

## Divergent evolution of membrane protein topology: The *Escherichia coli* RnfA and RnfE homologues

ANNIKA SÄÄF\*, MARIE JOHANSSON\*, ERIK WALLIN, AND GUNNAR VON HEIJNE†

Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden

Communicated by Jonathan Beckwith, Harvard Medical School, Boston, MA, May 11, 1999 (received for review March 2, 1999)

**ABSTRACT** Although the molecular evolution of protein tertiary structure and enzymatic activity has been studied for decades, little attention has been paid to the evolution of membrane protein topology. Here, we show that two closely related polytopic inner membrane proteins from *Escherichia coli* have evolved opposite orientations in the membrane, which apparently has been achieved by the selective redistribution of positively charged amino acids between the polar segments flanking the transmembrane stretches. This example of divergent evolution of membrane protein topology suggests that a complete inversion of membrane topology is possible with relatively few mutational changes even for proteins with multiple transmembrane segments.

It is well established that the most important determinant of membrane protein topology, both in prokaryotic and eukaryotic organisms, is the distribution of positively charged residues in the regions flanking the hydrophobic transmembrane (TM) segments (1, 2). This “positive inside” rule states that regions rich in positively charged residues tend to remain nontranslocated as a protein inserts into the membrane and has been found to hold not only for plasma membrane proteins but also for integral membrane proteins from thylakoids and mitochondria (3, 4). Strong support for the positive inside rule has come from protein engineering experiments, which have demonstrated that membrane protein topology can be changed or even fully inverted by moving positively charged residues from cytoplasmic to extra-cytoplasmic regions of the protein (5–8).

Whether such re-engineering of membrane protein topology also happens during evolution is unknown, as no clear example of strongly related polytopic proteins with opposite orientations in the membrane has been found to date. Here, we report that the *Escherichia coli* homologues of the RnfA and RnfE proteins from *Rhodobacter capsulatus* (two proteins belonging to a family of energy-coupling NADH oxidoreductases) form such a pair. They display more than 35% sequence identity across a stretch encompassing five TM segments, and yet have opposite orientations in the inner membrane as deduced from PhoA-fusion analysis. In spite of the high sequence similarity, the positively charged residues are distributed differently in the two proteins, and both proteins follow the positive inside rule. We conclude that evolution can impart different membrane topologies on strongly related proteins by reshuffling of positively charged residues. Our findings also should have implications for the possible function of the RnfA/E complex in electron transport.

### MATERIALS AND METHODS

**Enzymes and Chemicals.** Unless otherwise stated, all enzymes were from Promega. T7 DNA polymerase, *Taq* polymerase, Thermosequenase kit, and [<sup>35</sup>S]methionine were from

Amersham Pharmacia. T4 ligase was from GIBCO/BRL. Oligonucleotides were from CyberGene (Stockholm, Sweden). PhoA antiserum was from 5 Prime→3 Prime. The alkaline phosphatase chromogenic substrate PNPP (Sigma 104 phosphatase substrate) was from Sigma.

**Strains and Plasmids.** Experiments were performed in *E. coli* strains MC1061 [ $\Delta lacX74$ , *araD139*,  $\Delta(ara, leu)7697$ , *galU*, *galK*, *hsr*, *hsm*, *strA*] (9) and CC118 [ $\Delta(ara-leu)7697$   $\Delta lacX74$   $\Delta phoA20$  *galE* *galK* *thi* *rpsE* *rpoB* *argE(am)* *recA1*] (10). All constructs were expressed in *E. coli* from the pING1 plasmid (11) by induction with arabinose.

