this, translational fusion scanning remains a powerful technique for topology analysis and has been shown to produce results that are in very good agreement with both biochemical and high resolution structural studies [39,40].

4.2. Experimental results support a 10 helix topology for Acr3

In the present investigation we have made 42 translational fusion constructs to determine the transmembrane topology of the arsenite transporter Acr3 (Fig. 5). Initial topology predictions indicated that Acr3 would have either nine or ten TM segments. Following this, our first fusion constructs were aimed at discriminating between these two possibilities. Seven strong reporter protein signals corroborated the presence of the predicted first eight TM segments. In addition to this, we obtained two other strong signals from the variants with fusion points at the end of the loop following predicted TM helix H1 and at the end of the full length protein respectively. As both these fusion proteins were anchored in the membrane (Fig. 3A and B) this suggested that both the N-terminus and the C-terminus are located on the cytoplasmic side of the membrane. Consequently, Acr3 has to have an even number of TM segments. This essentially ruled out the nine TM helix model. Two fusion variants, 177-G and 240-G, produced ambiguous results relative to the 10 TM model. Instead of being on the extracellular side, the signals from these constructs suggested that they were located in the cytoplasm. In the case of 177-G these results could be accounted for if Acr3 had a topology with eight TM segments and a large cytoplasmic loop of ~90 amino acid residues instead of the predicted 10 TM topology. To

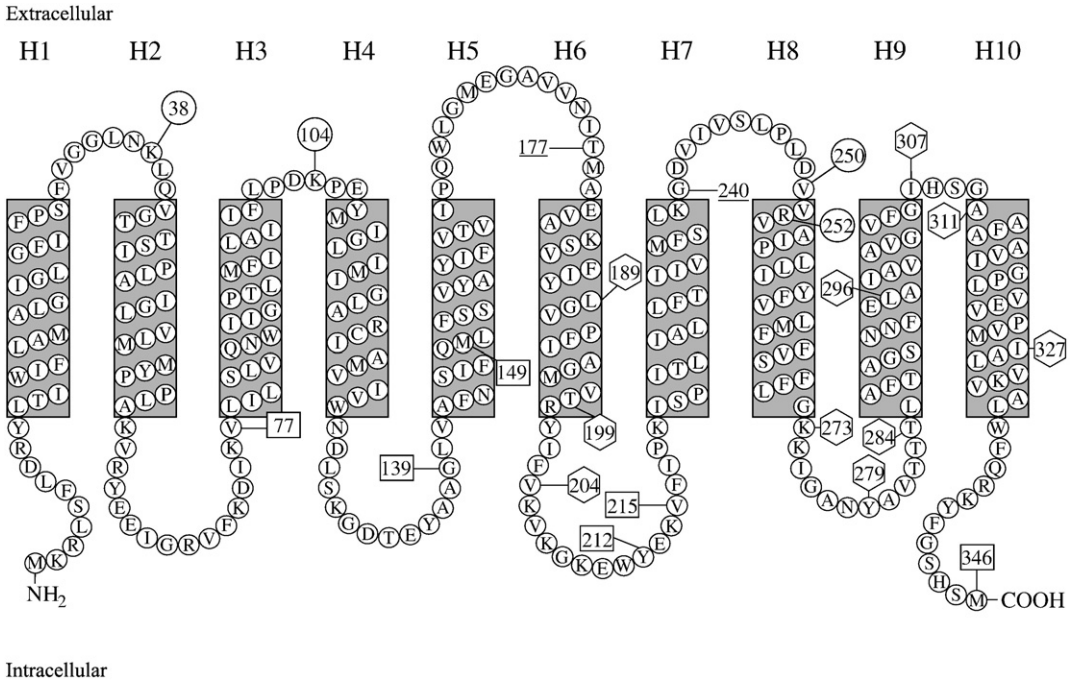

Fig. 5. 10 transmembrane helix topological model for the *B. subtilis* arsenite transporter Acr3 constructed on the basis of experimental results and from modelling including the locations of all fusion points of the study. Numbers in circles represent fusion points at which high PhoA activity was recorded. Numbers in squares represent fusion points at which high GFP activity was seen. Numbers in hexagons represent points for which reporter activities were too weak to be used for modelling. Underlined numbers show fusion points 177 and 240 whose GFP signals contradict the 10 TM model. The membrane spanning segments H1 through H10 in the figure correspond to the consensus helices resulting from comparison of results from prediction (Table 4).

