Mapping of Cystic Fibrosis Transmembrane Conductance Regulator Membrane Topology by Glycosylation Site Insertion*

(Received for publication, March 11, 1994, and in revised form, May 5, 1994)

Xiu-Bao Chang‡, Yue-Xian Hou, Timothy J. Jensen, and John R. Riordan
From the Research Institute, The Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada M5G 1X8

Technical difficulties in obtaining three-dimensional structures of intrinsic membrane proteins continues to limit understanding of their function. However, considerable insight can be gained from their two-dimensional topological arrangement in the lipid bilayer. Efficient molecular genetic approaches are available to discern the topology of prokaryotic but not of eukaryotic membrane proteins. The absolute asymmetry of the sidedness of their N-glycosylation was employed here to develop such a method using the cystic fibrosis transmembrane conductance regulator (CFTR). Insertion by in vitro mutagenesis of N-glycosylation consensus sequences (NXXT) in predicted cytoplasmic and extracytoplasmic loops between hydrophobic sequences capable of traversing the membrane established the membrane topology of CFTR. This provides the first experimental evaluation of the original topological model of CFTR based solely on hydrophathy algorithms and a method which may be generally applicable for the in vivo evaluation of the topology of other mammalian membrane proteins.

The two-dimensional disposition of integral membrane proteins in the lipid bilayer can be readily predicted from their sequences on the basis of hydrophathy algorithms (1, 2). However, although the validity of these predictions can be efficiently tested for prokaryotic proteins by the insertion of sequences coding for enzymes dependent for activity on exposure to the periplasmic space of Gram-negative bacteria (3-9), only more laborious and less definitive biochemical methods are available for eukaryotic membrane proteins. These include the labeling of putative transmembrane sequences (TMs)1 with lipid-soluble reagents (10) and the exposure of potential hydrophobic loops between the TMs to membrane-impermeable reagents (11), sequence-specific antibodies (12), or proteases (13). However, results from the application of these methods are not always clear cut nor in agreement with each other. For example, at least three distinctly different topological models have arisen from their application to the abundant and long studied myelin proteolipid protein (14-16). The glycosylation of membrane and secretory glycoproteins occurs only on the luminal side of the membrane of the endoplasmic reticulum and Golgi apparatus (17). We have now exploited the asymmetry of glycosylation to map the topology of the cystic fibrosis transmembrane conductance regulator (CFTR), a member of a large superfamily of membrane proteins involved in transport processes (18) and which when mutated leads to the disease cystic fibrosis (19). The feasibility of this strategy was indicated in its previous successful application to the testing of topological models of an integral membrane protein of the endoplasmic reticulum, hydroxymethylglutaryl-CoA reductase (20).

EXPERIMENTAL PROCEDURES

In order to insert N-glycosylation sequences in predicted extracellular loops and examine their accessibility to the glycosylation machinery, the first step was to eliminate the endogenous consensus glycosylation sequences in extracellular loop 4 (N894 and N900). A fragment in wild-type pNUT-CFTR (21, 22) was replaced by a PCR fragment (between HpaI site at nucleotide 2463 and DraIII site at nucleotide 3328) coding for the N894D and N900D changes to produce a CFTR with no consensus glycosylation sites on extracytoplasmic loops (designated ELO, Fig. 1B). A PCR fragment coding for the E115S change was introduced into ELO to generate a consensus glycosylation site on extracellular loop 1 (EL1, Fig. 1B). Four different PCR fragments containing insertions coding for peptides NTS, GS2NHS, HRNQS, and NTS were inserted into ELO at extracytoplasmic loops 2, 3, 5, and 6, respectively, to generate potential glycosylation sites on each putative extracellular loop (Fig. 1B). PCR fragments coding for the changes M156N, L953N, and V1065N were introduced into EL0 to generate potential glycosylation sites on cytoplasmic loop 1 (CL1, Fig. 1B), loop 3 (CL3, Fig. 1B), and loop 4 (CL4, Fig. 1A), respectively. The sequences of fragments generated by PCR were verified after insertion into the pNUT expression vector.