**DNA Techniques.** All plasmid constructs were confirmed by DNA sequencing using T7 DNA polymerase or the Thermosequenase kit. The *ydqQ* and *ORF193* genes were amplified by PCR from *E. coli* JM109 chromosomal DNA by using *Taq* polymerase or the Expand Long Template PCR system (Boehringer Mannheim). By use of appropriate PCR primers, a 5' *XhoI* and a 3' *KpnI* site were introduced in the regions flanking the amplified genes, and the initiator codon GTG in *ydqQ* was changed to ATG. The PCR products were cleaved with *XhoI* and *KpnI* and cloned behind the *ara* promoter in a *XhoI*–*KpnI*-restricted plasmid derived from pING1 containing a *lep* gene with a 5' *XhoI* site just upstream of the initiator ATG and a *KpnI* site in codon 78. Relevant parts of the *ydqQ* and *ORF193* genes were amplified by PCR from the pING1 plasmid with a 5' *SalI* and a 3' *KpnI* site encoded in the primers. Finally, the PCR *SalI*–*KpnI* fragment carrying the *lep* upstream region and the relevant *ydqQ* or *ORF193* segment was cloned into a previously constructed plasmid (12) carrying a *phoA* gene lacking the 5' segment coding for the signal sequence and the first five residues of the mature protein and immediately preceded by a *KpnI* site. In all constructs, an 18-aa linker (VPDSYTVQVASWTEPFPPFC) was present between the YdgQ or ORF193 part and the PhoA moiety.

**Expression of Fusion Proteins.** *E. coli* strain CC118 transformed with the pING1 vector carrying the relevant construct under control of the arabinose promoter was grown at 37°C in M9 minimal medium supplemented with 100  $\mu$ g/ml ampicillin, 0.5% fructose, 100  $\mu$ g/ml thiamin, and all amino acids (50  $\mu$ g/ml each) except methionine. An overnight culture was diluted 1:25 in fresh medium, shaken for 3.5 h at 37°C, induced with arabinose (0.2%) for 5 min, and labeled with [<sup>35</sup>S]methionine (75  $\mu$ Ci/ml). After 2 min, samples were acid-precipitated with trichloroacetic acid (10% final concentration), resuspended in 10 mM Tris/2% SDS, immunoprecipitated with antisera to PhoA, washed, and analyzed by SDS/PAGE. Gels were scanned in a FUJIX (Tokyo) Bas 1000 PhosphorImager and analyzed by using MACBAS software (version 2.31).

**PhoA Activity Assay.** Alkaline phosphatase activity was measured by growing strain CC118 transformed with the appropriate pING1-derived plasmids in liquid culture for 2 h

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviation: TM, transmembrane.

\*A.S. and M.J. contributed equally to this work.

†To whom reprint requests should be addressed. e-mail: gunnar@biokemi.su.se.

Table 1. BLASTP scan of the GenPept databank at Biology Workbench 3.0 using RnfA from *R. capsulatus* as query

Family	Score, bits	E value
<b>RnfA family</b>		
pir S39895 rnfA protein - <i>R. capsulatus</i>	340	3e-93
*gi 1787914 (AE00258) ORF193 - <i>E. coli</i>	206	1e-52
gi 1574535 (U32841) hypothetical protein - <i>Haemophilus influenzae</i>	199	7e-51
pir S65530 Nqr5 - <i>Vibrio alginolyticus</i>	132	1e-30
gi  1573126 (U32702) hypothetical protein - <i>H. influenzae</i>	131	3e-30
gi 3328694 (AE001300) Nqr5 - <i>Chlamydia trachomatis</i>	120	6e-27
<b>RnfE family</b>		
pir S39906 rnfE protein - <i>R. capsulatus</i>	88	4e-17
sp Q57020 YDGQ_HAEIN - <i>H. influenzae</i>	87	8e-17
*sp P77179 YDGQ_ECOLI - <i>E. coli</i>	82	2e-15
pir S65529 Nqr4 - <i>V. alginolyticus</i>	68	3e-11
gi 3328693 (AE001300) Nqr4 - <i>C. trachomatis</i>	66	2e-10
sp P43958 Y168_HAEIN - <i>H. influenzae</i>	64	6e-10

The E value is an estimate of the number of matches with the same or a higher score expected to be found by chance in the database. No other significant matches were found. The two *E. coli* proteins analyzed in this study are indicated by \*.

in the absence of arabinose and then for 1 h in the presence of 0.2% arabinose (13). Mean activity values were obtained from at least two independent measurements and were normalized by the rate of synthesis of the fusion protein determined by pulse labeling of arabinose-induced CC118 cells as described above. Normalized activities were calculated as:

$$A = (A_0 \times OD_{600} \times n_{Met}) / CPM,$$

where  $A_0$  is the measured activity,  $OD_{600}$  is the cell density at the time of pulse labeling,  $n_{Met}$  is the number of Met residues in the fusion protein, and CPM is the intensity of the relevant band measured on the PhosphorImager.

