Biology of the Cell (1997) **89**, 321-329

Original article

The aquaporin-Z water channel gene of *Escherichia coli*: Structure, organization and phylogeny

Giuseppe Calamita [a,b], Bettina Kempf [d], Kenneth E Rudd [e], Mélanie Bonhivers [b], Susanne Kneip [d], William R Bishai [c], Erhard Bremer [d] and Peter Agre [b]

[a] *Istituto di Fisiologia Generale, Università degli Studi di Bari, via Amendola, 165/A, 70126 Bari, Italy;*

[b] *Departments of Biological Chemistry and Medicine, School of Medicine, Johns Hopkins University;*

[c] *Department of Molecular Microbiology and Immunology, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, MD, USA;*

[d] *Fachbereich Biologie Laboratorium für Mikrobiologie, Philipps-Universität Marburg, Marburg, Germany;*

[e] *Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL, USA*

Aquaporin water channel proteins are found throughout the plant and animal kingdoms, but the first prokaryotic water channel gene, *aqpZ*, was only recently identified in wild type *Escherichia coli* (Calamita G *et al* (1995) *J Biol Chem* 270, 29063–29066). Here we define the organization of *aqpZ* in *E coli*, produce the AqpZ protein and compare the AqpZ phylogeny to that of some known bacterial homologs. Physical mapping and sequence analyses confirmed the location of *aqpZ* at minute 19.7 on the *E coli* chromosome where it is transcribed counterclockwise. The monocistronic nature of *aqpZ* was clearly indicated by the structural organization of its surrounding genes, *ybjD* and *ybjE'* and by the presence of a typical Rho-independent transcriptional terminator following the *aqpZ* stop codon. Computer sequence analysis indicated the –35/–10 region located 72 bases upstream of the *aqpZ* start codon as the most likely *aqpZ* promoter. A series of potential *cis*-regulatory elements were found in the 400 bp region preceding the *aqpZ* ORF. The AqpZ protein, produced under T7Φ10 control, showed a size of about 20 kDa by SDS-PAGE. Striking similarities were found between the *E coli aqpZ* and a gene included in the genome of the cyanobacterium *Synechocystis* sp PCC6803, a species permanently living in fresh water. These results may represent a fundamental step to characterize the regulation and the physiological features of the AqpZ water channel in prokaryotes.

prokaryotes / *Escherichia coli* / aquaporins / water transport / AqpZ

INTRODUCTION

Movement of water into or out of the cell is a basic process of life. The molecular mechanism by which water crosses cell membranes remained poorly understood until the discovery of the aquaporins, a family of water channel proteins found throughout nature (Preston *et al*, 1992). Seven mammalian aquaporins have been identified with distinct physiological roles, and mutations result in clinical disorders (reviewed by King and Agre, 1996). Multiple aquaporins have also been described in plants where they are believed to participate in transpiration and maintenance of cell turgor (reviewed by Chrispeels and Agre, 1994).

The first known prokaryotic aquaporin water channel gene, *aqpZ*, was recently identified in wild-type *E coli* (Calamita *et al*, 1995). Like plant and mammalian aquaporins, the predicted topology of the AqpZ protein has six transmembrane domains and five connecting loops (A to E), however the cytosolic NH_2- and $COOH^-$ termini are shorter and a 10-residue cassette was observed in exofacial loop C. Genomic Southern analyses revealed *aqpZ*-like sequences among Gram-negative and Gram-positive bacteria (Calamita *et al*, 1995). Genes whose

deduced proteins have sequences and topological similarities to AqpZ are present in the genomes of the Gram-negative bacterium *Haemophilus influenzae* (Fleischmann *et al*, 1995) and in a prokaryote lacking a cell wall and having a minimal set of functional genes such as *Mycoplasma genitalium* (Fraser *et al*, 1995). Bacteria also expresses *glpF*, an *aqpZ*-sequence homolog which is included within an operon (Heller *et al*, 1980; Sweet *et al*, 1990) and encodes a glycerol facilitator which is only weakly permeable to water (Maurel *et al*, 1994; Calamita *et al*, 1995). As revealed by phylogenetic analyses, functional distinctions between AqpZ and GlpF probably resulted from the ancient divergence in a primordial gene (Calamita *et al*, 1995; Park and Saier, 1996), however, *aqpZ* and *glpF* genes are not present in the genome of the archaeon *Methanococcus jannaschii* (Bult *et al*, 1996). Interestingly, a third sequence homolog, SmpX, is found in *Synechococcus* (Kashiwagi *et al*, 1995) but its phylogeny suggests a functional distinction from the functionally characterized aquaporins and glycerol facilitators (Park and Saier, 1996). Although aquaporin water channels seem to be largely distributed in prokaryotes, their genetic regulatory features and physiological role are unknown. The model defined with the present study may permit the characterization of the regulation and the physiological function of the aquaporin water channels in bacteria.

MATERIALS AND METHODS

Bacterial strains, media and growth conditions

E coli K-12 strains DH5α (Hanahan, 1983) or MC4100 (Casabadan, 1976) were employed for cloning studies. The *E coli* strain BL21(λDE3) was used for the AqpZ production under the T7Φ10 control (Studier and Moffat, 1986). Bacteria were grown aerobically at 37°C in LB broth or M9 minimal medium (Sambrook *et al*, 1989) supplemented with 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.2% casamino acids. When necessary, media included ampicillin or chloramphenicol at final concentrations of 50 μg/mL and 30 μg/mL, respectively.

