Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 49, Issue of December 8, pp. 29063–29066, 1995

Molecular Cloning and Characterization of AqpZ, a Water Channel from *Escherichia coli**

(Received for publication, September 25, 1995, and in revised form, October 6, 1995)

Giuseppe Calamita‡, William R. Bishai§¶, Gregory M. Preston‡§, William B. Guggino‖, and Peter Agre‡§**

From the Departments of ‡Biological Chemistry, §Medicine, and ‖Physiology, School of Medicine and the ¶Department of Molecular Microbiology and Immunology, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, Maryland 21205-2185

The aquaporin family of molecular water channels is widely expressed throughout the plant and animal kingdoms. No bacterial aquaporins are known; however, sequence-related bacterial genes have been identified that encode glycerol facilitators (*glpF*). By homology cloning, a novel aquaporin-related DNA (*aqpZ*) was identified that contained no surface *N*-glycosylation consensus. The *aqpZ* RNA was not identified in mammalian mRNA by Northern analysis and exhibited bacterial codon usage preferences. Southern analysis failed to demonstrate *aqpZ* in mammalian genomic DNA, whereas a strongly reactive DNA was present in chromosomal DNA from *Escherichia coli* and other bacterial species and did not correspond to *glpF*. The *aqpZ* DNA isolated from *E. coli* contained a 693-base pair open reading frame encoding a polypeptide 28–38% identical to known aquaporins. When compared with other aquaporins, *aqpZ* encodes a 10-residue insert preceding exofacial loop C, truncated NH_2 and COOH termini, and no cysteines at known mercury-sensitive sites. Expression of *aqpZ* cRNA conferred *Xenopus* oocytes with a 15-fold increase in osmotic water permeability, which was maximal after 5 days of expression, was not inhibited with $HgCl_2$, exhibited a low activation energy (E_a = 3.8 kcal/mol), and failed to transport nonionic solutes such as urea and glycerol. In contrast, oocytes expressing *glpF* transported glycerol but exhibited limited osmotic water permeability. Phylogenetic comparison of aquaporins and homologs revealed a large separation between *aqpZ* and *glpF*, consistent with an ancient gene divergence.

The entry and release of water from cells is a fundamental process of life. Although biophysical features of membrane water permeability were recognized, the molecular pathway of transmembrane water movement remained unknown until discovery of the aquaporins, a family of proteins that permits selective, osmotic flow of water across cell membranes (1). AQP1, the first recognized and characterized homolog, is abundant in mammalian red cells, renal proximal tubules, and other epithelia, where it is constitutively active (2). cDNAs encoding several other mammalian aquaporins have been isolated by homology cloning (3). In renal collecting duct AQP2 (4) is regulated by vasopressin (5, 6), and AQP3 is also slightly permeable to glycerol (7–9). Other aquaporins exist in brain (AQP4) (10, 11), salivary and lacrimal glands, cornea, and lung (AQP5) (12), and a weak water transporter is in lens (MIP) (13).

Aquaporins are expressed in diverse species. Plants express numerous different aquaporins in tonoplast or plasma membranes, and these proteins apparently play essential roles in maintenance of turgor and transpiration (14). A homologous sequence is known to be expressed in *Escherichia coli* and other bacteria but encodes GlpF, the glycerol facilitator (15), which exhibits minimal water permeability (16). Despite phylogenetic analyses suggesting their common prokaryotic origin, no aquaporins have been reported in bacteria.

Changes in the extracellular osmolarity elicit similar physical effects on cells throughout nature, and similarities are observed among the mechanisms of cellular osmoregulation. In bacteria, the cellular turgor is an essential feature based on the high osmotic pressure of the cytoplasm. Although some of the physiological and genetic responses to osmotic shifts among different bacteria are already known, the mechanisms of water transport during bacterial osmoregulation are poorly understood (17, 18). Here we report the isolation and the molecular characterization of AqpZ, the first prokaryotic water channel. The presence of this aquaporin in Gram-positive and Gram-negative species suggests a widespread existence of this aquaporin among bacteria and suggests structural explanations for the functional differences between the sequence-related proteins AqpZ and GlpF.

