CONFIDENTIAL

A Structural and Functional Characterization of ComEC

Hannah Foster

Candidacy Exam:
May 3, 2013
3pm
NW 195.05

Briana Burton, advisor

Hannah Foster, student

CONFIDENTIAL

Introduction:
The ability of a cell to take in resources and information is critical for life. However, the lipid bilayer making up cell membranes is impermeable, necessitating the use of channels made of integral membrane proteins (IMPs). These protein channels must have the ability to translocate a striking diversity of substrates, including ions, amino acids, metals, and large polymers, such as polypeptides. This has led to a comparably vast array of channel proteins responsible for the translocation of specific substrates. Of particular interest in the bacterial kingdom is the DNA uptake channel, which allows for the uptake of environmental DNA when bacterial cells are in the genetically-induced state referred to as competence.

Integral membrane proteins comprise 30-40% of the human genome (1). However, only about 400 of the nearly 90,000 protein structures published in the Protein Data Bank belong to IMPs (2, 3). The paucity of information regarding membrane proteins results, in particular, from the amphipathic nature of membrane proteins. IMPs are also frequently maintained at a very low concentration in the cell. However, while the 18 years after the publication of the first membrane protein structure (4) produced only around 75 IMP structures, improved techniques and new assays over the last 10-15 years are facilitating the study of IMPs. This has led to a greater understanding of a variety of channel proteins (2).

DNA uptake channels, however, are an essentially uncharacterized class of channels. Although natural transformation in bacteria was discovered nearly 90 years ago (5), characterizing these channels has been slow because of inadequate tools, the complexity of the competence system, and limited investigation by researchers. The competence system is complex, including a handful of proteins necessary for DNA transport through the peptidoglycan layer, several proteins, including the channel protein, involved in the translocation of DNA across the cell membrane, and proteins required for the processing of DNA following uptake (6; fig. 1a). ComEC is the highly-conserved DNA uptake channel in *Bacillus subtilis*, which has one of the best-characterized competence systems (7). ComEC is the best candidate for the channel protein based on the fact that it is a large, polytopic IMP (8; 9). Additionally, ComEC is absolutely required for DNA uptake (10). Scientists have little understanding of the structure and mechanism of ComEC, but the channel is probably quite distinctive, since the translocation of a large, hydrophilic molecule like DNA across the cell membrane is a very unique problem. Characterizing ComEC will provide new insight into the ways by which proteins address the problem of gating and specificity.

The goal of my project is to characterize the DNA uptake channel ComEC. Using novel and standard genetic and biochemical assays, I will 1) develop a structural model of ComEC using topology mapping and analysis of the path of DNA through the channel, 2) determine the oligomeric state of translocating ComEC, and 3) characterize the biophysical properties of ComEC.

SPECIFIC AIMS:

1) *Develop a structural model of ComEC using topology mapping and analysis of the path of DNA through the channel*

Topology mapping provides an accessible method for understanding the structure of proteins and simultaneously illuminates aspects of protein function. I will use genetic and biochemical methods to map the topology of ComEC and determine the path taken by DNA through the protein.

Researchers often use homologous proteins to aid in solving a protein's structure. Unfortunately, ComEC has no homology to any characterized structure. Previous research using ComEC-LacZ/PhoA fusion proteins has led to a proposed topology map (9; fig. 1b). The reporter system, however, yielded conflicting results in some regions of ComEC. Moreover, the published map neither agrees with a limited series of fluorescent protein fusions (H. Wang, unpublished data), nor the predicted topology from hydropathy analysis programs.

Site-directed Fluorescent Labeling:

I will use site-directed fluorescent labeling to identify external and internal residues of the channel. This approach has two primary advantages: it only requires the mutation of a single amino acid to cysteine, and labeling of the protein is performed after protein folding. Thus, the structure of ComEC is likely to remain unperturbed. I have identified 54 serines of 56 in ComEC, in addition to two threonines in areas under-represented by serines, where I will use site-directed-mutagenesis (SDM) to replace the serine with a cysteine. After confirming the functionality of the cysteine mutants using a transformation efficiency assay, I will proceed with the *in vivo* crosslinking of probes to the wildtype or mutant strains (Fig. 2). I will split a population of competent cells into two samples. One sample will be pretreated with the non-fluorescent thiol-reactive, cell-impermeable reagent 4-acetamido-4'-maleimidylstilbene-2-2'-disulfonic acid (AMS) to prevent further crosslinking of any externally-located cysteines. Subsequently, both samples will be incubated with Oregon Green 488 maleimide (OGM), a fluorescent, cell-impermeable thiol-reactive probe. I will then use a protein gel to detect fluorescence and a Western blot to determine the quantity of total protein. I will then compare the normalized fluorescence intensity of the two samples. If the non-pretreated sample exhibits greater fluorescence intensity than the ASM-pretreated sample, the cysteine is likely located extracellularly. However, if both samples show equal fluorescence, the location of intracellular cysteines or cysteines in transmembrane segments (TMSs) can be confirmed by permeabilizing the cells prior to treatment with the probes.