Plasmid constructions

Routine manipulations of DNA were done according to standard procedures (Sambrook *et al*, 1989). After each plasmid preparation, the insert was sequenced by double-strand dideoxynucleotide termination (US Biochemical Corp, USA). As previously reported (Calamita *et al*, 1995), the plasmid pGC94 was a pBS-SK(–) (Stratagene, USA) derivative carrying *aqpZeco*, a 3171 bp segment of the E coli K-12 chromosome containing the entire *aqpZ* gene and flanking regions. The *aqpZeco* sequence has been submitted to GenBank as an update to accession number U38664. A segment of DNA composed by the *aqpZ* coding region and its 100 bp of 3'-flanking region was amplified by PCR using pGC94 as template and GCBD2 (5'-TTGCTTAGCTCATGAAA-3') and GCBU11 (5'-CGGAATTCATGTTCAGAAAATTAGCAGCTG-3') as primers. The PCR product was blunted and digested with *Eco*RI at the 3'- and 5'-ends, purified, and cloned into a pBS-KS(+) vector which had been linearized with *Not*I, blunted and then digested with *Eco*RI. The resulting plasmid named pGC95 is available at the American Type Culture Collection (deposit no 87386). The plasmids pOPAZ and pAZE were derived from low copy number T7Φ10 promoter expression vectors (Dersch *et al*, 1994). A 1263 bp DNA segment containing the *aqpZ* ORF plus the 50 and 516 bp segments of the *aqpZ* 5'- and 3'-flanking regions was amplified by PCR from pGC94 by using GCBU15 (5'-CGGAATTCTATAAAACGACCATATTTT-3') and GCBD4 (5'-ATTAACAGCCCAGAAAA-3') as primers. The PCR fragment was digested with *Eco*RI and *Sma*I and inserted into the *Eco*RI/*Eco*RV site of pPD100 leading to pOPAZ. pAZE was prepared by inserting the *Eco*RI-*Sal*I fragment of pGC95 (*aqpZ* ORF plus the 100 bp of the 3'-flanking region) into the *Eco*RI/*Sal*I site of pPD110.

Sequence analyses

The pGC94 plasmid was employed to sequence both of the DNA strands of *aqpZeco*. The restriction map of *aqpZeco* was obtained with DNA Strider 1.2 program (CEA, France). The additional coding regions flanking the *aqpZ* ORF in *aqpZeco* were identified using the GeneMark program (Borodovsky *et al*, 1993). MacVector program (Oxford Molecular Group, USA) was used to inspect for *cis*-regulatory elements within the 500 bp preceding the *aqpZ* ORF. The presence of possible repetitive extragenic palindromic (REP) sequences following the *aqpZ* stop codon was investigated with SiteFinder, a computer program devised by Webb Miller and Kenneth Rudd. Possible extensive secondary structures of *aqpZeco* were explored with the RNA folding program mfold (Jaeger *et al*, 1989). Database searches for homologs of *aqpZ*, as well as of the surrounding genes *ybjD* and *ybjE*, were performed using the BLAST programs and the nr databases maintained at the National Center for Biotechnology Information (Altschul *et al*, 1990; Gish and States, 1993). The computer mapping of *aqpZ* on the *E coli* K-12 chromosome was performed with MapSearch program (Rudd *et al*, 1991) by aligning the *aqpZeco* restriction map with a recent version of the digital integrated genomic restriction map EcoMap8, which is derived from the map of Kohara (Kohara *et al*, 1987; Berlyn *et al*, 1996). MapSearch was also used to determine the probable direction of the *aqpZ* transcription on the *E coli* chromosome. Sequence alignments between AqpZ and its homologs were performed with the Lipman-Pearson method. Phylogenetic trees were obtained with the Lasergene program (DNASTAR, United Kingdom).

Physical mapping of *aqpZ* on the *E coli* chromosome

A digoxigenin(DIG)-labeled (Boehringer-Mannheim, Germany) *aqpZ* probe representing the whole *aqpZ* coding region was hybridized to 3 μL of all lysates of the Kohara miniset collection (Kohara *et al*, 1987) spotted onto a nylon membrane. Phage chromosomal DNA was

purified by using the Qiagen Lambda kit (Qiagen, Germany). Fine mapping of *aqpZ* was performed by Southern blot by hybridizing the *aqpZ* probe with the *Bgl*I- and *Pvu*I-digested DNA of Kohara phages 211, 212, and 213.

Production of the AqpZ polypeptide under T7Φ10 control

The *aqpZ*-encoded protein, AqpZ, was expressed by using the pOPAZ and pAZE constructs. In both of the plasmids, *aqpZ* was under the transcriptional control of the T7Φ10 promoter (Studier and Moffat, 1986). For the translation, pOPAZ carried the ribosome-binding site of *aqpZ* while pAZE carried that of *lacZ*. The plasmids were transformed into the BL21(λDE3) and the production of AqpZ was induced by adding isopropyl-1-thio-β-galactopyranoside (IPTG) to the cultures. The newly synthesized proteins were visualized as previously reported (Dersh *et al*, 1994).