explore this possibility experimentally we constructed another 10 fusion variants. Of these variants only two, 149-G and 212-G, gave significant reporter protein signals. Variant 149-G supported the previously observed intracellular location of loop 5 between H4 and H5 and variant 212-G confirmed the cytoplasmic location of loop 7 between H6 and H7 (Fig. 5) which variant 215-G had already indicated. Hence, no experimental information was obtained which supported the 8 TM model. We also attempted to find an alternative model which could account for the signal from variant 240-G. An eight TM model could also be obtained if fusion points 240 and 346 were constrained to cytoplasmic locations simultaneously. However, this model seemed highly unrealistic as it was directly contradicted by several (139-G, 212-G, 215-G, 250-P and 252-P) of the fusion protein signals. Hence, in our experiments we could not obtain any results consistent with an eight TM model. Another possibility is that the signals from variants 177-G and 240-G could arise from improper folding and retention of the reporter protein in the cytoplasm (c.f. [38]). In support of this idea we have four pieces of evidence obtained by sequence analysis and modelling. Firstly all topology prediction programmes converge on the prediction of eight transmembrane helices with only short connecting loops in the first 300 amino acids. Such consensus is usually only seen for reliable predictions (c.f. [41]). Secondly, we have found that the polypeptide chain segments comprising amino acids 140–180 and 220–270 contain only a few positively charged amino acids. K and R residues are well known to be of great importance for correct insertion of the polypeptide chain into the lipid bilayer [42]. Thirdly, the accuracy of prediction of TM helices of Acr3 seems to be lower from amino acid 142 through 240 than in other parts of the sequence. This is reflected in larger standard deviations of the starting points and end points of the consensus helices for H5 and H6 than for the other TM segments (Table 4, Fig. 6). Finally, modelling of the distance to the center of the bilayer for each amino acid residue yields a maximum distance of ~24 Å throughout the sequence of Acr3 with only 26 residues reaching farther from the membrane centre than 20 Å. In the Z-pred algorithm a distance exceeding 25 Å indicates that a residue is outside the lipid bilayer [28]. This may suggest that the extramembranous loops of Acr3 just barely protrude from the lipid bilayer. Taken together, the large standard deviation of H5 and H6 in combination with the very small predicted extramembranous domains could indicate that H5 and H6 may be significantly longer than the 20 amino acids required to traverse the membrane. Usually such hydrophobic mismatch between the helix length and the thickness of the bilayer results in a tilt of the TM helix relative to the membrane normal (c.f [43]). The forces that stabilize the orientation of transmembrane helices are very complex and poorly understood. Two factors that have been shown to be important for the orientation of TM helices, however, and which are critically dependent on the amino acid sequence of the protein are i) interactions between aromatic and positively charged amino acid side chains and membrane lipid head groups [44] and ii) interhelical contacts [45,46]. It is therefore possible that fusion variants 177-G and 240-G which may contain tilted helical segments in combination with a weak K+R signal may require the presence of a C-

terminal polypeptide segment or TM helix in order to become properly folded. This type of behaviour has previously been seen for the CitS citrate transporter from *Klebsiella pneumoniae* [38,39].

To conclude, our experimental data suggest that *B. subtilis* Acr3 has an even number of TM helices. In the absence of positive signals supporting an 8 TM topology, results from modelling lead us to favour a 10 TM helix topology for Acr3 with short extracellular and cytoplasmic loops (Fig. 5).