ComEC contains eight natural cysteines, which could react with the probes, obscuring the results. However, in past studies using this assay, researchers

demonstrated that it is not always necessary to remove native cysteines (11). I will test wildtype ComEC for background fluorescence. If it appears that one or more native cysteines may interfere with the assay, I will use SDM to replace the confounding cysteine(s) with a serine. Draskovic and Dubnau (2005) showed that six of the cysteines can be removed at least one, two, or three at a time without affecting the function of ComEC. The remaining two cysteines are required for ComEC function but are unlikely to interfere with labeling because they are involved in a disulphide bond (9, 11).

Also of note is the fact that wildtype *B. subtilis* cells only express competence genes in 5-10% of the population when naturally induced (12). To increase the competent proportion of the population, I will use a *comS*-inducible background for my experiments. ComS frees the competence master regulator, ComK, to activate transcription of itself and other competence genes (13; 14). Overexpression of ComS results in a higher proportion of free ComK, leading to a higher proportion of competent cells, while levels of competence proteins within individual cells remain within a physiological range.

Determining the location of residues in ComEC will allow me to combine our data with the topology map proposed previously and hydropathy analysis program predictions to develop a convincing topology map for ComEC. This will provide a better understanding of the distinctive protein structure that allows for DNA uptake.

DNA Crosslinking to Unnatural Amino Acids:

Determining the path of DNA through ComEC can help define protein structure and give insight into the mechanism of the protein. To do this, I will use unnatural amino acids to induce site-directed protein-DNA photo-crosslinking (SDPC) *in vivo*. The incorporation of unnatural amino acid (UAA) *p*-benzoyl-L-phenylalanine (pBpA) into proteins has been used to study protein-DNA interactions (15). SDPC will allow me to determine specific residues or regions of ComEC that interact with DNA. The technique involves utilizing a transfer RNA (tRNA) and an aminoacyl synthetase that recognize pBpA. Once the tRNA is charged with pBpA supplied through the cell culture media, the tRNA delivers the pBpA to a low-abundance nonsense codon, such as the amber (TAG) stop codon. This results in the suppression of the stop codon and the expression of the pBpA in the protein of interest. I will incorporate pBpA into ComEC at sites that may be interacting with DNA. The region inside the gray box in figure 1b highlights the most conserved region of the protein, so it follows that this region likely contains residues that interact with DNA. In addition, a high proportion of hydrophilic residues in a TMS may indicate the location of the aqueous pore, such as TMSs E, F, and G in figure 3. Following incorporation and confirmation of the functional mutants in a *comS*-inducible background, I will induce *comS*. I will then add DNA to the cells and expose the cells to ultraviolet light, causing crosslinking of pBpA to any DNA within a few angstroms of the UAA. For analysis, I will use an Electrophoretic Mobility Shift Assay (EMSA). DNA is cleaved

before transformation occurs (6), so the DNA fragments bound to ComEC may not all be equal in size. Thus, the ComEC-DNA crosslinked sample likely will not produce a distinctive band on the gel. However, the absence of a band can be used as an initial indicator that DNA is bound to ComEC when compared with wildtype ComEC and a non-crosslinked UAA mutant. For UAA mutants that appear to have crosslinked to DNA, the presence of DNA can later be confirmed by radiolabeling the ComEC-associated DNA

Our lab is currently creating a working UAA system for *B. subtilis*. Although it is possible that the UAA system will not work in *B. subtilis*, this is unlikely, since the UAA system has been utilized in many different cell types, including *Escherichia coli*, *Mycobacterium tuberculosis*, yeast, and mammalian cells (16; 17; 18; 19).

The information from the SDPC and DNA crosslinking assays combined will confirm TMSs, internal and external regions of the protein and identify residues or regions of the protein making up the aqueous channel.