RESULTS

Sequence analysis of *aqpZ*

A 3171 bp fragment of *E coli* chromosomal DNA containing the *aqpZ* gene from a wild-type bacterium was cloned into the *Pst*I-*Kpn*I site of pBS-SK(–) yielding the plasmid pGC94 (Calamita *et al*, 1995). The restriction map and the gene organization of *aqpZ* were determined with the open reading frame between positions 1775 and 2470 (fig 1). Computer sequence analyses revealed several potential *cis*-regulatory elements. Putative –35/–10 regions at positions 1306 (TTGATA-N_{17}-TTCAAC), 1485 (TTGCTG-N_{18}-CAAAAT), and 1703 (TCGTCA-N_{17}-TATAAA) reside 469, 290, and 72 bases upstream of the *aqpZ* start codon. Sequence analyses revealed that *aqpZ* lacks an easily recognizable Shine-Dalgarno region, but the GTGG 9 bp upstream of the ATG start codon is the most likely ribosome-binding site (RBS). The lack of a downstream box (Sprengart *et al*, 1990) among nucleotides +15 to +26 of the *aqpZ* coding region indicates the absence of this translational enhancer. Other potential prokaryotic *cis*-regulatory elements only partially matched the corresponding canonical consensi. These include catabolite activator protein sites (TGTGA at positions 1366, 1430, and 1749), and sequences corresponding to stationary sigma factor binding regions (three *rpoS* sites), oxidative stress, nitrogen starvation (*rpoN*), and late (*rpoH*) and early heat shock (*rpoE*) elements (data not shown).

Analysis of the *aqpZ* 3'-untranslated region revealed a likely Rho-independent transcriptional terminator located 8 bases downstream the *aqpZ* stop codon (fig 1). This hairpin with a calculated secondary structure of –16.7 kcal/mole is immediately followed by a string of seven unbroken T residues and is very likely to confer efficient transcription termination. This would prevent readthrough into the downstream *ybjE'* ORF. The region of *aqpZ* located between bp 2576 and 2794 (fig 1) is 98% identical to a 221 bp region located at positions 41926 through 42146 in ECOHU47 (Genbank U00007), a segment of *E coli* DNA located at minute 47.8 on the *E coli* K-12 chromosome (Berlyn *et al*, 1996). This repeated region is not a repetitive extragenic palindromic (REP) sequence (Dimri *et al*, 1992) and it is likely not to be transcribed, unlike REPs which are in mRNA sequences. Two potentially stable stem-loop structures were found in the 2594–2625 and 2730–2751 regions of the repeat (–25.5 and –14.2 kcal/mol) and are of unknown biological significance.

The genetic organization of the *aqpZ* region was further evaluated by searching for the presence of open reading frames flanking the *aqpZ* gene. Two new open reading frames are located 426 bp upstream and 494 bp downstream of the *aqpZ* (fig 1). A full length open reading frame (1257 bp) is located on the complementary DNA strand where it is transcribed in the direction opposite to *aqpZ*. A second partial open reading frame (207 bp) is located on the same DNA strand as *aqpZ* and is transcribed in the same direction. These upstream and downstream flanking genes are temporarily referred to here as *ybjD* and *ybjE'*, using the *E coli* ORF naming system (Berlyn *et al*, 1996). YbjD and Old (Myung and Calendar, 1995) align over the entire 418 amino acids of YbjD, with the strongest similarity in a 180 amino acid region (25.6% identity). Old has a 150 amino acid terminal domain not present in YbjD. The ATP binding motifs of YbjD are well-conserved, as is region II of this family. The short region IV motif appears to be absent from YbjD (data not shown). Purified Old protein has been shown to have both RNA and DNA exonuclease activities and YbjD may be a new *E coli* nuclease. Database searches did not reveal any convincing homolog to the N-terminal 69 amino acids of the partial YbjE' sequence. *ybjD* defines the 5'-end of the *cis*-regulatory region of *aqpZ* which should be located within positions 1350 and 1702 of *aqpZ*. The direction of transcription of *ybjD*, the presence of likely transcriptional terminator following the *aqpZ* stop codon, and the long intergenic segment of DNA (473 bp) between *aqpZ* and *ybjE'* indicate that the *aqpZ* gene is monocistronic.

Mapping of *aqpZ* on the *E coli* chromosome

Chromosomal mapping of *aqpZ* was undertaken by a computer database analysis and by physical methods. After our mapping studies were completed, Oshima *et al* (1996) reported the sequence of the 12.7–28.0 min region of the *E coli* K-12 chromo-

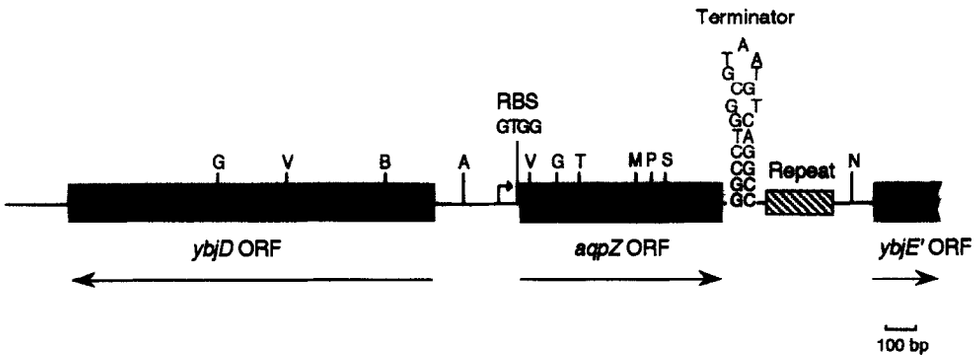

Fig 1. Restriction map and organization of the *aqpZ* gene in *E coli*. The adjacent open reading frames *ybjD* and *ybjE'* are located upstream and downstream of *aqpZ* with directions of transcription indicated *(arrows)*. The putative –35/–10 promoter *(angled arrow)*, ribosome-binding site (RBS) and terminator of transcription (Terminator) are represented. Intergenic sequences repeated at min 47.8 on the *E coli* chromosome are shown (Repeat). Restriction sites are denoted (A, *ApaLI*; B, *BglII*; G, *BglI*; M, *BsmI*; N, *NsiI*; P, *PvuI*; S, *SacII*; T, *SnaBI*; V, *PvuII*). The entire nucleotide sequence is accessible as GenBank accession number U38664.

some which included the *aqpZ* gene at min 19.9. Thus, we briefly describe part of our results of mapping as confirmatory *aqpZ* mapping data.