MATERIALS AND METHODS

DNA Preparations and Screening—Bovine eyes were obtained from a local slaughterhouse. Eye cups were isolated by dissection, and an oligo(dT) cDNA library was prepared in the *Xho*I/*Eco*RI site of the Uni-ZAP XR vector (ZAP-cDNA synthesis kit, Stratagene). Aliquots of the library were used for PCR[1] with nested oligonucleotide primers corresponding to the sequences surrounding the Asn-Pro-Ala motifs in the aquaporins (11, 12). A product of 387 bp was subcloned into the *Eco*RI-*Bam*HI site of pBS-KS(+) and sequenced by double-strand dideoxynucleotide termination (U. S. Biochemical Corp.). This DNA fragment was radiolabeled with [α-^{32}P]dCTP (Amersham Corp.) by random DNA labeling (Boehringer Mannheim) and was used used to screen 3×10^5 plaques from the cDNA library. A positive plaque designated "*aqpZ*" was isolated, and the related pBS-SK(−) phagemid containing a 3.24-kilobase insert was recovered by *in vivo* excision and sequenced.

Based on codon usage analysis of *aqpZ* (19), amplification of *E. coli* chromosomal DNA (DH5α strain) was undertaken by PCR. Specific oligonucleotides (GCBU3, 5′-TATCTGGTGAATGACTCGCC-3′ and GCBD2, 5′-TTGCTTAGCTCATGAAAGGA-3′) were designed corresponding to 5′- and 3′-untranslated sequences of *aqpZ* and were used as

* This work was supported in part by National Institutes of Health Grants HL33991, HL48268, DK32753, and AI36973. This work was presented at the Aquaporins and Epithelial Water Transport International Symposium, Manchester, United Kingdom, September 13, 1995 (Program p. 38 (abstract)). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U38664.

** To whom correspondence should be addressed: Dept. of Biological Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205. Fax: 410-955-3149.

[1] The abbreviations used are: PCR, polymerase chain reaction; bp, base pair(s).

primers (94 °C, 1 min; 51 °C, 1 min; 72 °C, 1 min; 30 cycles). The resulting 917-bp PCR band was subcloned into pBS-KS(+) and sequenced.

Genomic Southern Blot Analysis—Total genomic DNA from bovine spleen and chromosomal DNA from selected Gram-positive and -negative bacterial species were digested with *Bgl*I and *Ava*II, electrophoresed into 0.9% agarose gel, and transferred to a nylon membrane (GeneScreen Plus, DuPont). The blots were hybridized at low or high stringency in a 10% dextran sulfate solution (1 M NaCl, 1% SDS, 100 μg/ml salmon sperm DNA, 10% dextran sulfate, 50 mM sodium phosphate, pH 7) with ^{32}P-labeled DNA probes from coding sequences: 562-bp *Bgl*I-*Ava*II fragment of *aqpZ*, 251-bp *Apa*I-*Hin*dIII fragment of *glpF*, or the 837-bp *Eco*RI-*Sma*I fragment of *AQP1*.

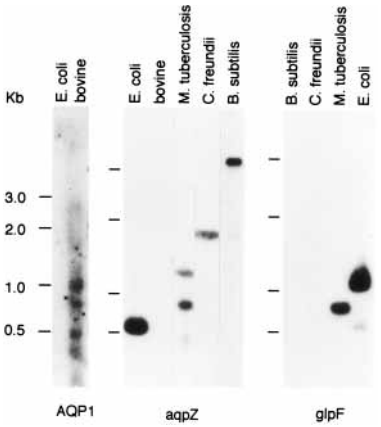

FIG. 1. **Genomic Southern analyses.** Mammalian genomic DNA (18 μg, bovine) and chromosomal DNA samples from selected bacterial species (50 ng, *E. coli DH5α*, *M. tuberculosis*, *C. freundii*, and *B. subtilis*) were digested with *Ava*II and *Bgl*I, electrophoresed into agarose gels, and transferred to blots. Probes corresponding to coding regions of human *AQP1* (837 bp) and *aqpZ* (562 bp) and *E. coli glpF* (251 bp) were labeled with ^{32}P and hybridized to the blots under conditions of medium stringency. *Kb*, kilobases.