2) *Determine the oligomeric state of translocating ComEC*

Oligomerization is necessary for the function of many channel proteins. Thus, an important aspect of the characterization of ComEC involves determining the oligomeric state of translocating ComEC.

Data from a crosslinking assay suggests that ComEC may form homodimers (9). However, the validity of the purported homodimer is questionable, since in this case, dimerization occurred *in vitro* where conditions may have led to ComEC interacting unnaturally. Moreover, the possibility of higher-order oligomerization has not been explored. Research from our lab suggests that ComEC must interact with another protein, such that transformation efficiency is poisoned with increasing concentration of non-functioning ComEC (hereafter, ComEC*) expressed with wildtype ComEC (H. Wang, unpublished data). It is unclear, however, whether this interacting protein is ComEC or another competence protein. Other competence proteins that may interact with ComEC, based on fluorescence assays and pull-downs, include ComEA, ComGA and probably ComFA (see fig. 1a; 20).

<u>Genetic Study Using ComEC*:</u>

I will utilize ComEC* to determine whether ComEC forms a hetero-oligomer or a homo-oligomer. To accomplish this, I will use a strain with ComEC* expressed under an IPTG-inducible promoter in a wildtype background to create several new strains, each containing one ectopically-expressed component of the competence system, including wildtype ComEC, ComEA, ComGA, and ComFA. I will then compare transformation efficiency among the strains. If increasing expression of one of these proteins rescues the dominant-negative phenotype of ComEC*, ComEC likely

interacts with the overexpressed protein for transformation to occur. The results of this assay will give us critical information about essential protein-protein interactions in the competence system.

It is possible that no protein alone will fully rescue the dominant negative phenotype. However, if any of the proteins partially rescue the phenotype, this suggests that protein may interact with ComEC, in addition to one or more other proteins. In this case, by simultaneously inducing the overexpression of multiple proteins, I will determine the combination of proteins able to fully recover transformation efficiency.

Fluorescence Co-localization with Fluorescently-labeled Proteins and DNA:

After I have determined the interacting protein(s), I will examine the functional significance of this interaction using fluorescence-labeling to determine whether the interaction of the proteins coincides with DNA uptake. I will label two of the interacting proteins with a CLIP or a SNAP tag, which will allow for the specific, simultaneous labeling of ComEC and the interacting protein in cells with different substrate probes (21). I will also label the DNA using the gfp-tagged tetracycline repressor (TetR), which forms a tight and specific bond with the tetracycline operator (*tetO*) sequence (22) encoded in the substrate DNA. This type of transporter-substrate characterization usually relies on the creation of a stalled translocation intermediate. The bulky fluorophores attached to the DNA may cause the DNA uptake machinery to stall. If so, and the interacting protein remain as oligomers during translocation, I will see an enrichment of the three probes co-localized when compared to a control using the non-translocating ComEC*. If only two fluorescent probes co-localize at one time but DNA can be visualized co-localizing with the labeled ComEC, this suggests that the oligomer separates for translocation. If, however, DNA never appears to co-localize with either protein or all three probes are not observed, this indicates a problem with the system: the TetR-GFP probe may be cleaved off or otherwise removed, may be allowed through the channel, or may prevent uptake altogether.

If the *tetO*/TetR system fails because the fluorophores do not stall DNA uptake, I will use fluorescent nucleotides to label the DNA and take advantage of a pBpA mutant from specific aim 1 to create a stalled translocation intermediate for visualization of the oligomer and DNA.

The assays described above will discern whether ComEC forms a homo-oligomer or a hetero-oligomer and will shed light on the necessity of this interaction for transformation.

3) *Characterize the biophysical properties of ComEC*

A biophysical characterization of channel proteins can provide important insight into the selectivity of the channel and the response of the channel to various ligands or stimuli. Currently, there is no available information on the biophysical properties of ComEC. Thus, data from this study will provide entirely novel information regarding the substrate specificity and gating of ComEC, in addition to answering the critical question of whether or not isolated ComEC is sufficient for DNA passage across the cell membrane.

Reconstitution of ComEC in Liposomes:

Particularly when characterizing the biophysical nature of an IMP, it is often helpful to reconstitute the protein in liposomes. Proteoliposomes provide an experimental setup lacking many of the confounding factors associated with research in cells, such as other cellular processes and interacting proteins. I will therefore reconstitute ComEC in liposomes in order to obtain data from the channel in a more controlled environment.