Chinese hamster ovary-K1 cells were cultured at 37 °C in 5% CO2. Subconfluent cells were transfected with either pNUT-CFTR or mutated versions in the presence of 20 mM HEPES, pH 7.05, 137 mM NaCl, 5 mM KCl, 0.7 mM Na2HPO4, 6 mM dextrose, and 125 mM CaCl2. The cells were shocked 5 h later with 25% glycerol. After a further 24 h, 50 μg methotrexate was added to the medium. Cells continued to grow in the selective medium for about 10 days. Surviving individual colonies were picked and amplified in the selective medium.

CFTR protein detection was performed as follows. Cultures grown from individual colonies were lysed with 1% sodium dodecyl sulfate (SDS). DNA was sheared by cycling through a 30-gauge needle prior to SDS-PAGE (6% polyacrylamide gel). To remove N-linked oligosaccharides by cleavage of the asparagine-N-acetylglucosamine linkage, the lysates were diluted with 10 volumes of buffer (50 mM Tris, pH 8.0, 25 mM EDTA, 1% Triton X-100, 1% SDS). The amounts of lystate protein loaded are indicated in the figure legends. Immunoblotting was done with a primary monoclonal antibody (M3A7) generated against a fusion protein containing residues 1197-1480 of CFTR (23). The secondary antibody was goat anti-mouse labeled with horseradish peroxidase. Chemiluminescence film detection was performed according to the manufacturer’s recommendation (Amersham Corp., ECL Kit).

CFTR chloride channel activity was examined using an iodide efflux assay (21). Confluent cells in 35-mm plates were loaded for 1 h at room

* This work was supported by the Cystic Fibrosis Foundations of Canada and the United States, The NIDDK, National Institutes of Health, and the Respiratory Health Network of Centers of Excellence.

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.

‡ To whom correspondence should be addressed: S. C. Johnson Research Center, Mayo Clinic Scottsdale, 13440 East Shea Blvd., Scottsdale, AZ 85259. Tel.: 602-301-6111; Fax: 602-301-7017.

1 The abbreviations used are: TM, transmembrane sequence; CFTR, cystic fibrosis transmembrane conductance regulator; EL, extracellular loop; CL, cytoplasmic loop; PCR, polymerase chain reaction.
Mapping of CFTR Membrane Topology

The cystic fibrosis transmembrane conductance regulator is predicted to contain six transmembrane segments (TM), six extracytoplasmic loops (EL), and three cytoplasmic loops (CL). Amino acids with putative N-glycosylation sites are indicated in lowercase letters, and conserved asparagine residues are indicated in uppercase letters. The numbers on the right indicate the original amino acid positions in the sequence. The bold italicized capital letters indicate intact predicted transmembrane segments.

**Results and Discussion**

The cytosolic fibrosis transmembrane conductance regulator is predicted to contain six TM segments separated by two cytoplasmic loops (CLs) and three extracytoplasmic loops (ELs) preceding the cytoplasmic nucleotide binding folds in each half of the molecule (schematic in Fig. 1A; Ref. 18). There is evidence that some of the TMs, most notably TM6, are directly involved in formation of the Cl-‐conducting pore of this anion channel (24, 25), and there are disease-causing mutations in some of those putative helices (26). In order to generate more elaborate models of how the helices assemble to form the pore structure, it is first necessary to experimentally evaluate the proposed two-dimensional arrangement of the protein. To do this we have inserted consensus N-glycosylation sequences (NXS/T) in each of the predicted extracytoplasmic and cytoplasmic loops and examined their ability to be glycosylated. As shown schematically in Fig. 1A, 11 such sites were already present in the native sequence. Only the two at positions N894 and N900 in predicted EL4 are used, consistent with the prediction of the original model. When these two sites are removed by in vitro mutagenesis, unglycosylated CFTR is still transported to the cell surface and functions as a chloride channel (27).