Expression in Xenopus Oocytes—The *Dra*I-*Sma*I fragment of *aqpZ* containing 98 bp of the 5′-untranslated sequence, 162 bp of the 3′-untranslated sequence, and the entire *aqpZ* open reading frame was blunt-ligated into the *Bgl*II site of the *Xenopus* expression construct pXβG, which contains the *Hin*dIII-*Pst*I insert of pSP64T in pBS-KS, and the pXβG-AQP1 construct was used as a control (20). The pXβG-glpF construct was a generous gift from Dr. C. Maurel (16). Capped cRNAs were synthesized *in vitro* by using T3 RNA polymerase after a digestion with *Bam*HI. Defolliculated stage V–VI oocytes from *Xenopus laevis* were injected with 50 nl of water or up to 25 ng of sample cRNAs and incubated in 200 mosM modified Barth's solution at 18 °C.

Osmotic Water Transport Assay—After incubation for 2–5 days, the oocytes were transferred to 70 mosM modified Barth's solution at 21 °C; oocyte swelling was monitored by videomicroscopy, and the coefficient of osmotic water permeability (P_f) was determined (20). The Arrhenius activation energy (E_a) was calculated from the P_f values obtained swelling the oocytes at 10, 20, and 30 °C.

Oocyte Substrate Uptake Assays—Uptake of [^{14}C]glycerol (specific activity, 1.6 GBq/mmol, Amersham Corp.) or [^{14}C]urea (specific activity, 2.0 GBq/mmol, Amersham Corp.) was measured after 4 days of incubation. Individual oocytes were added to 100 μl containing 200 mosM Barth's solution, 12 μM [^{14}C]glycerol, or 16 μM [^{14}C]urea and incubated for up to 30 min at 20 ± 2 °C. The oocytes were then washed rapidly five times with ice-cold Barth's solution and lysed in 5% SDS (200 μl/oocyte) for liquid scintillation counting.