To facilitate purification, I will use a functional ComEC C-terminally-tagged with myc3 (ComEC-myc3) for reconstitution. Detergent solubilization of ComEC-myc3 is the first step towards its reconstitution in proteoliposomes. It is impossible to predict exactly which detergent will effectively solubilize a particular protein, so solubilization must be attempted with a variety of detergents (23). Therefore, I will test for solubilization of ComEC-myc3 using a wide range of detergents, focusing mainly on milder nonionic and zwitterionic detergents. These detergents have proven more effective than ionic detergents at maintaining protein activity after solubilization because they do not irreversibly denature proteins (23). After determining an effective detergent, I will proceed with ComEC-myc3 purification via affinity chromatography, followed by a polishing step. I will mix the purified proteins with preformed liposomes and detergent, resulting in a mix of protein-lipid-detergent and lipid-detergent micelles. The detergent must then be removed for proper proteoliposome formation. This can be accomplished in a number of ways, but polysterene beads, such as Bio-Beads, are my preferred method because they are very effective at almost entirely removing detergents of all types in a controlled manner (24). The resulting proteoliposomes can then be used for further experimentation described below.

If the reconstitution of liposomes in ComEC is not possible, our lab has confirmed that ComEC can be expressed via viral transfection in SF9 insect cells. Thus, I will use these cells to complete the following experiments.

Electrophysiology using Port-a-Patch:

Whole-cell patch-clamping is often used to study biophysical characteristics of channel proteins, including ion channels, gap junctions, and even an RNA channel (25; 26; 27). However, traditional patch-clamping can be cumbersome and time-consuming. I will examine the biophysical characteristics of ComEC using a more efficient patch-clamping technique: the Port-a-Patch® system by Nanion,. Some very preliminary electrophysiological data from our lab using the Port-a-Patch® suggests that ComEC may open when exposed to DNA (unpublished results, fig. 4). I will continue these experiments by measuring fluctuations in membrane potential to determine under what conditions ComEC is open, if it responds differently to various DNA substrates, or if other proteins are required for channel gating.

The state of the ComEC channel may not be affected directly by the presence of DNA but rather, may require other proteins to open or close the channel. In this case, the above experiments may be un-enlightening. A more effective strategy in this situation could be the patch-clamping of spheroplasts, which has been accomplished by Nanion (28). This will allow comparison of electrophysiological responses to DNA of wildtype versus mutant ComEC in the presence or absence of certain protein components of the competence system.

Membrane permeability assay:

In addition to discerning the conditions under which ComEC is open, other critical information includes what types of molecules are allowed passage through the channel and whether or not the isolated channel is bidirectional. Since the 1970s, dyes have been utilized to measure membrane potential changes in ion channels (29), as well as gap-junction permeability (30). To examine the permeability of ComEC to various substrates, I will use ComEC proteoliposomes to test a variety of dyes with a range of sizes and charges, in addition to fluorescently-labeled nucleotides incorporated into oligonucleotides of different lengths.

These permeabilization experiments will provide valuable information about the size and charge of molecules allowed into the ComEC, the conditions that lead to channel opening, and whether or not other proteins may be involved in controlling the opening and closing of the channel.

Conclusion:

Together, Specific Aims 1 and 2 will provide the first complete model for the tertiary and quaternary structure of ComEC in the context of its DNA substrate. Specific Aim 3 will further elucidate the mechanistic and biophysical characteristics of ComEC by giving us information on the sufficiency of isolated ComEC for uptake, substrate specificity and the conditions under which ComEC is open and/or able to transport DNA. Information regarding substrate specificity may provide insight into the function of ComEC orthologs in non-competent species of bacteria, since the channel may have been adapted over time for other purposes. Moreover, this research may reveal possible uses for ComEC as a general or specific channel for research or biomedical purposes. If ComEC proves sufficient for the passage of DNA molecules into cells, this research could even lead to major breakthroughs in the transformation of cell types lacking efficient transformation. The delineation of the combined structural, mechanistic, and biophysical characteristics of ComEC will aid tremendously in creating a more comprehensive picture of this unique class of channels and point to exciting new research questions in the study of transformation.

