Identification of Sequence Determinants That Direct Different Intracellular Folding Pathways for Aquaporin-1 and Aquaporin-4*

Received for publication, January 10, 2000, and in revised form, June 22, 2000
Published, JBC Papers in Press, August 15, 2000, DOI 10.1074/jbc.M000165200

William Foster‡, Andrew Helm‡, Isaiah Turnbull‡, Harnik Gulati‡, Baoxue Yang§, Alan S. Verkman§, and William R. Skach‡

From the Division of Molecular Medicine, Oregon Health Sciences University, Portland, Oregon 97201 and the Cardiovascular Research Unit, University of California, San Francisco, California 94143

Homologous aquaporin water channels utilize different folding pathways to acquire their transmembrane (TM) topology in the endoplasmic reticulum (ER). AQP4 acquires each of its six TM segments via cotranslational translocation events, whereas AQP1 is initially synthesized with four TM segments and subsequently converted into a six membrane-spanning topology. To identify sequence determinants responsible for these pathways, peptide segments from AQP1 and AQP4 were systematically exchanged. Chimeric proteins were then truncated, fused to a C-terminal translocation reporter, and topology was analyzed by protease accessibility. In each chimeric context, TM1 initiated ER targeting and translocation. However, AQP4-TM2 cotranslationally terminated translocation, while AQP1-TM2 failed to terminate translocation and passed into the ER lumen. This difference in stop transfer activity was due to two determinants within the TM3–4 peptide loop that enabled AQP4-TM3 but not AQP1-TM3 to reinitiate translocation and cotranslationally span the membrane. Based on these findings, it was possible to convert AQP1 into a cotranslational biogenesis mode similar to that of AQP4 by substituting just two peptide regions at the N terminus of TM2 and the C terminus of TM3. Interestingly, each of these substitutions disrupted water channel activity. These data thus establish the structural basis for different AQF folding pathways and provide evidence that variations in cotranslational folding enable polytopic proteins to acquire and/or maintain primary sequence determinants necessary for function.

Aquaporins comprise a conserved family of membrane proteins that form water- and/or solute-selective channels in cellular membranes (1, 2). Hydropathy analysis, topologic studies and cryoelectron diffraction experiments indicate that aquaporins contain a hydrophobic core region with six transmembrane (TM) helices (3–6). Ten mammalian aquaporins have been identified to date. They exhibit distinct cellular and subcellular expression patterns and play important roles in regulating water homeostasis under a variety of normal and pathologic conditions (7–12). In cell membranes, aquaporins exist primarily as higher ordered structures. AQP1 forms stable tetramers, with each monomer comprising a functional channel (13–15). A closely related aquaporin, AQP4, also forms tetramers which are organized into large macromolecular arrays (16, 17). Both AQP1 and AQP4 form high capacity water-selective channels, suggesting that they share similar overall structural features (18).

Eukaryotic aquaporins are synthesized and assembled in the rough endoplasmic reticulum (ER). During this process TM segments are oriented with respect to the ER membrane and integrated into the lipid bilayer (19, 20). These early biogenesis events are likely mediated via specific interactions between the Sec61 ER translocation machinery and topogenic sequence determinants (e.g. signal and stop transfer sequences) encoded within the nascent polypeptide (21–25). Consistent with this, topogenic determinants have been identified in AQP1 and AQP4 and shown to direct specific translocation and integration events (19, 20).

Despite their homology and similar function, AQP1 and AQP4 encode topogenic determinants with very different translocation properties. AQP4 contains three internal signal anchor sequences and three stop transfer sequences that sequentially initiate and terminate polypeptide translocation as the nascent chain emerges from the ribosome (20). These determinants thus establish six TM segments in a vectoral and cotranslational manner. AQP1, in contrast, encodes only two distinct signal sequences (TM1 and TM5) and two stop transfer sequences (TM3 and TM6) that cotranslationally establish four TM segments (19). AQP1-TM2 and TM4 are initially directed into the ER lumen and cytosol, respectively. Recent studies have revealed that this four membrane-spanning structure represents a folding intermediate that undergoes a topological reorientation during and following the completion of AQP1 synthesis (26). Thus, whereas mature AQP1 and AQP4 both exhibit a similar six membrane-spanning topology, the translocation, integration, and folding events that establish this topology are quite different.

To determine why closely related proteins utilize different folding pathways, we used a series of chimeric AQP1-AQP4 proteins (diagrammed in Fig. 1) and compared the sequence determinants responsible for directing specific translocation events. Topologic analysis identified differences in two peptide regions. Hydrophilic residues flanking the N terminus of...
AQP1-TM2 (Asn\(^{49}\) and Lys\(^{51}\)) prevented TM2 from functioning as a stop transfer sequence and terminating translocation. Similarly, residues within the TM3–4 peptide loop enabled AQP4-TM3 but not AQP1-TM3 to reinitiate translocation and cotranslationally span the membrane. Exchanging only these two regions in AQP1 with corresponding residues from AQP4 converted AQP1 to a cotranslational biogenesis pathway. However, the resulting chimeras lacked water channel activity. Taken together, our results demonstrate how minor sequence variations can significantly affect polytopic protein folding and provide an example in which a novel posttranslational biogenesis pathway is utilized in order to conserve structural determinants which are critical for function.