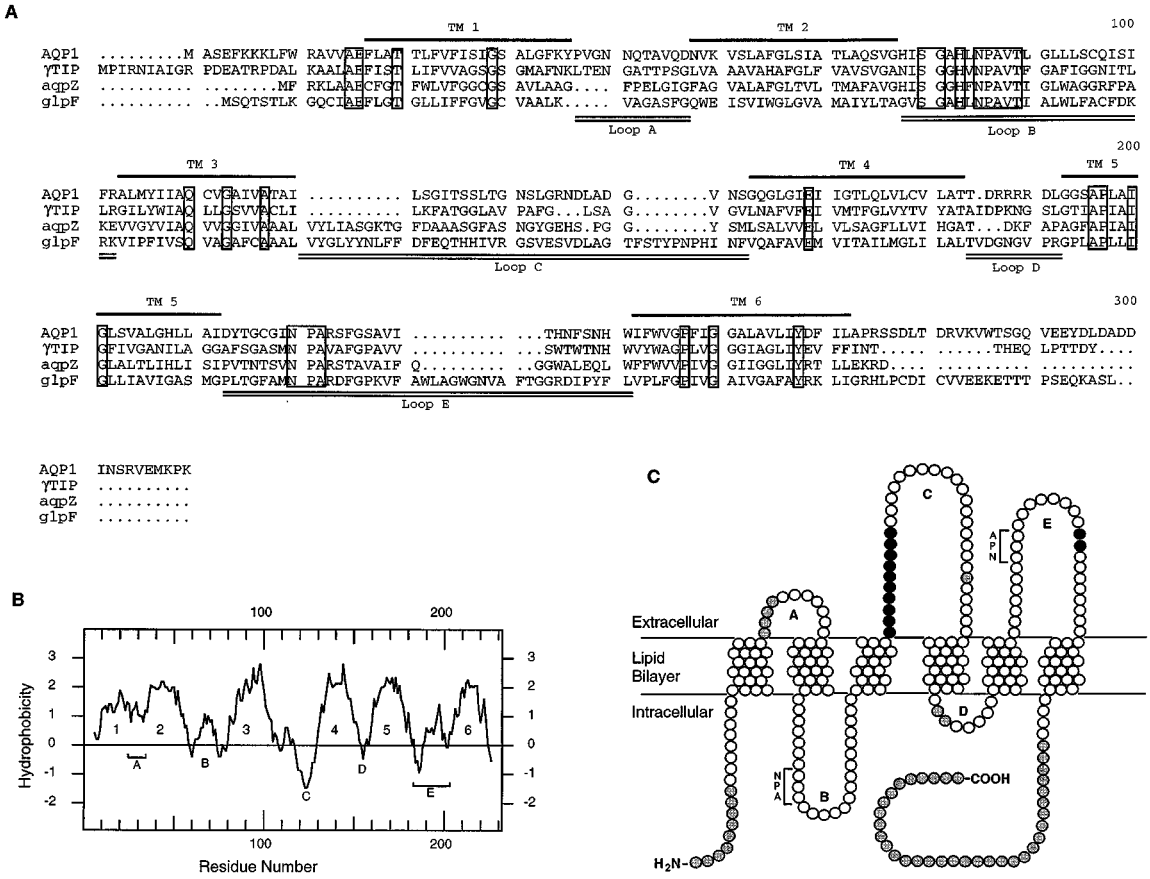

FIG. 2. **Comparative alignment of deduced amino acid sequence, hydrophobicity analysis, and predicted membrane topology of AqpZ protein.** *A*, deduced amino acid sequence alignments were performed by PILEUP program analysis of AqpZ, human AQP1, plant γTIP, and *E. coli* GlpF. *Single boldface lines* indicate the six predicted transmembrane domains (*TM 1* to *TM 6*). The polypeptide sequences joining transmembrane domains are *doubly underlined* (connecting loops A–E). Positions with fully conserved residues are enclosed. The gaps show the amino acid cassettes among the sequences in loops A, C, D, and E. *B*, Kyte-Doolittle hydrophobicity profile of AqpZ using a 7-residue window. *C*, predicted topology of AqpZ compared with AQP1. *Open circles* represent corresponding residues present in both structures; *closed circles* represent extra residues in AqpZ; *stippled circles* represent residues missing from AqpZ. The extracellular space corresponds to the *E. coli* periplasmic space.

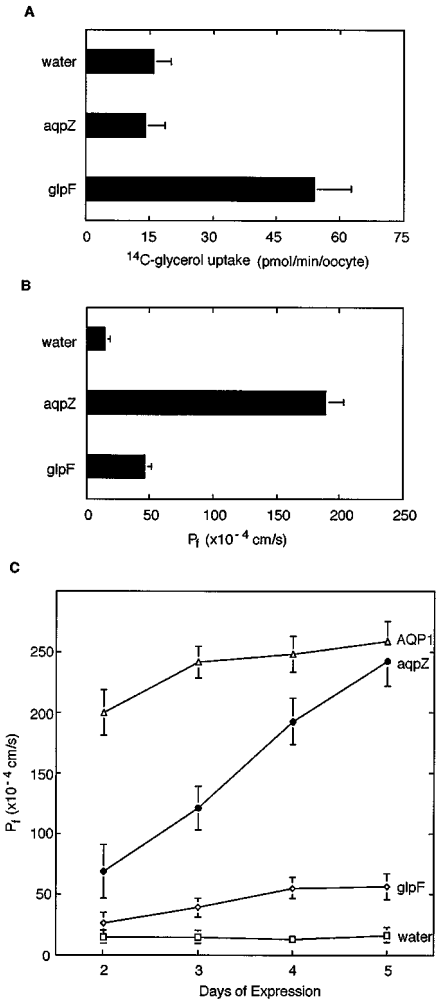

FIG. 3. **Comparative transport functional analyses.** *A*, [^{14}C]glycerol influx was measured in oocytes injected with water (control) and 25 ng of cRNA for *aqpZ* or *glpF* and measured after 4 days of expression (see "Materials and Methods"). *B*, osmotic water permeability (P_f) measured on oocytes injected similarly. *C*, osmotic water permeability of oocytes compared with time of expression. Oocytes were injected with water (control), 25 ng of cRNA for *aqpZ* or *glpF*, or 10 ng of cRNA for *AQP1* and incubated for 2–5 days before measurement of osmotic water permeability. Each *bar* or *data point* represents the mean ± S.E. of 8–12 individual oocytes.

Oocyte Voltage Clamp Studies—Ion conductance was measured on oocytes after 3–5 days of incubation by two-electrode voltage clamp (20).