Figures:

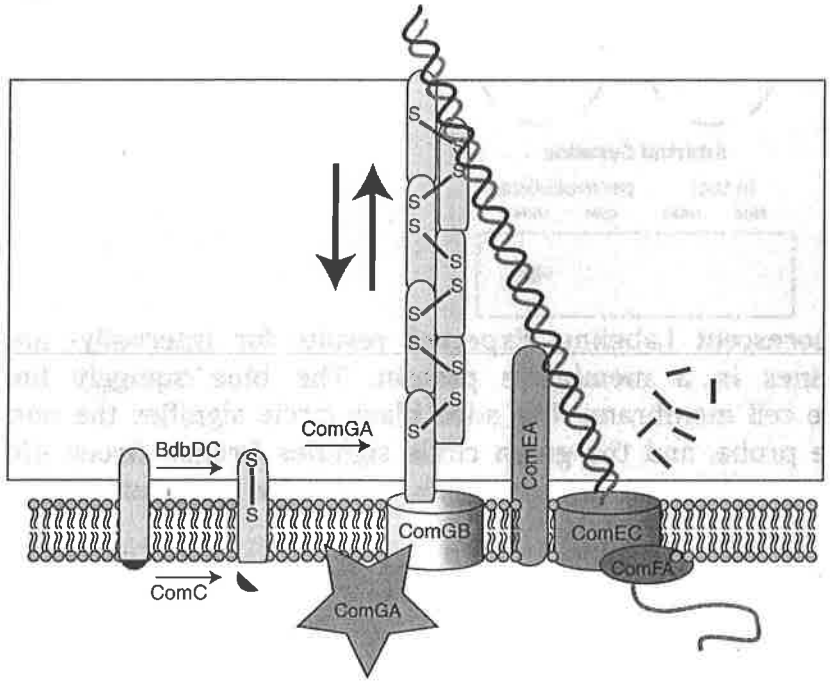

Fig. 1a. Components of the competence system in gram-positive bacteria. Proteins required for processing after DNA uptake are not illustrated here. (Dubnau, 2010).

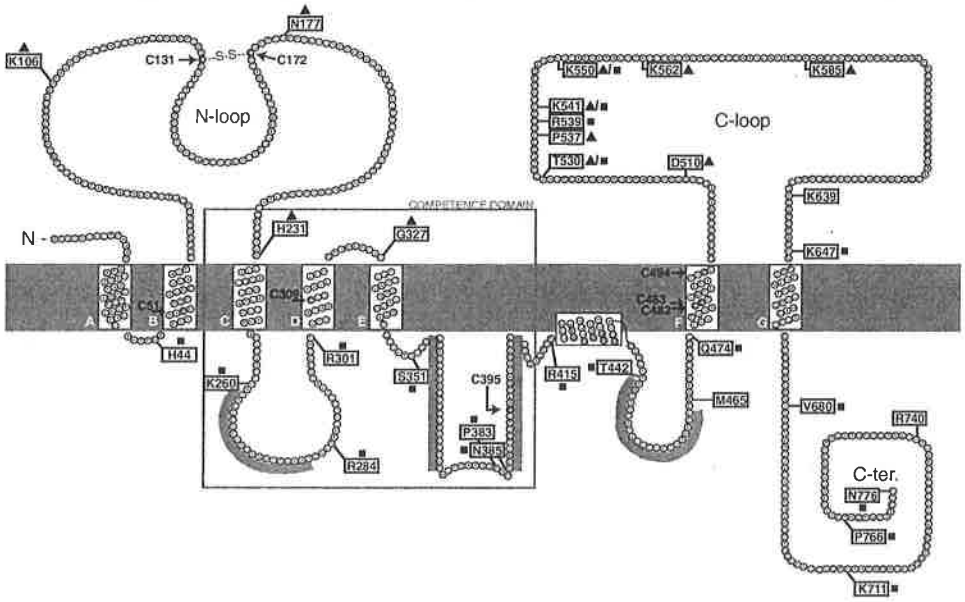

Fig. 1b. The topology as predicted by a random mutagenesis assay resulting in ComEC fusion proteins to PhoA or LacZ. The small black boxes signify fusion proteins that revealed high LacZ activity, while the black triangles signify high PhoA activity. The gray box labeled "competence domain" indicates the most highly-conserved region of the protein. The thick gray lines denote hydrophobic regions not predicted to be TMSs (Draskovic and Dubnau, 2005).