4.3. Topological model of the Acr3 family proteins

To extend the topological model that we obtained for Acr3 we also made topology predictions for five other known or putative arsenite transporters of the Acr3 family from *M. tuberculosis*, *S. cerevisiae*, *N. punctiforme*, *S. coelicolor* and *H. influenzae* [1,47,48], using seven prediction servers in single sequence mode and with aligned sequences using HMMTOP [19–24,26]. The proteins were chosen to include representatives from all three Acr3 subgroups from Gram positive and Gram negative bacteria and from fungi while minimizing sequence similarity [49]. The results from the different modelling programs were compared to obtain consensus helices for all six Acr3 family proteins. The Acr3 family proteins from the different organisms display between 24% and 74% sequence identity. Despite these significant differences, modelling indicates that the Acr3 family proteins all share a similar topology (Fig. 6). The crude features of the model suggest that, overall; loops are very short. In fact, for the *B. subtilis* Acr3 protein our modelling data indicate that the average loop length is similar to that of Bacteriorhodopsin which is recognized as having its polypeptide chain almost entirely buried in the lipid bilayer [50]. For all Acr3 proteins the cytoplasmic loops appear to be slightly longer than the extracellular loops. This would seem to fit well with the function of the protein, which is to collect arsenite in the interior of the cell and dispose of it in the periplasm or on the extracellular side. The eukaryotic Acr3 from yeast has in general longer loops than its bacterial counterparts. Interestingly, almost all conserved residues including the cysteine in position 119 (*B. subtilis* sequence as reference) are