MapSearch program (Rudd *et al*, 1991) aligned the restriction map of *aqpZ* to EcoMap8, an updated version of the *E coli* genomic restriction map (Kohara *et al*, 1987; Berlyn *et al*, 1996; KER, unpublished). The *aqpZ* gene was located at minute 19.7 on the *E coli* chromosome in the counterclockwise orientation (fig 2). MapSearch predicted the presence of *aqpZ* on miniset clones 211 and 212 of the Kohara collection of *E coli* genomic inserts. To experimentally verify the predicted location of *aqpZ* on the *E coli* chromosome, we used the *aqpZ* open reading frame to probe Southern blots of DNA from these bacteriophage clones. As predicted, appropriately sized restriction fragments obtained by restriction digests of the Kohara phages 211 and 212 DNA hybridized with the *aqpZ* probe, thus confirming our previous computer mapping results locating *aqpZ* at minute 19.7 on the *E coli* chromosome (fig 2). No hybridizations were observed with DNA from other Kohara phages, indicating the presence of a single copy of *aqpZ* on the *E coli* chromosome.

Expression of AqpZ protein

The pAZE and pOPAZ expression plasmids were constructed to produce and electrophoretically characterize the AqpZ polypeptide in *E coli*. The preparation of these plasmids is reported in *Materials and methods*. In pAZE, the *aqpZ* expression is mediated by the T7 transcription and translation initiation elements, while in pOPAZ the transcription and translation of *aqpZ* are controlled by the T7 promoter and the *aqpZ* ribosome binding site; the *aqpZ* coding region is followed by the homologous transcription terminator signal in both plasmids. These plasmids were used to transform the *E coli* strain BL21(λDE3) which carries a chromosomal copy of the structural gene for the T7 RNA polymerase under the control of the *lacPO* regulatory element (Studier and Moffat, 1986). Expression of *aqpZ* was induced by adding IPTG to the cultures and the translated protein was then visualized by Coomassie blue staining of the total cell extract separated by SDS-PAGE (fig 3). Controls were performed by expressing pPD100 and pPD110 or by omitting IPTG. A polypeptide of ~20 kDa was induced by IPTG (fig 3), and an identical pattern was also observed by expression under the T7 control when visualized by labeling with [^{35}S]methionine (data not shown). The induced polypeptide band at 20 kDa is slightly smaller than 23.6 kDa deduced from the *aqpZ* DNA sequence but is consistent with increased electrophoretic mobilities exhibited by other hydrophobic integral membrane proteins (Helenius and Simons, 1975).

Comparative analyses of AqpZ and its cyanobacterial homolog

Our previous genomic Southern analyses indicated the presence of *aqpZ*-like genes among both Gram-positive and Gram-negative bacterial species (Calamita *et al*, 1995). The hybridization patterns were not the same than those obtained by probing with the *E coli* glycerol facilitator gene, *glpF*. Thus, we speculated about the dual existence of the functionally distinct genes *aqpZ* and *glpF* as a frequent feature among bacteria.

Because of the recent sequencing of bacterial, yeast, plant and animal genomes, we searched for other *aqpZ* homologs by screening the AqpZ amino