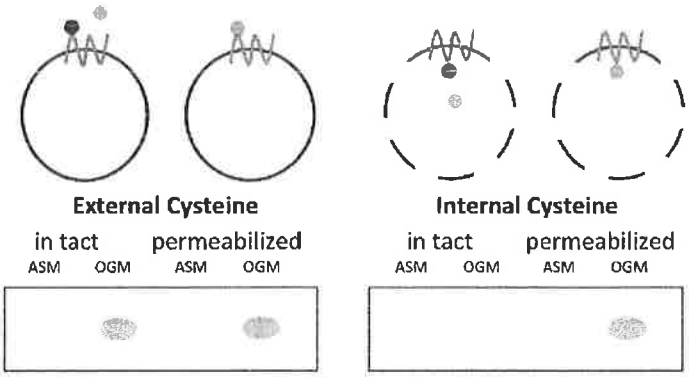

Fig. 2. Site-directed Fluorescent Labeling. Expected results for internally- and externally-located cysteines in a membrane protein. The blue squiggly line represents ComEC in the cell membrane. The solid black circle signifies the non-fluorescent crosslinkable probe, and the green circle signifies Oregon Green 488 Maleimide. Below are representative possible gels, with the gray solid ovals indicating fluorescence.

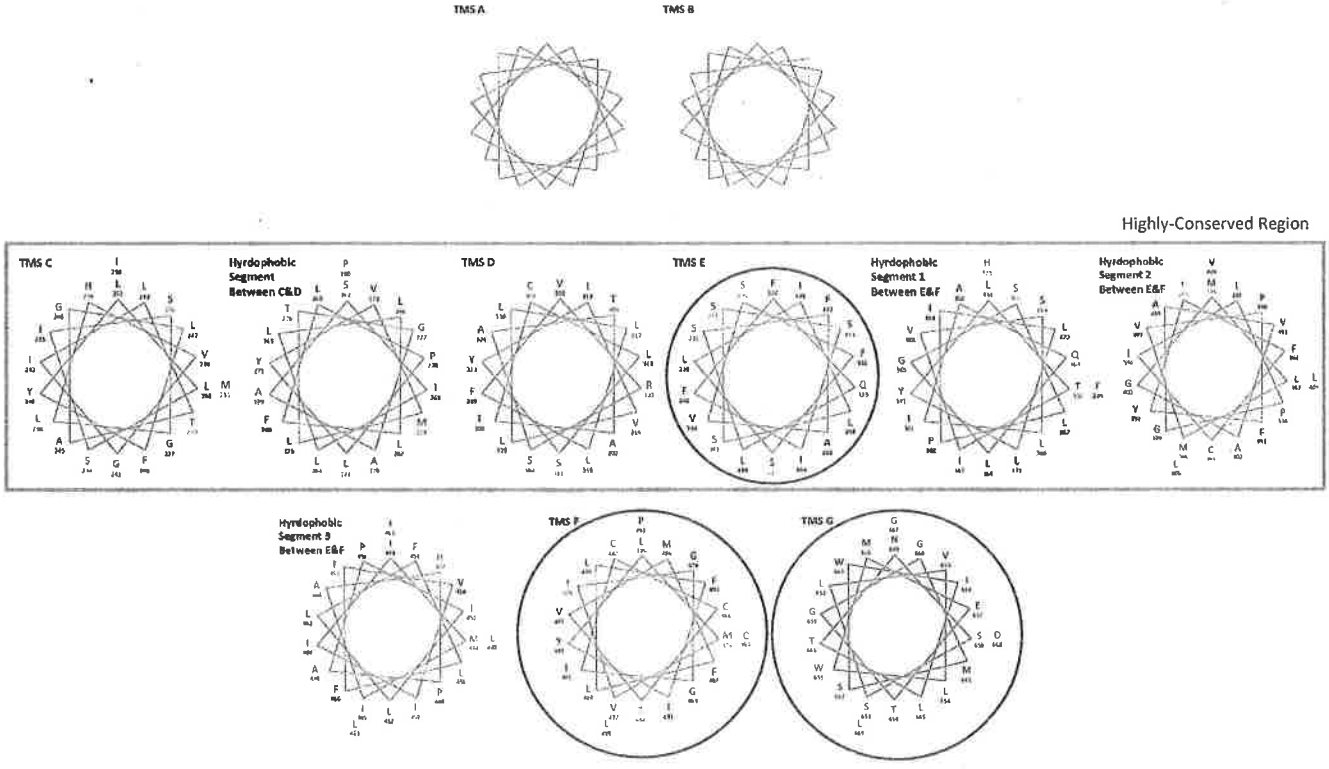

Fig. 3. Helical wheels of predicted TMSs. Hydrophilic residues are highlighted in red, and the red circles denote helical wheels with >33% hydrophilic residues. The large gray box indicates the most highly-conserved region of ComEC.