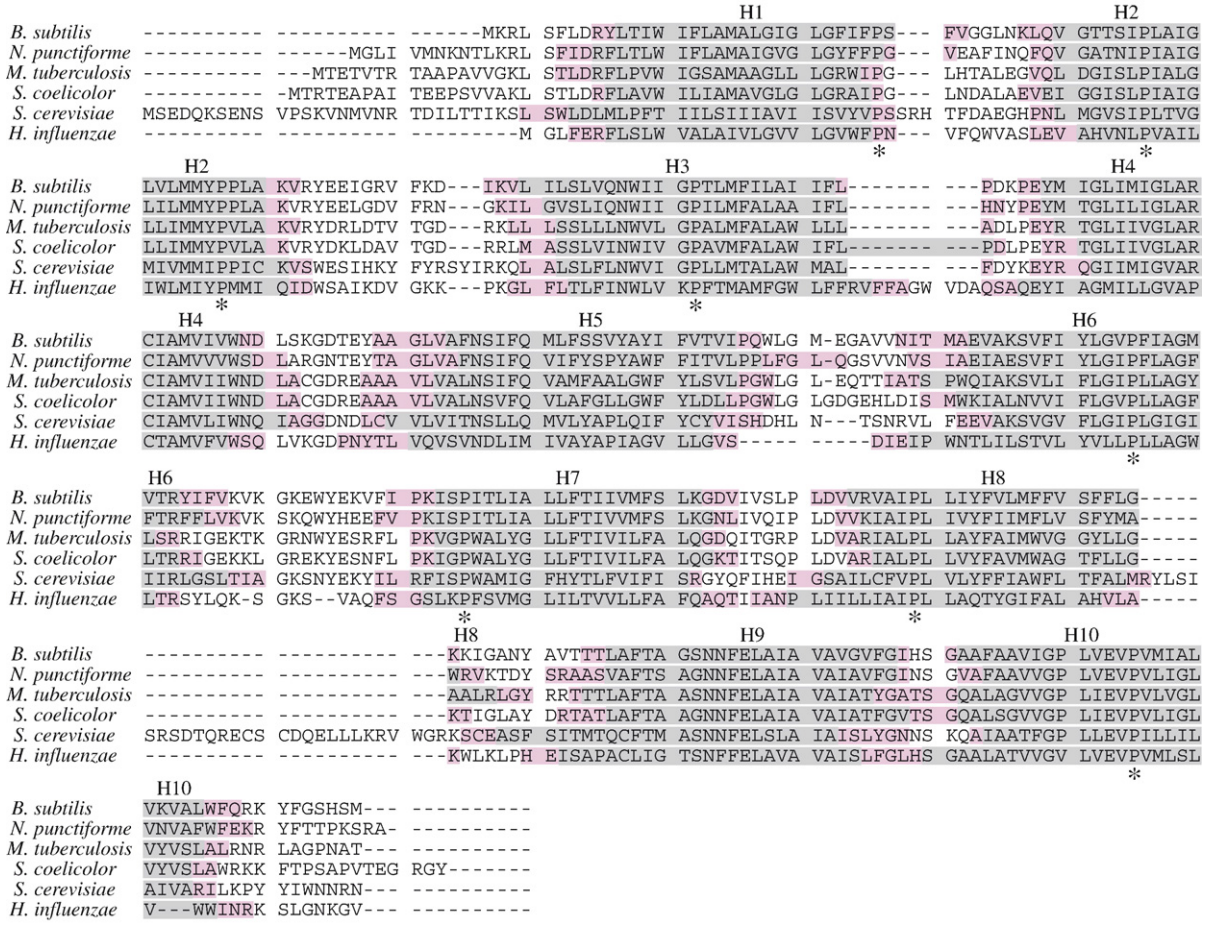

Fig. 6. Alignment between sequences of known or putative arsenite transporting Acr3 family proteins showing consensus helices H1–H10 (in grey shading) that are predicted for each protein to be transmembrane by all or a majority of the modelling programs. Standard deviations from the mean starting points and end points are shown in pink. The proteins in the alignment were chosen to include representatives from all three Acr3 subgroups while minimizing sequence similarity [49] and thus the alignment includes members from both Gram positive and Gram negative bacteria and the eukaryotic *S. cerevisiae*. (*M. tuberculosis* H37Rv (GI:15609780), *S. cerevisiae* (GI:2498103), *N. punctiforme* PCC 73102 (GI:23126620), *S. coelicolor* A3(2) (GI:6900906) and *H. influenzae* R2846 (GI:53732906)). For all Acr3 sequences approximately half of the proline residues are conserved. These are marked in the figure with asterisks (*). In the alignment the 145 amino acid residues which correspond to the cytoplasmic arsenate reductase moiety of the Acr3 family transporter from *M. tuberculosis*, were omitted.

located in transmembrane segments for all Acr3 family proteins. Two exceptions, Gly132 is in an intracellular loop between H4 and H5 and Ser 309 is in an extracellular loop between H9 and H10 in all models. Of the conserved residues 8 are prolines. This amounts to approximately half of the total number of prolines for all members of the alignment. Although prolines are not at all uncommon in transmembrane helices, this may suggest that the transmembrane helices of the Acr3 proteins are kinked in several places [51]. The next step is now to use the structural information obtained in the present investigation to probe the mechanism of transport by *B. subtilis* Acr3.

Acknowledgements

This work was supported by The Swedish Science Council, The Swegene consortium, Magnus Bergvalls stiftelse and Carl Tesdorpfs stiftelse. The authors also wish to express their sincere gratitude towards Prof. L. Hederstedt for discussion and for critical comments on the manuscript and to C. von Wachenfeldt for the generous gift of the *E. coli* extract containing GFPuv.