**MATERIALS AND METHODS**

**cDNA Construction—** AQP chimeras were engineered using the polymerase chain reaction (PCR) overlap extension method (27) (Vent DNA polymerase, New England Biolabs, Beverly, MA). Complimentary oligonucleotides were designed to span the desired fusion site between AQP1 and AQP4 proteins (Table I). These oligonucleotides together with flanking 5′ and 3′ primers were used to perform two initial PCR reactions using template plasmids pSP64.CHIP28 (AQP1) and SP64T-MIWC (AQP4) described previously (19, 20). The overlapping PCR fragments were then hybridized and used as templates for a third PCR amplification, thus “ligating” AQP1 and AQP4 sequences at the fusion site. Products of PCR 3 were digested with restriction enzymes (NcoI and AseI for AQP1, and HindIII and BstXI for AQP4) and ligated into appropriately digested and phosphatase-digested pSP64.CHIP28 and SP64T-MIWC vectors to regenerate full-length proteins. For constructs where internal peptide regions were mutated or exchanged, this process was repeated using wild type or mutant AQP chimeras as the DNA template. Specific residues exchanged are indicated in Table I, and locations of each fusion site are shown schematically in Fig. 1.

Fusion proteins containing the C-terminal translocation reporter were generated by amplifying wild type, mutant, or chimeric AQP coding sequences (10–15 cycles of PCR) using a sense oligonucleotide (SP6 promoter) and antisense oligonucleotides encoding a BstEII restriction site at AQP4 codons Val\(^{52}\), Pro\(^{77}\), Lys\(^{92}\), Val\(^{107}\), and Val\(^{139}\). Antisense oligonucleotides encoding a coding sequences (10–15 cycles of PCR) using a sense oligonucleotide were generated by amplifying wild type, mutant, or chimeric AQP templates. Specific residues exchanged are indicated in Table I, and locations of each fusion site are shown schematically in Fig. 1.

**Chimeric constructs**

Chimeric constructs are referred to by the region of AQP exchanged followed by the parent AQP protein. Fusion sites are indicated by l. Shown in parentheses are peptide regions that have been exchanged for internal segments. For example, AQP4-TM1/AQP1 indicates that AQP1 residues from the N terminus through TM1 have been replaced with corresponding residues from AQP4; AQP4/TM2/AQP1 indicates that AQP1-TM2 has been replaced with AQP4-TM2. Specific fusion sites designed in PCR oligonucleotides are shown at right. ECL-1 refers to the first extracellular loop.

| Plasmid name | AQP residues exchanged (Deleted → replaced with) | Sequence of fusion site (sense oligonucleotide) |
|-------------|--------------------------------------------------|-----------------------------------------------|
| AQP chimeras | AQP1/AQP4                                       | AQP4/AQP1                                      |
| AQP4-TM1/AQP1 | AQP4 M1-F35 → AQP4 M1-S40                         | AACTGGGGTGCCCTCA/AAATACCCGGTGGGG |
| AQP4-TM1-ECL1/AQP1 | AQP4 M1-K51 → AQP4 M1-D47                        | CCCCACCTCCGGCCG/CTGGCCTCGGCTCTT |
| AQP4-TM1-2/AQP1 | AQP4 M1-A73 → AQP4 M1-A72                         | CATCACAGCGTGCC/ACCTCAACCGGCCT |
| AQP1-TM1/AQP4 | AQP4 M1-S40 → AQP1 M1-F35                         | TCTGCCCTCGGCTTC/GAGAACCCCCCTACCT |
| AQP1-TM1-ECL1/AQP4 | AQP4 M1-D47 → AQP1 M1-K51                       | CAGGACACACGGTGAAG/ATGGTCCTCATCTCC |
| AQP1-TM1-2/AQP1 | AQP4 M1-G72 → AQP1 M1-A73                         | CATCACAGCGTGCC/ACCTCAACCGGCCT |
| AQP4/ECL1-TM2/AQP1 | AQP1 K36-A73 → AQP4 E41-G72                 | GGCTTC/GAGAAC. . . . . . . . GTGGCC/CACCTC |
| AQP4/TM2/AQP1 | AQP4 V52-A73 → AQP4 M48-G72                       | GTGAAG/ATGGTC. . . . . . . . GTGGCC/CACCTC |
| AQP4/T115-V140/AQP1 | AQP4 T116-A141 → AQP4 T115-V140              | GGCATC/ACCCAC. . . . . . . . CTGGTG/GAGATC |
| AQP1/TM2/AQP4 | AQP4 M48-G72 → AQP1 V52-A73                       | AQP4/AQP1                                      |
| Site-directed mutagenesis | AQP4 (N130M/K51)                                  | GAGACATGCTAAGATGTCCTGGCTGCGGCTCGG |
| AQP1 (LG35/C76/R58K) |                                     | GACTGGTGCAAGGGAAAGACTACATCTTCTGCAGC |
| AQP1 (G132N/N143T) |                                      | GATAACGTTAGATGTCCTGGCAGGCGTCG |
| AQP4 (N131Q) |                                              | GAGACAGCTACGAGCAGCCATGGCCTCCTGGTG |