RESULTS AND DISCUSSION

Bacterial DNAs Homologous to Aquaporins—The two highly conserved tandem repeats within the amino acid sequences from all known aquaporin water transport proteins have been exploited for isolation of new cDNAs by polymerase chain amplifications using degenerate oligonucleotide primers. This approach carries the risk of amplifying cDNAs from contaminating tissues or even bacteria. Searches for aquaporin cDNAs from bovine ocular tissues obtained from slaughterhouses repeatedly led to the amplification of a DNA lacking the characteristic *N*-glycosylation consensus. Northern blots of mRNA from bovine ocular tissues failed to hybridize when the PCR product was used as a probe, and when codon sequences were checked, the DNA exhibited a distinct preference for bacterial usage. This PCR product was used to isolate identical coding sequences from a bovine eye cup cDNA library and from *E. coli* genomic DNA and is now designated *aqpZ*.

Bovine genomic DNA as well as equivalent genomic samples of DNA from *E. coli* and other bacterial species were digested

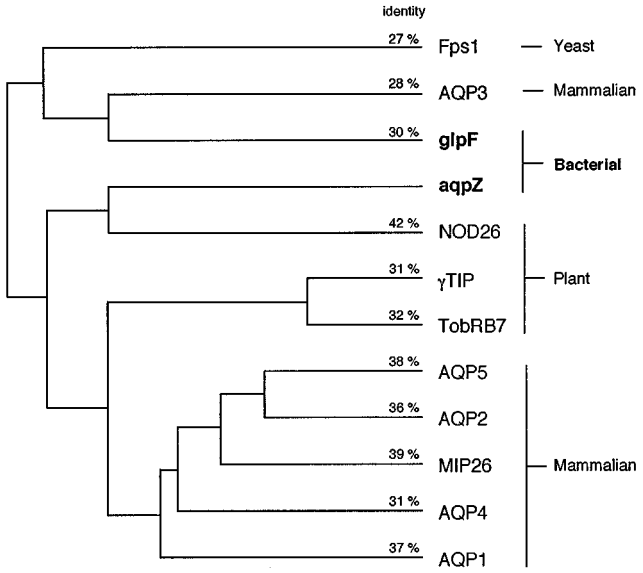

FIG. 4. **Phylogenetic analysis of AqpZ, other aquaporins, and homologs.** Deduced amino acid sequences from the indicated aquaporins and homologous proteins were aligned by PILEUP program analysis (version 7.1, Vax computer system). The percent identity compared with AqpZ is indicated.

with *Ava*II and *Bgl*I and evaluated by Southern analysis with probes corresponding to the human water channel protein (*AQP1*), the novel DNA (*aqpZ*), and the glycerol facilitator from *E. coli* (*glpF*). As expected, *AQP1* failed to hybridize with *E. coli* DNA but reacted with a multiple band in bovine DNA (Fig. 1, *left*). In contrast, the novel DNA (*aqpZ*) provided a striking band at 530 bp with *E. coli* DNA but failed to hybridize with bovine DNA (Fig. 1, *center*); this probe also hybridized with multiple bands in DNA from *Mycobacterium tuberculosis* (atypical Gram-positive), *Citrobacter freundii* (Gram-negative), and *Bacillus subtilis* (Gram-positive). Hybridization of the blot with *glpF* revealed a weak band in DNA from *B. subtilis* (not visible in Fig. 1), one of the bands in the *M. tuberculosis* sample, and gave a striking band of ~1100 bp in the *E. coli* sample (Fig. 1, *right*).

The deduced amino acid sequence of *aqpZ* was compared with sequences of mammalian (*AQP1*) and plant (*γTIP*) aquaporins and *E. coli glpF* (Fig. 2*A*). Alignments revealed the existence of residues known to be highly conserved in each of the transmembrane domains as well as the functionally important loops B and E. Hydropathy analysis of AqpZ was similar to that of other aquaporins (Fig. 2*B*) and is consistent with the existence of six transmembrane domains (21) and hydrophobic loops B and E, which may form the aqueous pore by dipping into the bilayer from opposite leaflets (22). The predicted topology of AqpZ contained distinctive features (Fig. 2*C*). The NH_2- and COOH-terminal domains of AqpZ were shorter than corresponding domains in plant and mammalian homologs. A single cassette encoding 10 extra hydrophobic residues was present at the junction of transmembrane domain 3 and exofacial loop C in AqpZ, and two cassettes were present in GlpF (Fig. 2*A*).