We employed this unglycosylated molecule as the starting point for the construction of individual variants with a single consensus N-glycosylation sequence in each loop (Fig. 1B). The extent of sequence perturbation employed to insert these sites ranged from a single amino acid change (E115S) in EL1 to the insertion of 6 additional residues in EL3. The rationale for these insertions was as follows. Because putative EL1 is relatively long (residues 103-117) and required only a single residue change at position 115 to create a consensus N-glycosylation site, this minimal change was made. On the other hand, the rationale for the insertion of 6 additional residues in EL3 was as follows. Because putative EL2 (residues 216-220) contained none of the site-determining residues, a complete set (NST) was inserted after residue 218. Because putative EL3 consisted of a single lysine residue at position 329, 6 residues were inserted so that the asparagine to be glycosylated would be flanked on either side by hydrophilic residues and hopefully accessible to oligosaccharidyl transferase. A similar rational was applied to putative EL5 which originally was comprised by only 2 residues.

A complete acceptor tripeptide was inserted into putative EL6 for the same reason as in EL2. Introduction of consensus sites into each of the putative cytoplasmic loops required only single residue changes as with EL1.

Fig. 2A shows that mature CFTR was formed and underwent complex glycosylation when a glycosylation site was present in any of the predicted extracellular loops. The amounts of the alternatively glycosylated CFTRs were consistently somewhat reduced compared with the wild type. In all cases, digestion with N-glycosidase F removed the oligosaccharide, converting the normal behavior of CFTR glycosylated at any of the ELs. On the other hand, the emu curve was consistently somewhat broadened and the duration of the response prolonged with unglycosylated CFTR (EL0 in Fig. 3), suggesting that the presence of carbohydrate attached at some extracytoplasmic site does influence the activity of the channel. Patch-clamp studies will be required to further characterize this influence as well as detect possible subtle changes in CFTR molecules glycosylated at alternative sites. However, the present data provide compelling evidence that each of the six predicted ELs is available for glycosylation and hence must be extracytoplasmically oriented.
Mapping of CFTR Membrane Topology

**Fig. 2.** Gel electrophoresis of wild-type, unglycosylated, and alternate site glycosylation variants of CFTR. A, immunoblots after SDS-PAGE of cell lysates expressing different CFTR glycosylation variants. Cells were lysed with 1% SDS. The amounts of total lysate protein loaded were as follows: EL4 (wild-type), 2 µg; EL variants, 10 µg; CL1, 10 µg, CL3, 20 µg; and CL4, 5 µg. B, effects of N-glycosidase F on alternately glycosylated CFTR. – or + sign indicates samples without or with N-glycosidase F treatment, respectively, as described under "Experimental Procedures." The amounts of protein loaded were: EL4 (wild-type), 2 µg; EL0, 25 µg; other EL variants, 5 µg. In both A and B, the samples were electrophoresed on 6% polyacrylamide gels which were electroblotted to nitrocellulose and probed with an anti-CFTR monoclonal antibody M3A7 (23). The secondary antibody was a goat anti-mouse labeled with horseradish peroxidase. The blots were processed by ECL (Amersham).

Insertion of the consensus glycosylation sites into the predicted cytoplasmic loops (CLs) had somewhat different consequences. The amounts of these proteins formed were considerably less than the wild-type, especially in the case of CL3. Most importantly for this study, however, only unglycosylated CFTR of the same apparent molecular size as the N894DN900D mutant (EL0) was detectable, indicating that the sites introduced into CL1, CL3, and CL4, like the site already present at N306 in CL2 of the wild-type molecule, were not available for glycosylation and were hence probably cytoplasmically oriented. The amount of forskolin-stimulated iodide efflux, like the amounts of protein, was substantially reduced compared with the wild type and the EL variants or even EL0 (Fig. 3; note different ordinate scales). All of the other nine potential glycosylation sequences (NXS/T) in the native molecule (Fig. 1A) are in protein domains believed for other reasons to be cytoplasmic; the fact that none of them is used is consistent with the idea that accessibility to the glycosylation machinery in the lumen of the endoplasmic reticulum and Golgi is the main determinant of glycosylation at these consensus sites.