**Fig. 1.** Plasmid construction schema. A, schematic representation of AQP1 and AQP4 indicating the location of key residues involved in constructing mutant, chimeric, and truncated fusion proteins. Predicted TM segments are represented by shaded ovals. Upper panel shows both the immature four membrane-spanning and mature six membrane-spanning topologies for AQP1. For convention, AQP1 polypeptide is shown in black, and AQP4 polypeptide is shown in gray. B, boundaries of TM segments and peptide loops exchanged in chimeric proteins are indicated for comparison. ECL-1 refers to the first extracellular loop.
FIG. 2. Protease protection of AQP1 and AQP4 C-terminal fusion proteins. A, location of truncation sites and fusion to the C-terminal P reporter. B and C, protease protection of fusion proteins derived from truncated wild type AQP4 (B) and AQP1 (C). Constructs were expressed in rabbit oocyte lysate supplemented with CRM and digested with PK in the presence or absence of Triton X-100 (det) prior to SDS-PAGE analysis. Arrows indicate glycosylated (lane 1) and protease-protected (lanes 2, 5, 8, and 14) polypeptides. Topology of fusion sites are indicated to the right of autoradiograms. Shaded ovals represent predicted transmembrane segments. AQP4 polypeptide is shown in gray; AQP1 polypeptide is shown in black. Utilized glycosylation sites are indicated as open circles.

PCR overlap extension were verified by DNA sequencing. Truncated constructs were verified by extensive restriction digestion and/or sequencing.

In Vitro Transcription, Translation, and Protease Digestion—mRNA was transcribed with SP6 RNA polymerase (Epicenter, Technologies, Madison, WI) using 2 μg of plasmid DNA in a 10-μl volume at 40 °C for 1 h as described (19). Transcription mixture was added directly to translation mixture containing [35S]methionine, 40% rabbit reticuloocyte lysate (29), and canine pancreas rough microsomal membranes prepared as described (30). Translation was carried out for 1 h at 24 °C (19, 29). Final concentration of membranes was A260 = 8.0, which resulted in >90% translocation and ~80% glycosylation of control proteins (31). For protease digestion, the translation mixture was aliquoted on ice, and CaCl2 was added to 10 mM final concentration (19). Protease K (PK) was then added (0.2 mg/ml) in the presence or absence of 1% Triton X-100. Samples were incubated on ice for 1 h, and residual protease was inactivated by rapid mixing with phenylmethylsulfonyl fluoride (10 mM) and heating to 100 °C in 10 volumes of 1% SDS, 0.1 M Tris, pH 8.0, for 5 min. Samples were then added directly to SDS loading buffer and analyzed by SDS-PAGE.

Oocyte Water Permeability—Stage VI Xenopus laevis oocytes were harvested and maintained as described previously (32). cRNA was transcribed from linearized plasmids using SP6 polymerase and microinjected into oocytes. Water permeability was determined 24–48 h following injection using a hypotonic swelling assay (32). Oocyte swelling was measured in response to a 20-fold dilution of the extracellular medium. Translocation efficiency was determined by correcting for the fractional methionine content of protease-protected peptide fragments relative to full-length polypeptides. Figs. were prepared using an Agfa Studio Scan II transmission scanner and Adobe Photoshop software.

RESULTS

AQP1 and AQP4 Topogenic Determinants Direct Different Cotranslational Topology—To compare the cotranslational translocation events that generate AQP1 and AQP4 topology, polypeptides were truncated at homologous sites following TM1, TM2, or TM3 and ligated to a C-terminal translocation reporter (P) derived from bovine prolactin (~15 kDa in size). Fusion sites and resulting constructs are diagrammed in Figs. 1 and 2A. Topology of the reporter was then determined by protease accessibility following expression in rabbit reticuloocyte lysate supplemented with canine pancreas rough microsomes (CRM). Under these conditions, the reporter is protected from exogenous protease only when it resides in the ER lumen and only in the absence of detergent (28, 33). Because the reporter domain encodes no intrinsic topogenic information and faithfully follows topogenic signals (19, 28), it thus reflects the cotranslational topology of each fusion site at a particular stage of protein synthesis.