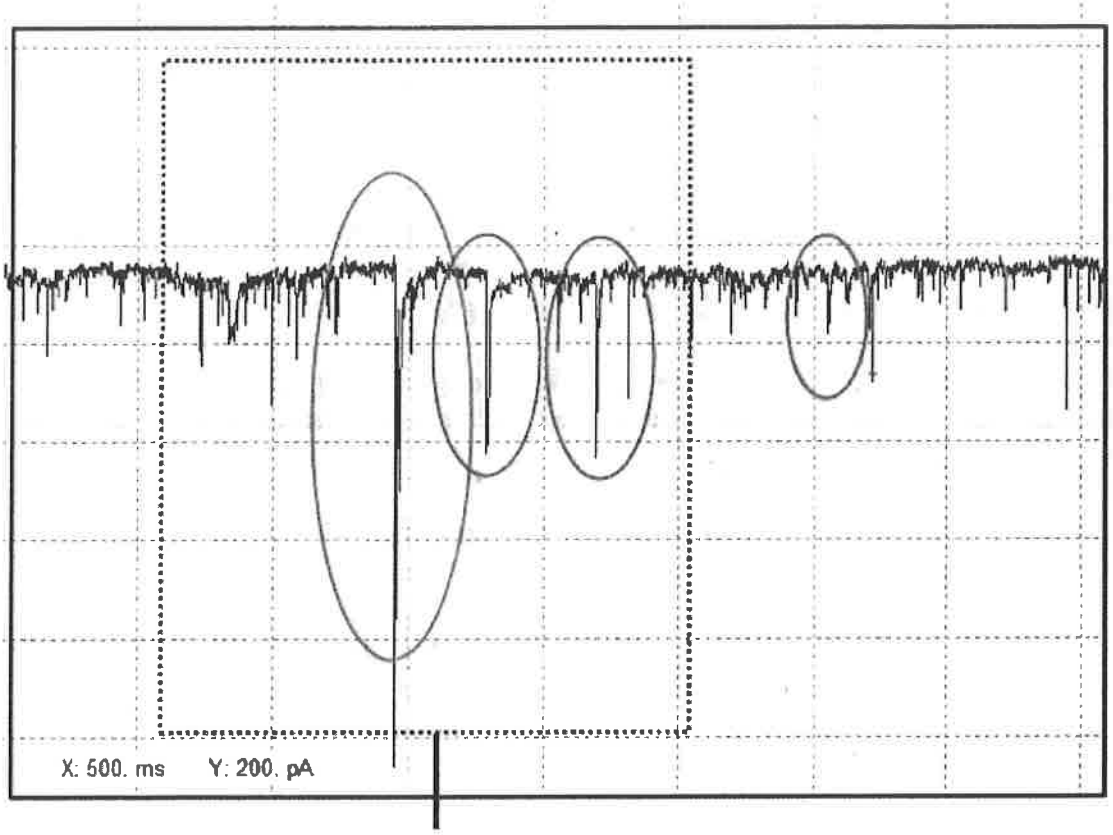

Fig. 4. Changes in conductance states indicate channel(s) opening. ssDNA was added to ComEC-transfected SF9 cells. Circles of different colors denote translocation events of different magnitudes. When ssDNA was added to untransfected cells, similar events were not observed.

Works Cited:

1) Watts, A. (2005). Solid-State NMR in drug design and discovery for membrane-embedded targets. Nat Rev Drug Discovery *4*, 555-568.

2) White, S.H. (2013). Membrane proteins of known 3D structure (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html).

3) RCSB Protein Data Bank. http://www.rcsb.org/pdb/home/home.do

4) Deisenhofer, J., S.J. Remington, and W. Steigemann. (1985). Experience with various techniques for the refinement of protein structures. Methods in Enzymology. 115:303-323.

5) Griffith, F. (1928). The Significance of Pneumococcal Types. *Journal of Hygiene.* 27:113-159.

6) Dubnau, D and B. Burton. (2010). Membrane-associated DNA Transport Machines. Cold Spring Harbor Laboratory Press.

7) Chen, I, P.J. Christie, and D. Dubnau. (2005). The Ins and Outs of DNA Transfer in Bacteria. *Science.* 310:1456-1460.

8) Hahn, J., G. Inamine, Y. Kozlov, and D. Dubnau. (1993). Characterization of *comE*, a late competence operon of *Bacillus subtilis* required for the binding and uptake of transforming DNA. *Molecular Microbiology.* 10:99-111.