**Topology Prediction and Sequence Alignment.** Topology predictions were done by using TOPPRED II (14), TMHMM (15), and TOPPRED-ALIGN (a version of TOPPRED that derives a

consensus prediction from a set of aligned sequences; our unpublished work). Sequence alignments were done by using BLASTP (16), CLUSTALW (17), and FASTA (18) as implemented in Biology Workbench 3.0 at <http://biology.ncsa.uiuc.edu/>. Default parameter settings were used in all cases.

## RESULTS

**RnfA and RnfE Are Homologues, Yet Have Opposite Predicted Topologies.** During an attempt to improve a previously developed method for prediction of membrane protein topology, TOPPRED (14, 19), by including information from multiple alignments of related proteins, we noticed that the consensus prediction for a multiple alignment based on the *R. capsulatus* RnfA protein was correct in having six predicted TM segments, but was incorrect in terms of predicted overall orientation in

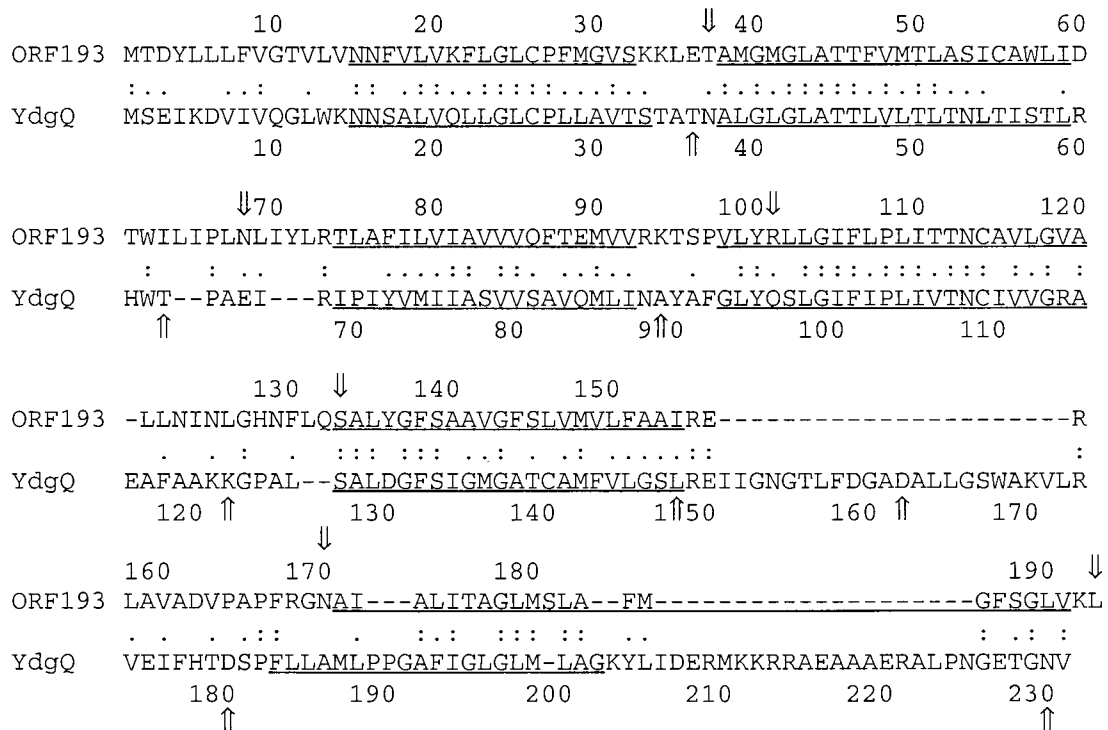


FIG. 1. FASTA alignment of *E. coli* ORF193 (an RnfA homologue) and YdgQ (an RnfE homologue). Identities are indicated by colons (:), and conservative substitutions by periods (.). Hydrophobic TM segments are underlined. The positions of the PhoA fusion joints are shown by arrows.