Fig. 2 shows topology of the reporter engineered at five sequential fusion sites in AQP4 (gray) and AQP1 (black). Consistent with previous studies, each construct truncated after TM1 (residues Val44–Val72 (AQP4) or Val50–Val78 (AQP1)) generated two protease-protected fragments (Fig. 2, B and C, lanes 1–3, downward arrows). For AQP4, we previously showed that the larger fragment represents the full-length construct, while the smaller fragment resulted from an internal cleavage event, likely at a cryptic signal peptidase recognition site (20). Similarly, the three bands generated from plasmid AQP1.52.P represent full-length glycosylated, full-length non-glycosylated, and signal peptide-cleaved polypeptides (Fig. 2C, lane 1, top, middle, and bottom bands, respectively) (19). Based on the efficiency of PK protection, TM1 C-terminal flanking residues resided in the ER lumen in greater than 70% of AQP1 and AQP4 nascent chains. Non-glycosylated AQP1 chains were not protected and represent polypeptides that either failed to target to the ER or failed to effectively translocate the reporter.
Previous experiments have confirmed that the N termini of both AQP1 and AQP4 reside in the cytosol, although they are inaccessible to protease (19, 20).

AQP4 fusion sites at residues Gly^{72} and Lys^{92} in AQP4, resided in the cytosol and were protease accessible (Fig. 2B, lanes 4–9). Residue Val^{129}, at the C terminus of TM3, was also initially cytosolic (Fig. 2B, lanes 10–12). However, after synthesis of 20 additional residues, the fusion site at Val^{140} was translocated into the ER lumen in >50% of nascent chains as demonstrated by N-linked glycosylation at residue Asn^{131} and generation of a protease-protected, 28-kDa glycosylated polypeptide fragment (Fig. 2B, lanes 13 and 14, downward arrows). Glycosylation was confirmed by translation in the absence of CRM and/or in the presence of a tripeptide inhibitor of oligosaccharyltransferase (Ref. 20 and data not shown). Note that PK cleavage within the TM2–3 connecting loop would be expected to generate a fragment containing the 15-kDa P reporter, 3 kDa of N-linked carbohydrates, and 7 kDa of AQP4 polypeptide. Thus the protected fragment migrates slightly slower than its predicted size. The intensity of the fragment reflects that it contains only 44% of initial methionine residues. Taken together, these results support a cotranslational biogenesis model in which AQP4-TM1 targets the nascent chain-ribosome complex to the ER membrane and initiates translocation; TM2 terminates translocation, and TM3 (together with its C-terminal flanking residues) reinitiates translocation and spans the membrane in a type II topology.

In contrast to AQP4, AQP1 fusion sites within the TM2–3 peptide loop (residues Pro^{77} and Arg^{90}) were directed into a protease-protected environment (Fig. 2C, lanes 4–9). Both uncleaved as well as signal peptidase-cleaved nascent chains remained protected from PK digestion, indicating that AQP1-TM2 failed to terminate translocation. Moreover, following the synthesis of TM3, residues Val^{107} and Leu^{139} both remained accessible to protease, indicating that TM3 spanned the membrane with its N terminus in the ER lumen and C terminus in the cytosol (Fig. 2C, lanes 10–15). Thus as AQP1 topogenic determinants emerge from the ribosome, they direct the nascent chain into a very different transmembrane orientation than corresponding determinants encoded within AQP4. Rather than spanning the membrane in a type I topology, AQP1-TM2 passes into the ER lumen. AQP1-TM3 then terminates translocation and initially spans the membrane in a type I topology (see diagrams in Fig. 2, B and C).

**Effect of AQP4 Residues on AQP1 Topology**—Previous studies of AQP1 and AQP4 have demonstrated no difference in the cotranslational topology of peptide loops C-terminal to TM3. However, after synthesis of TM3, residues Val^{107} and Leu^{139} both remained accessible to protease, indicating that TM3 spanned the membrane with its N terminus in the ER lumen and C terminus in the cytosol (Fig. 2C, lanes 10–15). Thus as AQP1 topogenic determinants emerge from the ribosome, they direct the nascent chain into a very different transmembrane orientation than corresponding determinants encoded within AQP4. Rather than spanning the membrane in a type I topology, AQP1-TM2 passes into the ER lumen. AQP1-TM3 then terminates translocation and initially spans the membrane in a type I topology (see diagrams in Fig. 2, B and C).

**Effect of AQP4 Residues on AQP1 Topology**—We next tested the reciprocal effects of replacing AQP4 residues with corresponding residues from AQP1. Exchanging TM1 or TM1-ECL1 had no effect on AQP4 topology (Fig. 4, A and B). In each case TM2 terminated translocation, and TM3 reinitiated translocation. When TM1 and TM2 were exchanged together, however, TM2 translocated into the ER lumen (Fig. 4C, lanes 1–3). In this context, residue Val^{129} was initially oriented in the cytosol, indicating that AQP4-TM3 could also terminate translocation (Fig. 4C, lanes 4–6). However, after synthesis of TM3 through residue Val^{140}, approximately 33% of nascent chains became doubly glycosylated, indicating that both Asn^{122} and Asn^{135} resided in the ER (Fig. 4C, lanes 7–9). In addition, PK digestion of these latter chains generated the 28-kDa protected fragment, demonstrating that TM3 spanned the membrane in a type II topology (Fig. 4C, lane 8). Thus, TM3 not only translocated its C-terminal residues into the ER lumen, it also reoriented N-terminal flanking residues from the ER lumen back into the cytosol. Of note, translocation was less efficient (<50%) than in wild type AQP4 (Fig. 2B), suggesting that TM3 topology was at least partially dependent on whether or not TM2 terminated translocation.