References

[1] M. Ghosh, J. Shen, B.P. Rosen, Pathways of As(III) detoxification in *Saccharomyces cerevisiae*, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 5001–5006.
[2] Y.L. Meng, Z. Liu, B.P. Rosen, As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*, J. Biol. Chem. 279 (2004) 18334–18341.
[3] T. Sato, Y. Kobayashi, The ars operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite, J. Bacteriol. 180 (1998) 1655–1661.
[4] H.C. Yang, J. Cheng, T.M. Finan, B.P. Rosen, H. Bhattacharjee, Novel pathway for arsenic detoxification in the legume symbiont *Sinorhizobium meliloti*, J. Bacteriol. 187 (2005) 6991–6997.
[5] L.S. Tisa, B.P. Rosen, Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein, J. Biol. Chem. 265 (1990) 190–194.
[6] S. Dey, B.P. Rosen, Dual mode of energy coupling by the oxyanion-translocating ArsB protein, J. Bacteriol. 177 (1995) 385–389.
[7] J. Abramson, S. Iwata, H.R. Kaback, Lactose permease as a paradigm for membrane transport proteins (Review), Mol. Membr. Biol. 21 (2004) 227–236.
[8] J. Wu, L.S. Tisa, B.P. Rosen, Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase, J. Biol. Chem. 267 (1992) 12570–12576.
[9] A.R. Achour, P. Bauda, P. Billard, Diversity of arsenite transporter genes from arsenic-resistant soil bacteria, Res. Microbiol. (2006).
[10] D. Drew, D. Sjöstrand, J. Nilsson, T. Urbig, C.N. Chin, J.W. de Gier, G. von Heijne, Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 2690–2695.
[11] J. Sambrook, D. Russel, Molecular Cloning: A Laboratory Manual, 3 ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001.
[12] D. Ish-Horowicz, J.F. Burke, Rapid and efficient cosmid cloning, Nucleic. Acids Res. 9 (1981) 2989–2998.
[13] C. von Wachenfeldt, L. Hederstedt, *Bacillus subtilis* holo-cytochrome *c*-550 can be synthesised in aerobic *Escherichia coli*, FEBS Lett. 270 (1990) 147–151.
[14] M. Rapp, D. Drew, D.O. Daley, J. Nilsson, T. Carvalho, K. Melen, J.W. De Gier, G. Von Heijne, Experimentally based topology models for *E. coli* inner membrane proteins, Protein Sci. 13 (2004) 937–945.
[15] A. Lewin, L. Hederstedt, Compact archaeal variant of heme A synthase, FEBS Lett. 580 (2006) 5351–5356.
[16] H. Schägger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem. 166 (1987) 368–379.
[17] J. Bengtsson, C. von Wachenfeldt, L. Winstedt, P. Nygaard, L. Hederstedt, CtaG is required for formation of active cytochrome *c* oxidase in *Bacillus subtilis*, Microbiology 150 (2004) 415–425.
[18] M. Cserzo, E. Wallin, I. Simon, G. von Heijne, A. Elofsson, Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method, Protein Eng. 10 (1997) 673–676.
[19] G.E. Tusnady, I. Simon, Principles governing amino acid composition of integral membrane proteins: application to topology prediction, J. Mol. Biol. 283 (1998) 489–506.
[20] D.T. Jones, W.R. Taylor, J.M. Thornton, A model recognition approach to the prediction of all-helical membrane protein structure and topology, Biochemistry 33 (1994) 3038–3049.
[21] B. Rost, P. Fariselli, R. Casadio, Topology prediction for helical transmembrane proteins at 86% accuracy, Protein Sci. 5 (1996) 1704–1718.
[22] T. Hirokawa, S. Boon-Chieng, S. Mitaku, SOSUI: classification and secondary structure prediction system for membrane proteins, Bioinformatics 14 (1998) 378–379.
[23] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[24] R.Y. Kahsay, G. Gao, L. Liao, An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes, Bioinformatics 21 (2005) 1853–1858.
[25] K. Hofmann, W. Stoffel, TMbase — A database of membrane spanning proteins segments, Biol. Chem. Hoppe–Seyler 374 (1993) 166.
[26] M.G. Claros, G. von Heijne, TopPred II: an improved software for membrane protein structure predictions, Comput. Appl. Biosci. 10 (1994) 685–686.
[27] K. Melen, A. Krogh, G. von Heijne, Reliability measures for membrane protein topology prediction algorithms, J. Mol. Biol. 327 (2003) 735–744.
[28] E. Granseth, H. Viklund, A. Elofsson, ZPRED: predicting the distance to the membrane center for residues in alpha-helical membrane proteins, Bioinformatics 22 (2006) e191–e196.
[29] S. Kumar, K. Tamura, M. Nei, MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment, Brief Bioinform. 5 (2004) 150–163.
[30] Y. Akiyama, K. Ito, Folding and assembly of bacterial alkaline phosphatase *in vitro* and *in vivo*, J. Biol. Chem. 268 (1993) 8146–8150.
[31] A.I. Derman, J. Beckwith, *Escherichia coli* alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm, J. Bacteriol. 173 (1991) 7719–7722.
[32] B.J. Feilmeier, G. Iseminger, D. Schroeder, H. Webber, G.J. Phillips, Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*, J. Bacteriol. 182 (2000) 4068–4076.
[33] E.B. Duffy, B. Barquera, Membrane topology mapping of the Na+-pumping NADH: quinone oxidoreductase from *Vibrio cholerae* by PhoA-green fluorescent protein fusion analysis, J. Bacteriol. 188 (2006) 8343–8351.
[34] S.M. Gandlur, L. Wei, J. Levine, J. Russell, P. Kaur, Membrane topology of the DrrB protein of the doxorubicin transporter of *Streptomyces peucetius*, J. Biol. Chem. 279 (2004) 27799–27806.
[35] H. Andersson, E. Bakker, G. von Heijne, Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in *Escherichia coli*, J. Biol. Chem. 267 (1992) 1491–1495.
[36] A.R. Diaz, M.C. Mansilla, A.J. Vila, D. de Mendoza, Membrane topology of the acyl-lipid desaturase from *Bacillus subtilis*, J. Biol. Chem. 277 (2002) 48099–48106.
[37] V. Chepuri, R.B. Gennis, The use of gene fusions to determine the topology of all of the subunits of the cytochrome *o* terminal oxidase complex of *Escherichia coli*, J. Biol. Chem. 265 (1990) 12978–12986.