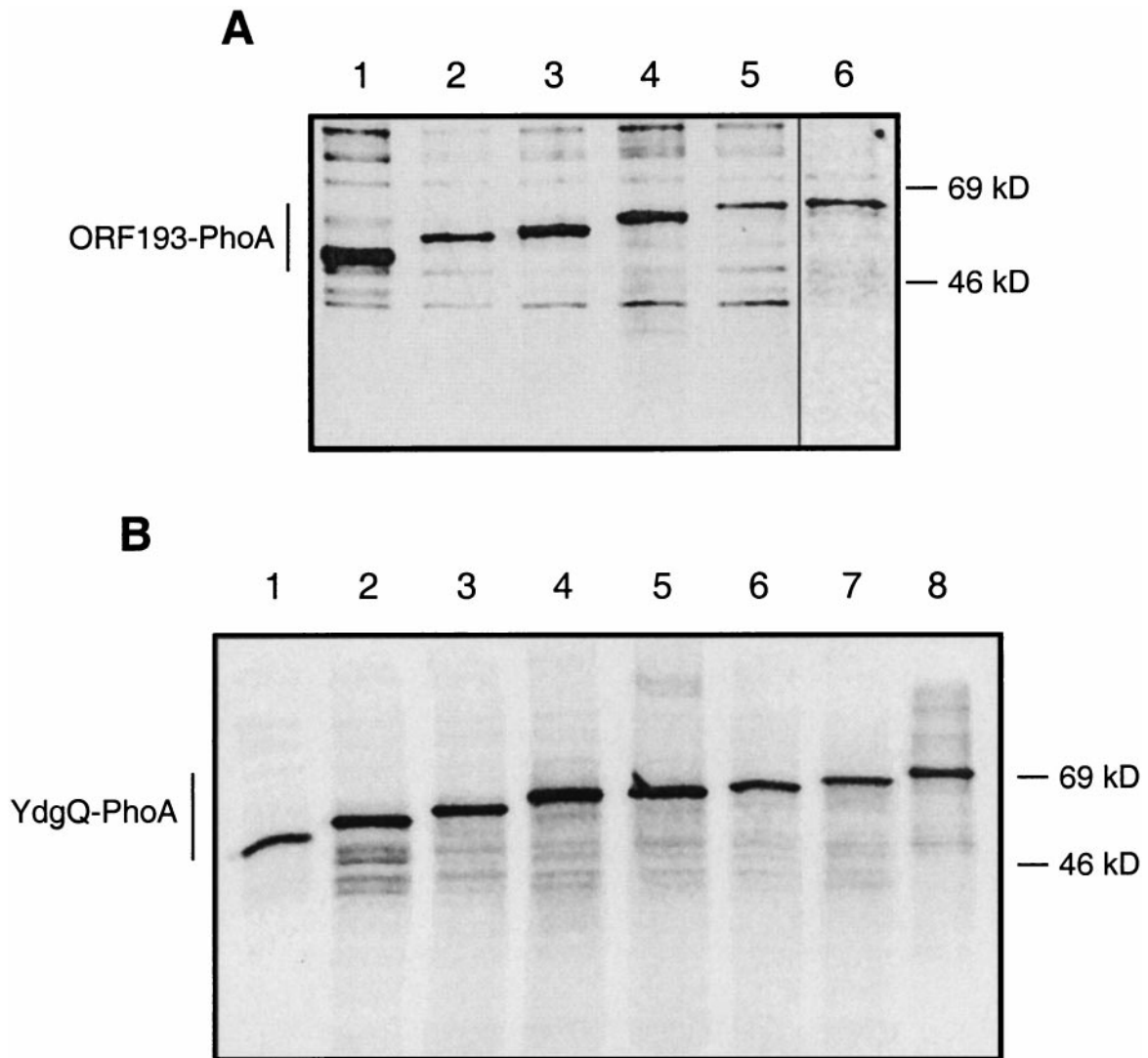


FIG. 2. Expression of PhoA fusions in *E. coli* CC118 cells. Cells were labeled for 2 min by [<sup>35</sup>S]methionine, and fusion proteins were immunoprecipitated by a PhoA antiserum. (A) ORF193-PhoA fusions. Lane 1: ORF193<sup>37</sup>-PhoA; lane 2: ORF193<sup>68</sup>-PhoA; lane 3: ORF193<sup>101</sup>-PhoA; lane 4: ORF193<sup>133</sup>-PhoA; lane 5: ORF193<sup>170</sup>-PhoA; lane 6: ORF193<sup>193</sup>-PhoA. ORF193<sup>193</sup>-PhoA is from a different gel, but its position relative to ORF193<sup>170</sup>-PhoA is correct. (B) YdgQ-PhoA fusions. Lane 1: YdgQ<sup>36</sup>-PhoA; lane 2: YdgQ<sup>63</sup>-PhoA; lane 3: YdgQ<sup>89</sup>-PhoA; lane 4: YdgQ<sup>122</sup>-PhoA; lane 5: YdgQ<sup>148</sup>-PhoA; lane 6: YdgQ<sup>162</sup>-PhoA; lane 7: YdgQ<sup>180</sup>-PhoA; lane 8: YdgQ<sup>230</sup>-PhoA.

the inner bacterial membrane: N<sub>in</sub>-C<sub>in</sub> instead of the experimentally determined N<sub>out</sub>-C<sub>out</sub> orientation (20). A more careful analysis revealed that our BLAST-based selection of proteins to be included in the multiple alignment contained not only the expected RnfA homologues, but also a family of RnfE homologues with highly significant BLAST scores (Table 1). To our surprise, when the topology prediction was carried out only on the RnfA family members (individually or together), the topology predicted from the positive inside rule was correct, whereas the RnfE family members were all strongly predicted to have the opposite N<sub>in</sub>-C<sub>in</sub> topology (data not shown). In the full multiple alignment, the RnfE homologues happened to dominate the prediction, and the "consensus" topology thus also was predicted as N<sub>in</sub>-C<sub>in</sub>.