**Role of TM2 in Directing AQP Topology**—Results from Figs.
Ala73 rather than Gly72 as a result of the
ments.

Diagrams
dicated as in Fig. 3.

sites listed in Table I.

Interestingly, in all contexts where TM2 terminated transloca-

2

consensus site at the corresponding location in AQP1 (G132N,

Asn49 and Lys51 in AQP1 versus Met48 and Leu50 in AQP4 (Fig.

8)

lanes 1–3

lanes 10–12

lanes 16–21

lanes 7–9


case in which it is presented. We noted that TM2 sequences

are highly homologous with the exception of two residues at the

N terminus of the predicted membrane-spanning segment, Asn49

and Lys51 in AQP1 versus Met48 and Leu50 in AQP4 (Fig.

6A). As a result, the core of AQP1-TM2 is less hydrophobic and

three residues shorter than AQP4-TM2. To test whether these

residues were responsible for differences in TM2 stop transfer

activity, Asn49/Lys51 in AQP1 were mutated to Met48/Leu50

(referred to as ML). Topologic analysis of fusion proteins con-

firmed that AQP1-TM2(ML) efficiently terminated transloca-

ation and exhibited a transmembrane topology identical to that

observed for AQP1-TM3 (Fig. 7, lanes 4–6 and 1–3, respec-

tively). TM3 and its N-terminal flanking residues are highly

homologous with the exception of Cys85, Arg86, and Lys87 in

AQP1, which correspond to Leu85, Cys87, and Gln88 in AQP1

(Fig. 7A). Only weak homology is present in the TM3–4 peptide

loop region (residues Thr116–Ile141 versus Thr115–Val125–Val140).

As shown in Fig. 7B, introduction of L85C/C87R/Q88K (CRK)

mutations into AQP1 had a minor effect on TM3 signal sequence

activity (30% translocation efficiency) (lanes 7–9). Exchange of

the TM3 C-terminal flanking residues increased translocation

efficiency to 50% (lanes 10–12), while exchanging both N- and

C-terminal flanking residues increased translocation efficiency
to 85%, similar to that observed for AQP4 (lanes 13–15). For

unclear reasons, hybrid TM3 constructs were cleaved to vari-

ous extents by signal peptidase, but cleavage did not correlate

with translocation efficiency. We also noted that AQP4 encodes

a glycosylation site 17 residues downstream of TM3, whereas

AQP1 does not. Removal of the glycosylation consensus site

from AQP4 (mutation N131Q) had only a minor inhibitory

effect on TM3 signal sequence activity, and introduction of a

consensus site at the corresponding location in AQP1 (G132N,

N134T) had essentially no effect (lanes 16–21). Thus structural

determinants within the TM3–4 peptide loop rather than gly-

cosylation per se appear to be responsible for differences in

TM3 topology.

TM3 C-terminal Flanking Residues Play a Key Role in AQP1

Biogenesis—To confirm the importance of TM3 flanking resi-
dues in AQP1 biogenesis, full-length AQP1 constructs contain-
ing the CRK mutations and/or the TM3–4 peptide loop substi-
tution were used to generate fusion proteins. As expected, the

CRK mutations alone did not effect TM3 topology (Fig. 8A).

In the presence of the ML substitution, where AQP1-TM2 termi-
nates translocation, CRK-TM3 reinitiated translocation in only
25% of nascent chains (Fig. 8B). Remarkably, exchange of the TM3–4 peptide loop resulted in glycosylation of residue Asn131 and increased protease protection of the P-reporter (60%) both in the absence and presence of the CRK mutations (Fig. 7C, lanes 1–3 and 4–6, respectively). In combination with ML substitutions, insertion of AQP4 residues Thr115–Val140 again resulted in glycosylation and translocation of the TM3–4 peptide loop, and translocation was slightly enhanced by the CRK substitutions (80% glycosylation and 70% protection of the reporter (compare Fig. 8D with Fig. 6B). For unclear reasons, the TM2–3 peptide loop was not accessible to protease in these latter constructs even though TM2 spanned the membrane. As was observed for TM3 alone, the N-linked glycosylation site had essentially no effect (Fig. 6E). Taken together, these results indicate that TM3 flanking residues, particularly within the TM3–4 peptide loop, play an important role in directing the cotranslational topology of TM3 during AQP1 synthesis.