9) Draskovic, I. and D. Dubnau. (2005). Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Molecular Microbiology.* 55:881-896.

10) Hahn, J., M. Albano, and D. Dubnau. (1987). Isolation and characterization of Tn917lac-generated competence mutants of *Bacillus subtilis. Journal of Bacteriology.* 169:3104-3109.

11) Henderson, N.S., S.S.K. So, C. Martin, R. Kulkarni, and D.G. Thanassi. (2004). Topology of the Outer Membrane Usher PapC Determined by Site-directed Fluorescence Labeling. *The Journal of Biological Chemistry.* 279:53747-53754.

12) Haijema, B.J., J. Hahn, J. Haynes, and D. Dubnau. (2001). A ComGA-dependent Checkpoint Limits Growth during the Escape from Competence. *Molecular Microbiology*. 40(1): 52- 64.

13) Turgay, K., L.W. Hamoen, G. Venema, et al. (1997). Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the

activity of ComK, the competence transcription factor of *Bacillus subtilis. Genes and Development.* 11:119-128.

14) van Sinderen, D., A. Luttinger, L. Kong, D. Dubnau, G. Venema, and L. Hamoen. (1994). *comK* encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis. Molecular Microbiology.* 15:455-462.

15) Lee, H.S., R.D. Dimla, P.G. Schultz. (2009). Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. Bioorganic and Medicinal Chemistry Letters. 19:5222-5224.

16) Chin, J.W., A.B. Martin, D.S. King, L. Wang, and P.G. Schultz. (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America*. 99:11020-11024.

17) Wang, F., S. Robbins, J. Guo, W. Shen, and P.G. Schultz. (2010). Genetic Incorporation of Unnatural Amino Acids into Proteins in *Mycobacterium tuberculosis. PLOS One.* 5:e9354.

18) Chin, J.W., T.A. Cropp, J.C. Anderson, M. Mukherji, Z. Zhang, and P.G. Schultz. (2003). An Expanded Eukaryotic Genetic Code. *Science.* 301:964-967.

19) Zhang, Z.W., L. Alfonta, F. Tian, B. Bursulaya, S. Uryu, D.S. King, and P.G. Shultz. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America.* 101:8882-8887.

20) Kramer, N., J. Hahn, and D. Dubnau. (2007). Multiple interactions among the competence proteins of *Bacillus subtilis. Molecular Microbiology.* 65:454-464.

21) Gautier, A., A. Juillerat, C. Heinis, I.R. Corrêa, Jr., M. Kindermann, F. Beaufils, and K. Johnsson. (2008). An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry and Biology*. 15:128-136.

22)Berens, C. and W. Hillen. (2003). Gene regulation by tetracyclines. *European Journal of Biochemistry.* 270:3109-3121.

23) Kalipatnapu, S. and A. Chattopadhyay. (2005). Membrane protein solubilization: Recent advances and challenges in solubilization of serotonin(1A) receptors. *IUBMB Life.* 57:505-512.

24) Rigaud, J. and D. Lévy. (2003). Reconstitution of membrane proteins in liposomes. Methods in Enzymology. 372:65-86.

25) Hamill, O.P., A. Marty, E. Neher, B. Sakmann, and F.J. Sigworth. (1981). Improved Patch-Clamp Techniques for High-Resolution Current Recordings from Cells and Cell-Free Membrane Patches. *European Journal of Physiology*. 391:85-100.

26) Elfgang, C., R. Eckert, H. Lichtenberg-Fraté, A. Betterweck, O. Traub, R.A. Klein, D.F. Hülser, and K. Willecke. (1995). Specific Permeability and Selective Formation of Gap Junction Channels in Connexin-transfected HeLa Cells. *The Journal of Cell Biology*. 129:805-817.

27) Shih, J.D. and C.P. Hunter. (2011). SID-1 is a dsRNA-selective dsRNA-gated channel. *RNA*. 17:1057-1065.

28) Nanion Technologies. Preparation and recordings on Nanion's Port-a-Patch of native *E. coli* Spheroplasts. (http://www.nanion.de/images/stories/pdf/Port-a-patch_Spheroplasts.pdf) downloaded 2013-04-16.

29)Waggoner, A.S. (1979). Dye indicators of membrane potential. Annual Reviews in Biophysical Bioengineering. 8:47-68.

30) Eckert, R. (2006). Gap-Junctional Single-Channel Permeability for Fluorescent Tracers in Mammalian Cell Cultures. Biophysical Journal. 91:565-579.