Further analysis revealed that RnfA and RnfE, as well as their *E. coli* homologues ORF193 (PID:g1787914) and YdgQ (SwissProt:P77179), displayed more than 35% identity throughout a segment encompassing the first five predicted TM segments (up to residue 156 in ORF193) (Fig. 1). Such a high degree of sequence identity generally is taken to imply closely related tertiary structure (21) and, among other things, membrane topology. To resolve this apparent dilemma, we

decided to determine the membrane topology of ORF193 and YdgQ by using the PhoA-fusion approach (22).

**Determination of the Topology of the *E. coli* RnfA and RnfE Homologues by PhoA Fusions.** A series of PhoA fusions were made both to ORF193 and YdgQ. As recommended (23), the fusion joints generally were chosen near the C-terminal end of the putative periplasmic and cytoplasmic loops (Fig. 1). All fusions could be expressed in the *phoA*<sup>-</sup> strain CC118 (10), could be immunoprecipitated by a polyclonal PhoA antiserum, and were of the expected sizes (Fig. 2). Alkaline phosphatase activities were measured according to ref. 13 and are indicated in Fig. 3.

The results for the RnfA homologue ORF193 are in perfect agreement with those of an earlier study where the topology of RnfA was determined by expressing RnfA-PhoA fusions in *R. capsulatus* (20) and provide further support for the proposed six-TM N<sub>out</sub>-C<sub>out</sub> topology. Strikingly, the results for the RnfE homologue YdgQ also indicate a six-TM topology, but with the opposite N<sub>in</sub>-C<sub>in</sub> orientation in the inner membrane. The results are unequivocal for the first five TM segments, but there is some ambiguity concerning the most C-terminal region. A fusion immediately after TM5