Functional Consequences of AQP1 Biogenesis Pathway—Sequence variations at the N terminus of TM2 and the C terminus of TM3 play a major role in directing AQP translocation events. We therefore tested whether these peptide regions might also be involved in other aspects of AQP1 physiology. Water channel activity of full-length, chimeric, and mutant AQP proteins was examined in microinjected X. laevis oocytes using an osmotic induced swelling assay (see “Materials and Methods”). Results shown in Fig. 9 demonstrate that all substitutions affecting AQP1 biogenesis events (e.g. substituting TM2, introducing ML mutations, or exchanging the TM3–4 peptide loop) reduced oocyte water permeability to base-line level. In contrast, CRK substitutions or exchange of TM1-ECL1, neither of which significantly affected AQP biogenesis, had only a partial effect on activity, reducing the Pf to roughly 40% of wild type. These results demonstrate that sequence variations responsible for different biogenesis pathways also play important roles in the overall acquisition of water channel function.

DISCUSSION

Previous studies have demonstrated that AQP1 and AQP4 exhibit different initial topologic conformations in the ER membrane (19, 20). We now define the translocation events and identify specific sequence variations that give rise to these alternate structures. During AQP synthesis, the first TM segment targets the nascent chain-ribosome complex to the ER membrane and initiates polypeptide translocation. For AQP4, topology of TM2 and TM3 is then acquired sequentially via independent stop transfer and signal anchor activities. For AQP1, however, TM2 transiently passes into the ER lumen and TM3 terminates translocation and initially spans the membrane in a Type I topology. As a result, AQP1 cotranslationally acquires only four transmembrane segments, and TM2 and TM3 must be posttranslationally reoriented during subsequent folding events (26). Analysis of AQP1/AQP4 chimeras identified specific residues at the N terminus of TM2 and the C terminus of TM3 that directly influenced TM2 stop transfer activity and TM3 signal sequence activity. Exchanging only these two peptide regions enabled AQP1 to cotranslationally acquire its mature, six membrane-spanning topology. However, substitutions that altered AQP1 biogenesis events also disrupted water channel function. This suggests that early
events of polytopic protein topogenesis are, in part, constrained
by structural features needed for later aspects of protein mat-
uration and/or function.

In the endoplasmic reticulum, initial protein topology is es-

tablished through interactions between topogenic determi-
nants, the ribosome and ER translocation machinery (21, 23–
25). Failure of TM2 to terminate translocation during AQP1
biogenesis suggests that TM2 is unable to disrupt the ribo-
some-membrane junction and direct the elongating nascent
chain into the cytosol (23, 24, 37, 38). Consistent with this,
AQP1-TM2 was completely translocated into the ER lumen
when it was independently engineered into an otherwise sec-
tory protein (19). One possibility is that hydrophilic residues
Asn<sub>49</sub> and/or Lys<sub>51</sub> at the N terminus of AQP1-TM2 interfere
with receptor-mediated interactions that are involved in trans-
location termination (39–41). Alternatively, Asn<sub>49</sub>/Lys<sub>51</sub>
might simply decrease TM2 hydrophobicity below a critical
threshold needed for recognition by the ribosome and/or translocation
channel (31, 37, 42). Because AQP1-TM2 fails to terminate
translocation, TM3 likely emerges from the ribosome into an
open translocation channel where it functions as a stop trans-
fer sequence. This behavior is consistent with other topogenic
determinants whose function may be influenced by their mode
of presentation to ER translocation machinery (31, 33, 43, 44).
However, AQP1-TM3 also failed to achieve its expected type II
topology when it was preceded by a TM segment with stop
transfer activity (e.g. AQP4-TM2). The inability of TM3 to
function as a signal (anchor) sequence, together with the lack of
TM2 stop transfer activity, therefore explains why neither TM
segment cotranslationally achieves its proper topology during
AQP1 biogenesis.

Surprisingly, when AQP4-TM3 was engineered downstream
of AQP1-TM2, it not only translocated its C-terminal flanking
residues into the ER lumen but also oriented its N terminus
toward the cytosol (Fig. 4). Thus under appropriate condi-
tions, a strong signal anchor sequence can both initiate
polypeptide translocation into the ER lumen and also direct a
translocated peptide loop back into the cytosol. This finding is
similar to that recently observed by Goder et al. (45) and
provides further evidence that the translocation channel can
simultaneously accommodate and integrate topogenic informa-
tion encoded within multiple TM helices (25). Further studies
are needed to determine how these unusual topogenic determi-
nants coordinate the gating of both ends of the translocation
channel without mixing cytosolic and ER lumenal contents (37,
46).

For both AQP1 and AQP4, TM1 efficiently targets the nas-
cent chain to the ER membrane and initiates translocation.
Because TM1 sequences did not influence the subsequent to-
pogenic behavior of TM2 and TM3, initial membrane targeting
events do not appear to contribute significantly to different
biogenesis pathways. Moreover, once the ribosome is targeted
to the ER membrane it remains docked during the synthesis
of short cytosolic loops (23, 44, 47). It is therefore likely that TM3