Transport Functions—cRNAs corresponding to *aqpZ* and *glpF* were injected into *Xenopus* oocytes and assessed for permeability to glycerol or water (Fig. 3). As expected, water-injected control oocytes showed limited permeability to isotopically labeled glycerol. Oocytes injected with cRNA encoding AqpZ exhibited glycerol permeability equivalent to the control oocytes, whereas oocytes injected with cRNA encoding GlpF had a severalfold rise in glycerol permeability (Fig. 3*A*). In contrast, oocytes expressing AqpZ exhibited a marked increase

in osmotic water permeability (P_f), while oocytes expressing GlpF featured only a modest increase in P_f (Fig. 3*B*). Similar to known aquaporin water transporters, the AqpZ oocytes exhibited an Arrhenius activation energy of 3.8 ± 1.5 kcal/mol, and consistent with the lack of a cysteine preceding the downstream Asn-Pro-Ala motif (21), incubation of AqpZ oocytes in $HgCl_2$ did not significantly lower P_f (data not shown). In other studies of AqpZ oocytes, no newly expressed conductances were detected when transmembrane voltage was clamped at several different potentials, and no increased uptake of isotopically labeled urea was measured (data not shown).

Previous studies reported an even lower P_f of oocytes injected with *glpF* cRNA and incubated for only 2–3 days (16), so a time course of expression and water transport was performed. After 2 days of expression, the AqpZ oocytes exhibited P_f values significantly below those expressing AQP1 cRNA, but after additional days of incubation, the water permeability of AqpZ oocytes rose significantly, whereas the permeability of GlpF oocytes remained low (Fig. 3*C*). While it is not known why the *aqpZ* cRNA was expressed more slowly, it may reflect the different preferences in codon usage or it may be a result of defective membrane targeting of AqpZ protein, which lacks a long COOH-terminal cytoplasmic domain that is apparently needed for efficient transit to the plasmalemma (22). Nevertheless, it is apparent that despite highly related deduced amino acid sequences, the two bacterial proteins AqpZ and GlpF perform distinct transport functions.

Evolution of the Aquaporin Homologs—Sequence comparisons of *aqpZ*, *glpF*, and genes for plant and animal aquaporins revealed a wide phylogenetic separation of the two bacterial genes. As indicated (Fig. 4) most members of the aquaporin family and related proteins contain amino acid sequences with highly conserved motifs (Fig. 2*A*), but overall the sequences were only 27–42% identical. Although evolutionary distances may only be inferred by computer sequence comparisons, the study demonstrates that *aqpZ* and *glpF* may have diverged very early; the yeast gene *FPS1*, which compensates for suppressed growth on fermentable sugars (23), and the product of mammalian gene *AQP3*, which transports a small amount of glycerol in addition to water (7–9), may have originated from a common prokaryotic gene related to *glpF*. Likewise, the plant and the other mammalian aquaporins may have originated from a prokaryotic gene related to *aqpZ*. The highest homology is shared with *NOD26*, a plant gene expressed in peribacteroid root nodules (24).

The relative simplicity of bacterial genomes is consistent with functional requirements of all genes. While it is premature to conclude what role AqpZ plays in bacterial metabolism, it seems highly likely that it may be expressed for management of osmotic stress (17, 18), cell growth and division, or desiccation, which may be involved in chronic dormancy. The presence of two related genes, *aqpZ* and *glpF*, encoding proteins with distinct transport functions implies the need for both, and the sequence differences between these genes may provide structural insight into their distinctive transport functions. For example, the presence of only one cassette insert in loop C of AqpZ but two cassette inserts in loop C and another cassette insert in loop E of GlpF may be of structural significance. Indeed, genomic Southern analysis (Fig. 1) is consistent with the widespread existence of *aqpZ*-related genes among bacterial species. In contrast to plant genomes that contain numerous aquaporin genes,[2] when the entire sequence of the *Hemophilus influenzae* gene was recently sequenced (25), only two genes were identified that are sequence-related to the aquaporins. Of note, the *H. influenzae* gene *glpF 690* and the *E. coli glpF* are contained within operons encoding other known glycerol metabolic enzymes, whereas the other *H. influenzae* gene, *glpF 1017*, and *E. coli aqpZ* do not exist within operons. These observations lead us to speculate that dual existence of the functionally distinct genes *aqpZ* and *glpF* may be a general feature of bacterial species.