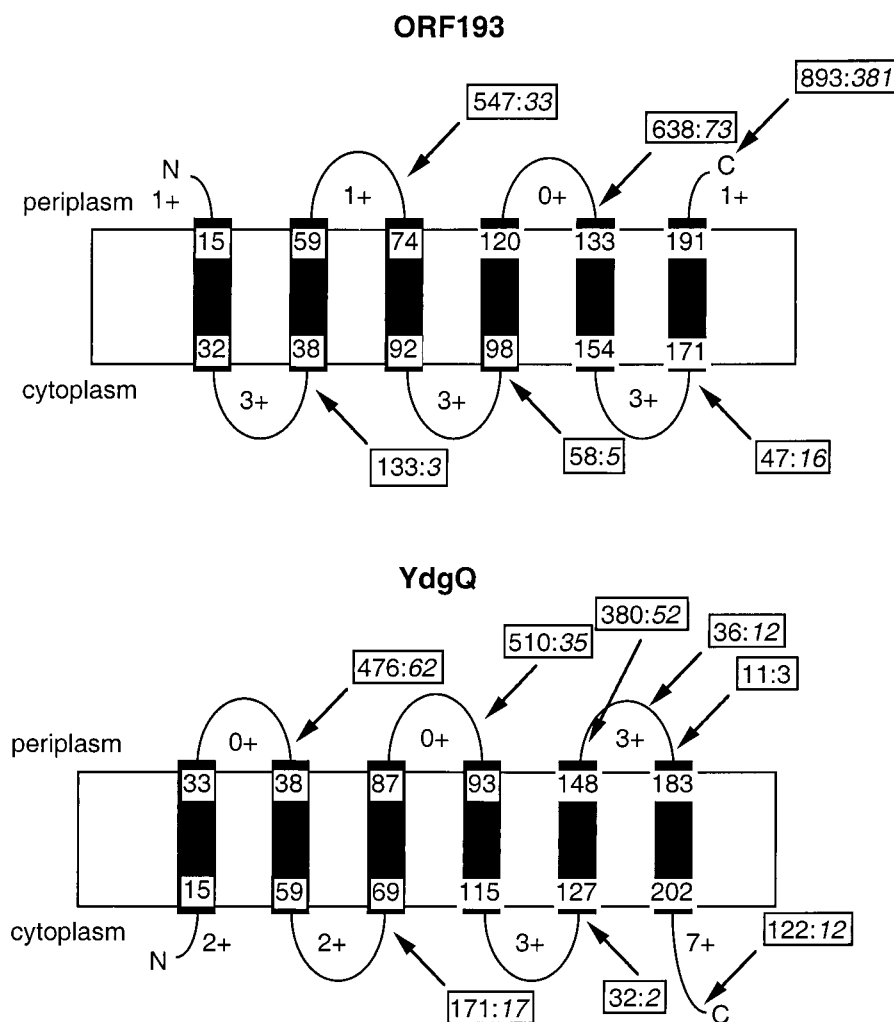


FIG. 3. Topology models for ORF193 (Upper) and YdgQ (Lower). PhoA fusions are indicated by arrows, together with their respective measured and normalized alkaline phosphatase activities (measured/normalized; the latter are expressed in arbitrary units). Approximate ends of the TM segments as well as the number of positively charged residues (Lys+Arg) in the tails and loops are indicated.

has a high activity as expected, but two additional fusions in the putative loop between TM5 and TM6 have a low activity. A fusion to the C-terminal end of YdgQ again has a low activity, suggesting a cytoplasmic location. Considering that the region between TM5 and TM6 contains no strong candidate TM segment (Fig. 1) we feel that the six-TM model shown in Fig. 3 is the most likely one, although we cannot completely rule out that TM6 is located between residues 151 and 172 rather than at the location proposed in the figure. Possibly, the 151–172 segment is just hydrophobic enough to anchor in the membrane in the YdgQ<sup>180</sup>-PhoA fusion but may be “pushed out” to the periplasmic side when followed by the much more hydrophobic 183–202 segment in fusion YdgQ<sup>230</sup>-PhoA.

In any case, considering the high degree of sequence identity in the TM1-TM5 region, we conclude that the ORF193/YdgQ pair has undergone divergent evolution of topology.

## DISCUSSION

It has long been known that positively charged residues are important determinants of membrane protein topology, and hence under selective pressure to maintain the topology required for the proper function of any given protein. The ORF193/YdgQ pair studied here is an example of divergent evolution of topology, where two highly related polytopic proteins have evolved opposite orientations in the membrane.