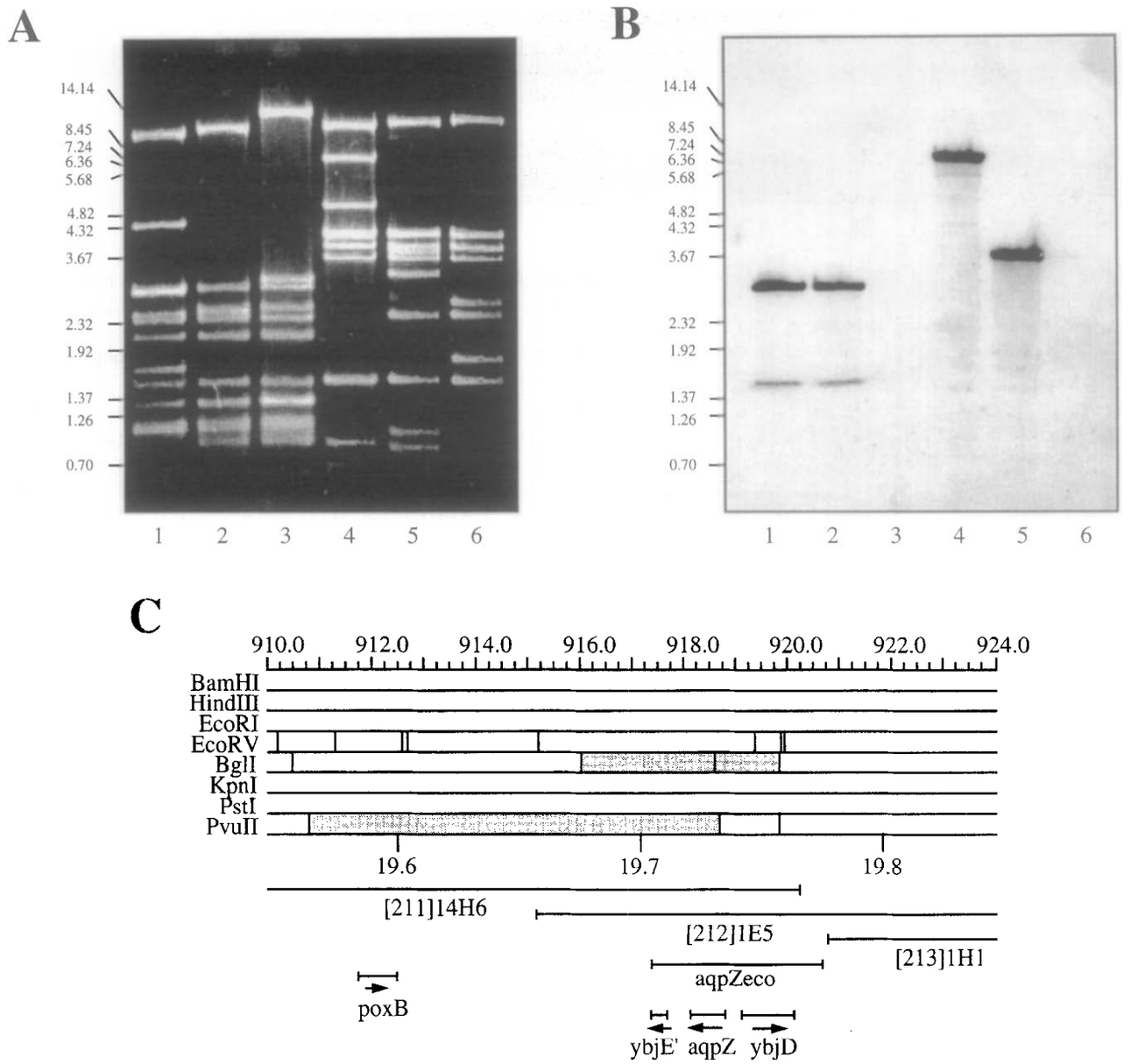

Fig 2. Physical mapping of *aqpZ* on *E coli* K-12 chromosome. **A.** Restriction pattern of the Kohara phages 211 (lanes 1 and 4), 212 (lanes 2 and 5) and 213 (lanes 3 and 6) after 10 μg of each DNA was digested with *BglI* (lanes 1–3) or *PvuII* (lanes 4–6) and electrophoresed into 1% agarose gel. Size markers (in kb) are on the left. **B.** Southern analysis of blot from **A** after hybridization with the *BglI-AvaII* fragment of the *aqpZ* open reading frame. **C.** Portion of the *E coli* genomic restriction map showing the location of *aqpZ* on the chromosome. Grey segments represent the restriction fragments hybridizing with the *aqpZ* probe in **B.** The top scale is the genomic address in kilobases; the lower scale is in minutes (length of the chromosome ÷ 100). Physical locations of chromosomal inserts in Kohara phages 211, 212, and 213, and position of *aqpZ* open reading frame are depicted.

acid sequence against the available database by using the BLAST programs (Altschul *et al*, 1990; Gish and States, 1993). A strikingly high homology (87.8% similarity) was found between the *E coli* AqpZ and the deduced amino acid sequence of the *aqpZ* gene included in the genome of the cyanobacterium *Synechocystis* species PCC6803 (Kaneko *et al*, 1996). As *aqpZ* in *E coli*, the structure of the *Synechocystis* chromosomal region including the *aqpZ* ORF clearly suggests a monocistronic transcription for this gene (data not shown). Figure 4 shows the Lipman-Pearson alignment of the deduced amino acid sequences of *E coli* and *Synechocystis* sp PCC6803 AqpZ. Similar to AqpZ, the Kyte-Doolittle hydrophobicity profile of the *Synechocystis* AqpZ predicts a membrane protein having six transmembrane domains (TM 1 to 6) and five connecting loops (A to E) (fig 4). However, the *Synechocystis* homolog

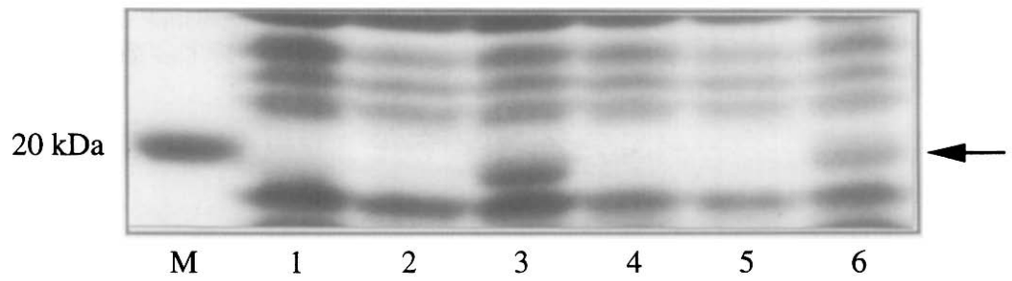

Fig 3. Expression of the *aqpZ*-encoded protein under the control of the T7Φ10 promoter. The expression was performed in strain BL21(λDE3), carrying plasmids pPD100 (induced, lane 1), pOPAZ (uninduced, lane 2; induced, lane 3), pPD110 (induced, lane 4) and pAZE (uninduced, lane 5; induced, lane 6). The proteins were electrophoretically separated on a 15% SDS-polyacrylamide gel and visualized by staining with Coomassie brilliant blue. The position of the *aqpZ*-encoded protein or AqpZ is indicated by the *arrow*. Lane M contains the molecular mass marker of 20 kDa.