---

**Fig. 8. Role of TM3 flanking residues in directing AQP1 topology.** Diagrams above autoradiograms indicate AQP1 (black) and AQP4 (gray) residues and engineered AQP1 ML and CRK mutations. Truncation sites are shown above autoradiograms. Topology of fusion proteins is shown beneath autoradiograms. **Fig. 9. Water channel activity of mutant and chimeric AQP1 constructs.** Plasmids encoding full-length wild type (WT), mutant, and chimeric AQP1 constructs were linearized, transcribed, and expressed in mature oocytes. Regions of AQP1 that were mutated or replaced with corresponding residues from AQP4 are shown at left. 24 h follow-
ing injection, oocyte water permeability ($P_f$) was determined by osmotic induced swelling in response to a 20-fold dilution of medium. Shown are results from a representative experiment. $n$ = 5–8 oocytes in each group ± S.E.
flanking residues facilitate translocation by augmenting post-targeting events such as Sec61 gating (37, 48–51) and/or lateral exit of the TM segment from Sec61 into the lipid bilayer (52, 53). For several chimeric constructs we found that the P reporter was incompletely protected from protease even in chains where the adjacent N-linked consensus site, Asn131, was glycosylated (Figs. 4 and 5). This observation raises the possibility that AQP1-TM3, and to a lesser extent AQP4-TM3, may fall back into the cytosol because it is unable to stably span the membrane in the absence of TM4-TM6. A similar finding has been observed in experimentally engineered polytopic proteins (43, 45) and may partially explain why glycosylation scanning and protease protection analysis may occasionally yield different topologic results. Interestingly, insertion or removal of the Asn131 glycosylation had little effect on TM3 topology. Thus in contrast to previous studies where glycosylation may favor certain transmembrane orientations by inhibiting retrograde translocation (45, 54), TM3 C-terminal flanking residues appear to influence topology through a direct effect on the ER translocation machinery.

Why then should homologous proteins with similar function and predicted structure utilize different folding pathways? Our results suggest that the answer lies in the close relationship between determinants required for protein function and determinants that influence protein topogenesis. In the simplest case, polytopic protein topology is established cotranslationally from N to C terminus through the sequential action of alternating, independent signal and stop transfer sequences (28, 43, 44, 55). Any residue that interfered with these independent topogenic activities would thus disrupt protein folding. However, variations on this cotranslational model have been increasingly observed in a variety of engineered and naturally occurring polytopic proteins (23, 31, 45, 56–61). For example, translocation across and integration into the bilayer may involve cooperative interactions between topogenic determinants that have either failed to acquire or lost the ability to independently carry out specific biogenesis activities (31, 33, 58, 59, 61–66). Topology may also be established posttranslationally by downstream signal sequences (59, 67, 68) or during late events of protein folding that likely involve intramolecular interactions within the nascent chain itself, as has been demonstrated for Sec61p (60). In the case of AQP1, the topogenic activities of TM2 and TM3 are directly influenced by residues required for protein function. This finding is consistent with the high degree of conservation of the TM2 and TM3 flanking residues across different species (Fig. 10). Similarly, charged residues in the N terminus of the cystic fibrosis transmembrane conductance regulator disrupted TM1 signal sequence activity and required that a posttranslational pathway properly orient TM1 in the membrane (Refs. 59 and 69, and data not shown). Our data thus support a model in which alternate folding pathways provide an evolutionary advantage by enabling polytopic proteins to acquire sequence diversity that would not be possible if topogenesis were restricted solely to cotranslational events.

Our studies provide evidence that transient translocation of cystolic loops occurs in normal and efficient physiologic folding pathways. This finding contrasts recent studies in which unusual or inefficient topogenic determinants have been implicated in protein degradation. In truncated forms of cystic fibrosis transmembrane conductance regulator and the Na,K-ATPase, predicted cystolic peptide loops and/or TM segments translocate into the ER lumen (70, 71). In both cases transient exposure of peptide segments to the luminal environment has been proposed to facilitate recognition by ER quality control machinery. AQP1 is not a substrate for ER associated degradation and it is efficiently converted from an immature four membrane-spanning topology to its mature six membrane-spanning topology in cells (26). However, AQP1 topological reorientation was highly dependent on the synthesis of all six TM segments. Truncated AQP1 proteins containing four or five TM segments were markedly less efficient at reorienting TM2 and TM3 and were rapidly degraded (data not shown). Thus, polytopic protein topogenesis may involve complex folding information, and removing portions of this information may contribute to protein misfolding even among physically distant TM segments. Understanding how intramolecular interactions within the nascent polypeptide as well as intermolecular interactions with the ER translocation machinery determine the final topology and ultimate fate of polytopic membrane proteins remains a significant challenge in this area.

Acknowledgments—We thank K. Moss and B. Watson for technical assistance, and Dr. L. Musil and members of the Skach laboratory for helpful comments.