In addition to establishing the sidedness of their intertransmembrane loops, their mutagenesis provided some insight into the effects of their perturbation on the biosynthetic processing and function of CFTR. Despite the fact the ELs are quite short in the wild-type molecule, indicating minimal cell surface exposure, it was possible to insert up to 6-mer peptides without seriously compromising protein maturation and channel function. Hence it should be possible to insert epitope tags at these sites to generate CFTR which is easily detectable on intact cells with antibodies of high specificity and avidity. Even though all of the cystic fibrosis-causing natural mutations known to interfere with processing of the protein are located in putative cytoplasmic domains (28, 29), mutagenesis of the CLs here did not appear to have a great effect on the amounts of protein formed compared with their EL0 parent. However, only single amino acid substitutions were made. Hence, although it can be...
inferred that these changes influenced channel function, because stimulated anion permeability was substantially reduced compared with EL0, they seemed to have little further effect on synthesis or processing.

Experiments in which the insertion of the related multidrug resistance P-glycoprotein into microsomal membranes on in vitro translation suggested that some of the 12 predicted helices might not span the membrane (30, 31). However, different deviations from the predicted 12-TM model was suggested in each of these two studies. Skatch et al. (31) claimed that TM8 and TM9 do not cross the membrane, whereas Zhang et al. (30) proposed that there are two populations of P-glycoprotein molecules, the larger of which has all 12 TMs spanning the membrane, the smaller having TMs 3, 5, 8, and 10 excluded from the membrane. In our experiments in which CFTR was synthesized in a functional state in living cells, there was evidence of only a 12-TM form. It may be of interest to re-examine P-glycoprotein using the strategy that we have employed. This approach should be widely applicable in the topological mapping of eukaryotic integral membrane proteins, regardless of additional factors which make sites more or less optimal for glycosylation in addition to just the minimal consensus sequence and accessibility to glycosyltransferases. A recent comprehensive scan of the known positions of N-linked oligosaccharide chains in more than 250 intrinsic membrane glycoproteins has suggested that in nature only single ELs of a minimal size are utilized and with greater frequency toward the N terminus of the molecules (32). The present study, however, shows that experimentally it is possible to insert sites in extremely short putative ELs regardless of their position along the polypeptide. With respect to the possibility of glycosylation at more than one EL, preliminary experiments, in which consensus sequences were introduced into EL2 of a CFTR still possessing the N894 site in EL4, indicated that glycosylation occurred at only one of the sites, although we have not yet identified which one is used. Further analysis of this and other multiple site constructs should provide definitive resolution of this issue.

These findings provide the first rigorous analysis of the normal two-dimensional topology of CFTR in living cells and present a general strategy for studies of other cell surface proteins. Confirmation that all 12 hydrophobic helical segments in CFTR probably do traverse the membrane means that they all will have to be accounted for in the deduction of the overall structure and probably in the elucidation of the mechanism of channel regulation.

As a method for mapping topology, this strategy represents a substantial improvement over more commonly used biochemical methods. The major advantages are 1) that the reporter consensus glycosylation sequences used are minimally perturbing of the native sequence, because they are naturally occurring and widely distributed in many proteins whether they are used or not; 2) the readout is done in living cells in which the protein is in a functional state. This contrasts with the assay of the insertion of in vitro translated polypeptide fragments into isolated microsomes where the functional integrity of the protein cannot be monitored and with the various biochemical methods of detecting exposure to reagents nominally either restricted to or excluded from the bilayer (10–13). An original application of glycosylation site insertion aided in assessing the membrane topology of hydroxymethylglutaryl-CoA reductase (20). However, in that case the detection of the utilization of sites inserted at specific places in the sequence was not performed under conditions where the protein was known to be functional and in its native location. In our case all the glycosylation site variants met these important criteria.

Acknowledgment—We thank Dr. John W. Hanrahan for comments on the manuscript.

REFERENCES