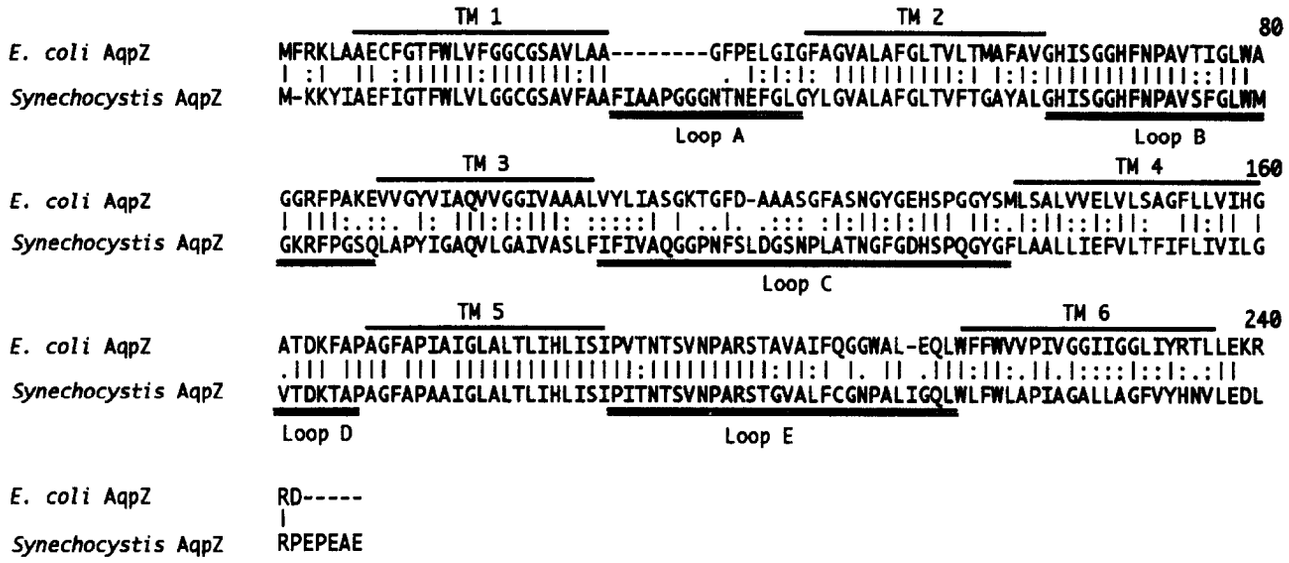

Fig 4. Comparative alignment of deduced amino acid sequences of *E coli* and *Synechocystis* species PCC6803 AqpZ. Deduced amino acid sequence alignments were performed by Lipman-Pearson program analysis. Presumed bilayer-spanning domains are indicated by single boldface lines (TM1–TM6). The polypeptide sequences joining transmembrane domains are doubly underlined (connecting loops A–E). Identical amino acids are joined with straight lines, whereas similar amino acids are connected with dots or colons based on polarity and charge.

contained distinct features since the exofacial loop A and the carboxy terminus were longer than the corresponding domains in AqpZ. Interestingly, compared to all the known aquaporins and members of the MIP family (Park and Saier, 1996) the deduced amino acid sequence of the *Synechocystis* sp PCC6803 AqpZ contains a third NPA motif located 10 residues downstream of the second highly conserved NPA in loop E.

A phylogenetic analysis was performed to study the evolutionary origins and the possible functional relationship between the *E coli* and cyanobacterial AqpZ. Some known prokaryotic homologs (Park and Saier, 1996) and an animal (AQP1) and a plant (γTIP) aquaporin were also included in the phylogenetic tree. As shown in figure 6, the cyanobacterial AqpZ is the closest homolog of *E coli* AqpZ and falls in the aquaporin branch by suggesting a water channel function for the *Synechocystis* AqpZ protein. Interestingly, another cyanobacterial sequence homolog, the *Synechococcus* sp PCC7942 protein SmpX (Kashiwagi *et al*, 1995), is not included in the aquaporin branch or the glycerol facilitator branch of the phylogenetic tree and thus may have a distinct function (fig 5).

DISCUSSION

Identification of bacterial water channel proteins may now permit molecular investigations of pro-