[38] M. van Geest, J.S. Lolkema, Transmembrane segment (TMS) VIII of the Na(+)/Citrate transporter CitS requires downstream TMS IX for insertion in the *Escherichia coli* membrane, J. Biol. Chem. 274 (1999) 29705–29711.
[39] J. Calamia, C. Manoil, Lac permease of *Escherichia coli*: topology and sequence elements promoting membrane insertion, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 4937–4941.
[40] J. Abramson, I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*, Science 301 (2003) 610–615.
[41] J. Nilsson, B. Persson, G. Von Heijne, Prediction of partial membrane protein topologies using a consensus approach, Protein Sci. 11 (2002) 2974–2980.
[42] G. von Heijne, Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues, Nature 341 (1989) 456–458.
[43] S.H. Park, S.J. Opella, Tilt angle of a trans-membrane helix is determined by hydrophobic mismatch, J. Mol. Biol. 350 (2005) 310–318.
[44] R.B. Koehorst, R.B. Spruijt, F.J. Vergeldt, M.A. Hemminga, Lipid bilayer topology of the transmembrane alpha-helix of M13 Major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy, Biophys. J. 87 (2004) 1445–1455.
[45] M. Eilers, S.C. Shekar, T. Shieh, S.O. Smith, P.J. Fleming, Internal packing of helical membrane proteins, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 5796–5801.
[46] H.R. Kaback, J. Wu, From membrane to molecule to the third amino acid from the left with a membrane transport protein, Q. Rev. Biophys. 30 (1997) 333–364.
[47] E. Ordonez, M. Letek, N. Valbuena, J.A. Gil, L.M. Mateos, Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032, Appl. Environ. Microbiol. 71 (2005) 6206–6215.
[48] L. Wang, S. Chen, X. Xiao, X. Huang, D. You, X. Zhou, Z. Deng, arsRBOCT arsenic resistance system encoded by linear plasmid pHZ227 in *Streptomyces* sp. strain FR-008, Appl. Environ. Microbiol. 72 (2006) 3738–3742.
[49] N.M. Mansour, M. Sawhney, D.G. Tamang, C. Vogl, M.H. Saier, The bile/arsenite/riboflavin transporter (BART) superfamily, FEBS J. 274 (2007) 612–629.
[50] H. Belrhali, P. Nollert, A. Royant, C. Menzel, J.P. Rosenbusch, E.M. Landau, E. Pebay-Peyroula, Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution, Structure 7 (1999) 909–917.
[51] S. Yohannan, S. Faham, D. Yang, J.P. Whitelegge, J.U. Bowie, The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 959–963.
[52] R. Roth, C. Hagerhall, Transmembrane orientation and topology of the NADH:quinone oxidoreductase putative quinone binding subunit NuoH, Biochim. Biophys. Acta 1504 (2001) 352–362.
[53] B.P. Cormack, R.H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP), Gene 173 (1996) 33–38.