Acknowledgments—We thank Drs. Hamilton Smith, Luis Pablo Cid-Soto, Rajkumar Patil, and James Dick for important contributions and discussions.

REFERENCES

1. Chrispeels, M. J., and Agre, P. (1994) *Trends Biochem. Sci.* **19,** 421–425
2. Agre, P., Preston, G. M., Smith, B. L., Jung, J. S., Raina, S., Moon, C., Guggino, W. B., and Nielsen, S. (1993) *Am. J. Physiol.* **265,** F463–F476
3. Knepper, M. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6255–6258
4. Fushimi, K., Uchida, S., Hara, Y., Hirata, Y., Marumo, F., and Sasaki, S. (1993) *Nature* **361,** 549–552
5. Nielsen, S., DiGiovanni, S. R., Christensen, E. I., Knepper, M. A., and Harris, H. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 11663–11667
6. Deen, P. M. T., Verdijk, M. A. J., Knoers, N. V. A. M., Wieringa, B., Monnens, L. A. H., van Os, C. H., and van Oost, B. A. (1994) *Science* **264,** 92–95
7. Ishibashi, K., Sasaki, S., Fushimi, K., Uchida, S., Kuwahara, M., Saito, H., Furukawa, T., Nakajima, K., Yamaguchi, Y., Gojobori, T., and Marumo, F. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 13052–13056
8. Echevarria, M., Windhager, E. E., and Frindt, G. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10997–11001
9. Ma, T., Frigeri, A., Hasegawa, H., and Verkman, A. S. (1994) *J. Biol. Chem.* **269,** 21845–21849
10. Hasegawa, H., Ma, T., Skach, W., Matthay, M. A., and Verkman, A. S. (1994) *J. Biol. Chem.* **269,** 5497–5500
11. Jung, J. S., Bhat, R. V., Preston, G. M., Guggino, W. B., Baraban, J. M., and Agre, P. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 13052–13056
12. Raina, S., Preston, G. M., Guggino, W. B., and Agre, P. (1995) *J. Biol. Chem.* **270,** 1908–1912
13. Mulders, S. M., Preston, G. M., Deen, P. M. T., Guggino, W. B., van Os, C. H., and Agre, P. (1995) *J. Biol. Chem.* **270,** 9010–9016
14. Maurel, C., Reizer, J., Schroeder, J. I., and Chrispeels, M. J. (1993) *EMBO J.* **12,** 2241–2247
15. Heller, K. B., Lin, E. C. C., and Wilson, T. H. (1980) *J. Bacteriol.* **144,** 274–278
16. Maurel, C., Reizer, J., Schroeder, J. I., Chrispeels, M. J., and Saier, M. (1994) *J. Biol. Chem.* **269,** 11869–11872
17. Csonka, L. N., and Hanson, A. D. (1991) *Annu. Rev. Microbiol.* **45,** 569–606
18. Lucht, J. M., and Bremer, E. (1994) *FEMS Microbiol. Rev.* **14,** 3–20
19. Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T., and Ikemura, T. (1990) *Nucleic Acids Res.* **18,** 2367–2411
20. Preston, G. M., Piazza-Carroll, T., Guggino, W. B., and Agre, P. (1992) *Science* **256,** 385–387
21. Preston, G. M., Jung, J. S., Guggino, W. B., and Agre, P. (1993) *J. Biol. Chem.* **269,** 1668–1673
22. Jung, J. S., Preston, G. M., Smith, B. L., Guggino W. B., and Agre, P. (1994) *J. Biol. Chem.* **269,** 14648–14654
23. Van Aelst, L., Hohmann, S., Zimmermann, F. K., Jans, A. W. H., and Thevelein, J. M. (1991) *EMBO J.* **10,** 2095–2104
24. Weaver, C. D., Crombie, B., Stacey, G., and Roberts, D. M. (1991) *Plant Physiol.* **95,** 222–227
25. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) *Science* **269,** 496–512

[2] M. Chrispeels, personal communication.