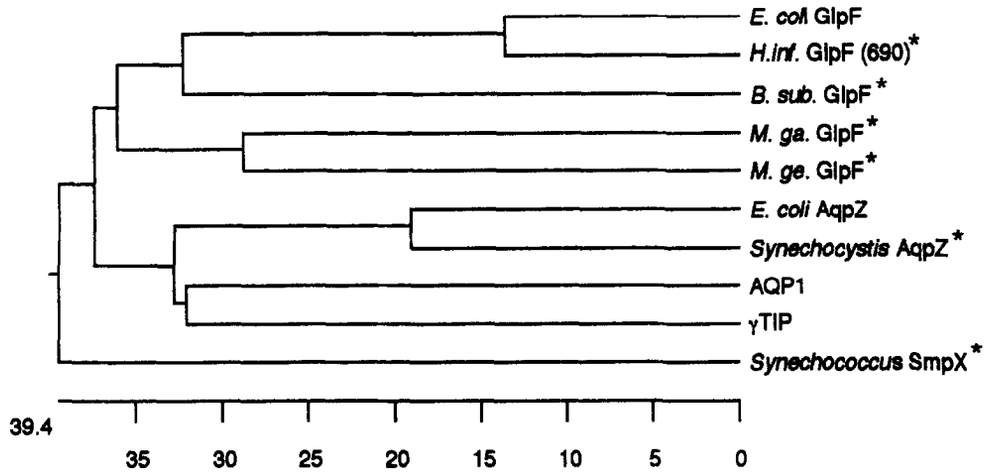

Fig 5. Comparative phylogenetic analysis. Deduced amino acid sequences from indicated aquaporins and homologous proteins were aligned by the Clustal method. AqpZ was compared to some sequenced bacterial homologs and to a plant (γTIP) and an animal (AQP1) aquaporin. Phylogenetic tree was constructed with Lasergene computer program. Branch lengths are proportional to evolutionary distances. Lower scale reports the genetic distances. Abbreviations are: *E coli, Escherichia coli; H inf, Haemophilus influenzae; M ga, Mycoplasma gallisepticum; M ge, Mycoplasma genitalium; Synechocystis, Synechocystis* species PCC6803; *Synechococcus, Synechococcus* species PCC7942. The amino acid sequence of *H inf* GlpF (690) was deduced from the ORF 690 of the *H influenzae* genome (Fleischmann *et al*, 1995). *Asterisk* (*) indicates homologs whose function is not experimentally demonstrated, yet.

karyotic membrane water permeability, and study of the origin and evolution of the aquaporin family of proteins may now be undertaken with a greater perspective. The existence of water channels had previously been postulated in prokaryotes because of the demonstrated movements of water across the bacterial cytoplasmic membrane (Csonka and Hanson, 1991) and from phylogenetic analyses (Reizer *et al*, 1993). These predictions have been verified by recognition of the bacterial water channel protein AqpZ in *E coli* (Calamita *et al*, 1995). Therefore, these studies were initiated to define the genetic structure of *aqpZ* in the *E coli* chromosome, to produce and characterize the *aqpZ*-encoded protein and to perform a phylogenetic analysis of AqpZ and its known homologs among bacteria.

The chromosomal location of *aqpZ*, structural organization of the surrounding genes, and identification of a putative Rho-independent transcription terminator region have been accomplished (figs 1, 2). These findings indicate that *aqpZ* is monocistronic, a feature consistent with the general nature of the aquaporins which are known to be constitutively activated without any accessory proteins. The monocistronic nature further distinguishes *aqpZ* from its *E coli* sequence-homolog, *glpF* (Heller *et al*, 1980; Sweet *et al*, 1990), which encodes a glycerol transporter and is contained within an operon comprising several genes involved in glycerol metabolism, thus supporting the hypothesis of a large phylogenetic separation between *aqpZ* and *glpF* (Calamita *et al*, 1995; Park and Saier, 1996). The genome of the archeon *Methanococcus jannaschii* does not contain genes sharing similarities with *aqpZ* or *glpF* (Bult *et al*, 1996), suggesting that the divergence may have occurred in a primordial gene which originated among eubacteria. The dual existence of functionally distinct paralogues such as aquaporins and glycerol facilitators may be a feature conserved throughout eubacterial genomes.

The *cis*-regulatory region of *aqpZ* should reside within the 400 bases preceding the translation initiation site, since the ORF of the *ybjD* gene begins 426 bases upstream of the *aqpZ* ORF and is transcribed in the opposite direction. Potential activities of the *aqpZ cis*-regulatory elements identified by computer search may now be evaluated experimentally. Among the potential *aqpZ* promoters, the -35/-10 element located 72 bases upstream of the *aqpZ* ORF seems most likely to represent the *aqpZ* promoter region. Nevertheless, in spite of a good overall score, the -10 hexamer TATAAA does not contain all the primary three base pairs characterizing the canonical consensus, TA---T. In addition, the -35 region, TCGTCA, separated by the usual 17 bases from the -10 region, also partially differs from the consensus TTG---.

The presence of a typical *aqpZ* gene in the genome of the cyanobacterium *Synechocystis* sp PCC6803 (Kaneko *et al*, 1996) is consistent with the large distribution of aquaporin water channel genes

among bacteria. Phylogenetic analysis and sequence alignment studies suggest a water channel function for the *Synechocystis aqpZ*-encoded protein although this needs to be experimentally proven. Speculatively, the presence of a typical *aqpZ* gene in a fresh water prokaryotic species such as *Synechocystis* sp PCC6803 may suggest a role for the aquaporin-Z water channels in the osmoregulatory mechanisms of bacteria being in hypoosmotic environments.

In conclusion, in this work we define the structure of the aquaporin-Z water channel gene and express its encoded protein in *E coli*. These results represent an important step for future works aimed to study the biochemical features of the AqpZ polypeptide and to characterize the regulation and the physiological function of the aquaporin water channels in bacteria. In addition, the *aqpZ* gene and AqpZ protein may serve as suitable molecular model for studying genetic, transport, and structural features relevant to the entire aquaporin family of water channels.

ACKNOWLEDGMENTS

We thank Drs Angela Ludovico, Antonietta Monte, Maria Svelto, Giovanna Valenti, and Paul Blount for important contributions and discussions. This work was supported in part by the National Institute of Health grants HL33991 and HL48268 (PA) and AI36973 (WRB), Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie (EB), Foundation pour la Recherche Médicale and Bourse de formation à l'étranger INSERM (MB). The nucleotide sequence reported in this paper has been submitted to GenBank with accession number U38664.