REFERENCES
1. Agre, P., Bonhivers, M., and Borgia, M. J. (1998) J. Biol. Chem. 273, 14659–14662
2. Verkman, A., van Hoek, A., Ma, T., Frigeri, A., Skach, W., Mitra, A., Tamarrappo, B., and Farinas, J. (1996) Am. J. Physiol. 270, C12–C30
3. Preston, G. M., and Agre, P. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 11110–11114
4. Preston, G., Jung, J., Guggino, W., and Agre, P. (1994) J. Biol. Chem. 269, 1668–1673
5. Cheng, A., van Hoek, A. N., Yeager, M., Verkman, A. S., and Mitra, A. K. (1997) Nature 387, 627–630
6. Walz, T., Hiraim, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P., and A. E. (1997) Nature 387, 624–627
7. Frokiaer, J., Marples, D., Knepper, M., and Nielsen, S. (1998) Am. J. Med. Sci. 316, 291–299
8. Nielsen, S., Kwon, T-H., Christensen, B., Promeneur, D., Frokiaer, J., and Marples, D. (1998) J. Am. Soc. Nephrol. 9, 1047–1063
9. Verkman, A., and A., E. (1997) J. Biol. Chem. 272, 624–627
10. Schnermann, J., Chou, J., Ma, T., Knepper, M., and Verkman, A. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 9660–9664
11. Brown, C., Katsura, T., Kawashima, M., Verkman, A., and Saholic, I. (1995) Histochem. Cell Biol. 104, 1–9
12. King, L., and Agre, P. (1996) Annu. Rev. Physiol. 58, 619–648
13. Shi, L.-B., Skach, W., and Verkman, A. (1994) J. Biol. Chem. 269, 10417–10422
14. Smith, B. L., and Agre, P. (1991) J. Biol. Chem. 266, 6407–6415
15. Verkman, A., van Hoek, D., Sable, I., Valenti, G., van Hoek, A. N., Ma, T., and Verkman, A. S. (1993) J. Cell Biol. 125, 605–618
16. Yang, B., Brown, D., and Verkman, A. (1996) J. Biol. Chem. 271, 4577–4580
17. Verkman, A., and Verkman, A. K. (1997) Cell 97, 2855–60
18. Yang, B., and Verkman, A. (1997) J. Biol. Chem. 272, 16140–6
19. Skach, W., Shi, L.-B., Calayag, M. C., Frigeri, A., Lingappa, V., and Verkman, A. (1994) J. Cell Biol. 125, 805–815
20. Shi, L.-B., Skach, W., and Verkman, A. (1995) Biochemistry 34, 8250–8256
21. Laird, V., and High, S. (1997) Trends. Cell Biol. 7, 1983–1989
22. Johnson, A. (1997) Trends Cell Biol. 7, 90–94
23. Mothes, W., Heinrich, S., Graf, R., Nilsson, I., von Heijne, G., Brunner, J., and Rapoport, T. (1997) Cell 95, 523–533
24. Biel, E. (1998) Trends Biochem. Sci. 23, 51–55
25. Hegde, R., and Lingappa, V. (1999) J. Biol. Chem. 274, 10245–10250
26. Lu, Y., Turnbull, I., Bragin, A., Carveth, K., Verkman, A., and Skach, W. (2000) Mol. Biol. Cell, 11, 2973–2985
27. Ho, N., Hunt, H., Horton, R., Pullen, J., and Pease, R. (1989) Gene (Amst.) 77, 51–59
28. Rothman, R. E., Andrews, D. W., Calayag, M. C., and Lingappa, V. R. (1988) J. Biol. Chem. 263, 10470–10489
Determinants of AQP1 and AQP4 Topogenesis

29. Skach, W. (1998) Methods Enzymol. 292, 265–277
30. Walter, P., and Blobel, G. (1983) Methods Enzymol. 96, 4–93
31. Moss, K., Helm, A., Lu, Y., Bragin, A., and Skach, W. (1998) Mol. Biol. Cell 9, 2681–2097
32. Zhang, R., and Verkman, A. S. (1991) Am. J. Physiol. 260, 26–34
33. Skach, W., and Lingappa, V. (1994) Cancer Res. 54, 1329–1329
34. Nilsson, I., and von Heijne, G. (1993) J. Biol. Chem. 268, 5798–5801
35. Popov, M., Tam, L., Li, J., and Reithmeier, R. (1997) J. Biol. Chem. 272, 18325–18332
36. Landolt-Marticorena, C., and Reithmeier, R. (1994) Biochim. Biophys. Acta 1202, 184–194
37. Liao, S., Lin, J., Do, H., and Johnson, A. (1997) J. Biol. Chem. 272, 18325–18332
38. Hegde, R., Voigt, S., Rapoport, T., and Lingappa, V. (1998) Cell 92, 621–631
39. Yost, C. S., Lopez, C. D., Prusiner, S. B., Meyers, R. M., and Lingappa, V. R. (1990) J. Cell Biol. 116, 801–808
40. Kuroiwa, T., Sakaguchi, M., Omura, T., and Mihara, K. (1996) J. Biol. Chem. 271, 6423–6428
41. Xiong, X., Bragin, A., Widdicombe, J., Cohn, J., and Skach, W. (1997) J. Cell Biol. 139, 725–736
42. Tector, M., and Hartl, F. (1999) EMBO J. 18, 6290–6298
43. Beguin, P., Hasler, U., Staub, O., and Geering, K. (2001) Mol. Biol. Cell 11, 1657–1672