From the alignment presented in Fig. 1, it is clear that the TM segments are much more conserved (46% sequence identity not counting the region beyond TM5) than the connecting loops and the N- and C-terminal tails (14% sequence identity not counting the region beyond TM5), and that a critical difference between the loops/tails of the two proteins is precisely their content of positively charged residues. In RnfA and its homologues, the TM1/2, TM3/4, and TM5/6 loops are rich in lysine and arginine, whereas in the RnfE family the N and C tails as well as the TM2/3 and TM4/5 loops carry more positive charge. In contrast, there is no consistent difference in the distribution of negatively charged residues between the two families.

Although the exact function of RnfA and RnfE is not known, they are thought to be involved in mediating electron transport between electron transfer systems in the inner membrane and the cytoplasmic nitrogenase system, possibly being components of an energy-dependent ferredoxin reductase (20). Given their opposing topologies, they most likely form an unusual, quasi-symmetrical complex across the inner bacterial membrane, which should have interesting implications for the function of the RnfA/E system.

Assuming that such a symmetrical arrangement is functionally significant, a possible scenario for the evolution of the RnfA/E complex is that an ancestral protein with a balanced distribution of positively charged residues was able to insert with a mixed N<sub>in</sub>-C<sub>in</sub> and N<sub>out</sub>-C<sub>out</sub> topology (as has been shown

for deliberately engineered proteins; refs. 6 and 7) and thus could form a perfectly symmetrical complex. After a gene duplication, the two resulting proteins were free to evolve unique but opposite topologies, possibly allowing further functional refinement of the enzyme complex. In any case, the RnfA/E pair provides a striking example of divergent evolution on the level of protein topology rather than tertiary structure or enzymatic activity.

This work was supported by grants from the Swedish Cancer Foundation, the Swedish Natural and Technical Sciences Research Councils, and the Göran Gustafsson Foundation to G.v.H.

1. von Heijne, G. (1996) *Progr. Biophys. Mol. Biol.* **66**, 113–139.
2. Spiess, M. (1995) *FEBS Lett.* **369**, 76–79.
3. Gavel, Y. & von Heijne, G. (1992) *Eur. J. Biochem.* **205**, 1207–1215.
4. Gavel, Y., Steppuhn, J., Herrmann, R. & von Heijne, G. (1991) *FEBS Lett.* **282**, 41–46.
5. von Heijne, G. (1989) *Nature (London)* **341**, 456–458.
6. Nilsson, I. M. & von Heijne, G. (1990) *Cell* **62**, 1135–1141.
7. Gafvelin, G. & von Heijne, G. (1994) *Cell* **77**, 401–412.
8. Gafvelin, G., Sakaguchi, M., Andersson, H. & von Heijne, G. (1997) *J. Biol. Chem.* **272**, 6119–6127.
9. Dalbey, R. E. & Wickner, W. (1986) *J. Biol. Chem.* **261**, 13844–13849.
10. Lee, E. & Manoil, C. (1994) *J. Biol. Chem.* **269**, 28822–28828.
11. Johnston, S., Lee, J. H. & Ray, D. S. (1985) *Gene* **34**, 137–145.
12. Whitley, P., Nilsson, I. & von Heijne, G. (1994) *Nat. Struct. Biol.* **1**, 858–862.
13. Manoil, C. (1991) *Methods Cell Biol.* **34**, 61–75.
14. Claros, M. G. & von Heijne, G. (1994) *Comput. Appl. Biosci.* **10**, 685–686.
15. Sonnhammer, E., von Heijne, G. & Krogh, A. (1998) *Intell. Syst. Mol. Biol.* **6**, 175–182.
16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
18. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
19. von Heijne, G. (1992) *J. Mol. Biol.* **225**, 487–494.
20. Kumagai, H., Fujiwara, T., Matsubara, H. & Saeki, K. (1997) *Biochemistry* **36**, 5509–5521.
21. Sander, C. & Schneider, R. (1991) *Proteins* **9**, 56–68.
22. Manoil, C. & Beckwith, J. (1986) *Science* **233**, 1403–1408.
23. Boyd, D., Traxler, B. & Beckwith, J. (1993) *J. Bacteriol.* **175**, 553–556.