REFERENCES

Altschul SF, Gish W, Miller W, Myers W and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215, 403–410

Berlyn M, Brooks-Low K and Rudd KE (1996) Integrated linkage map of *Escherichia coli* K-12. *In: Escherichia coli and Salmonella: Cellular and Molecular Biology, 2nd edn* (Neidhart FC, Curtiss III R, Ingrham JL, Lin ECC, Brooks-Low K, Magasik B, Reznikoff W, Riley M, Schaechter M and Umbarger HE, eds) ASM Press, Washington, DC, edition 9, 1715–1902

Borodovsky M and McIninch J (1993). GENEMARK: parallel gene recognition for both DNA strands. *Comput Chem* 17, 123–133

Bult CJ, White O, Olsen G, Zhou L, Fleischmann RD, Sutton GG, Blake J. A, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb J, Adams MD, Reich CI, Overbeek R, Kikness EF, Weinstock KG, Merrick J, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley, JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk H, Fraser CM, Smith HO, Woese CR and Venter JC (1996) Complete genome sequence of the Methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073

Calamita G, Bishai WR, Preston GM, Guggino WB and Agre P (1995) Molecular cloning and characterization of AqpZ, a water channel from *Escherichia coli*. *J Biol Chem* 270, 49, 29063–29066

Casabadan MJ (1976). Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J Mol Biol* 104, 541–555

Chrispeels MJ and Agre P (1994) Aquaporins: water channel proteins of plant and animal cells. *Trends Biochem Sci* 19, 421–425

Csonka LN and Hanson AD (1991) Prokaryotic osmoregulation: genetics and physiology. *Annu Rev Microbiol* 45, 569–606

Dersch P, Fsihi H and Bremer E (1994) Low-copy-number T7 vectors for selective gene expression and efficient protein overproduction in *Escherichia coli*. *FEMS Microbiol Lett* 123, 19–26

Dimri GP, Rudd KE, Morgan MK, Bayat H and Ferro-Luzzi Ames G (1992) Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. *J Bacteriol* 174, 4583–4593

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J, Dougherthy, BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu L, Glodek A, Kelley JM, Weidman JF, Phillips C A, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM. Gnehm CL. McDonald LA, Small K V, Fraser CM, Smith HO and Venter JC (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann, RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchmann J L, Weidman J. F., Small KV, Sandusky M, Furhmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomm J, Dougherty BA, Bott KF, Hu P, Lucier TS, Peterson SN, Smith HO, Hutchison III CA and Venter JC (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403

Gish W and States DJ (1993). Identification of protein coding regions by database similarity search. *Nat Gen* 3, 266–272

Hanahan D (1983). Sudies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557–580

Helenius A and Simons K (1975) Solubilization of membranes by detergents. *Biochim Biophys Acta* 415, 29–79

Heller KB, Lin ECC and Wilson TH (1980) Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J Bacteriol* 144, 274–278

Jaeger JA, Turner DH and Zuker M (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* 86, 7706–7710

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M and Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3, 3, 109–136

Kashiwagi S, Kanamaru K and Mizuno T (1995) A *Synechococcus* gene encoding a putative pore-forming intrinsic membrane protein. *Biochim Biophys Acta* 1237, 189–192

King LS and Agre P (1996) Pathophysiology of the aquaporins water channels. *Annu Rev Physiol* 58, 619–648

Kohara Y, Akiyama K and Isono K (1987) The physical map of the whole *E coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 50, 495–508

Maurel C, Reizer J, Schroeder JI, Chrispeels MJ and Saier MH Jr (1994) Functional characterization of the *Escherichia coli* glycerol facilitator, GlpF, in *Xenopus* oocytes. *J Biol Chem* 269, 11869–11872

Myung H and Calendar R (1995) The *old* exonuclease of bacteriophage P2. *J Bacteriol* 177, 497–501

Oshima T, Aiba H, Baba T, Fujita K, Hayashi K, Honjo A, Ikemoto K, Inada T, Itoh T, Kajihara M, Kanai K, Kashimoto K, Kimura S, Kitagawa M, Makino K, Masuda S, Miki T, Mizobuchi K, Mori H, Motomura K, Nakamura Y, Nashimoto H, Nishio Y, Saito N, Sampei G, Seki Y, Tagami H, Takemoto K, Wada C, Yamamoto Y, Yano M and Horiuchi T (1996) A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7–28.0 min region on the linkage map. *DNA Res* 3, 137–155

Park JH and Saier MH Jr (1996) A phylogenetic characterization of the MIP family of transmembrane channel proteins. *J Membr Biol* 153, 171–180

Preston GM, Piazza-Carroll T, Guggino WB and Agre P (1992) Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* 256, 385–387

Reizer J, Reizer A and Saier MH (1993) The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the repeated halves of the proteins. *Crit Rev Biochem Mol Biol* 28, 235–257

Rudd KE, Miller W, Werner C, Ostell J, Tolstoshev C and Satterfield SG (1991) Mapping sequenced *E coli* genes by computer: software, strategies and examples. *Nucleic Acids Res* 19, 637–647

Sambrook J, Fritsch EF and Maniatis TE (1989) *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY

Sprengart ML, Fatscher, HP and Fuchs E (1990) The initiation of translation in *E coli*: apparent base pairing between the 16S rRNA and downstream sequences of the mRNA. *Nucleic Acids Res* 18, 1719–1723

Studier FW and Moffat BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression. *J Mol Biol* 189, 113–130

Sweet G, Gandor C, Voegele R, Wittekindt N, Beuerle J, Truniger V, Lin ECC and Boos W (1990) Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product. *J Bacteriol* 172, 424–430

Received 23 June 1997; accepted 11 September 1997