THE PUZZLE OF CLONING SHARK (*SQUALUS ACANTHIAS*) CFTR cDNA: MULTIPLE TOXIC SEQUENCE MOTIFS MAY BE KEY TO ISOLATING A FULL LENGTH EXPRESSION CLONE

Yijian Fei[1], Ali Poyan Mehr[1], and John N. Forrest, Jr.[1]
[1]Department of Medicine, Yale University School of Medicine, New Haven, CT 06520

The cystic fibrosis transmembrane conductance regulator (CFTR) is a cyclic AMP-regulated chloride channel protein. Numerous mutations in CFTR have been identified as genetic causes of cystic fibrosis, one of the most common inherited fatal human diseases. Research on CFTR has great significance for understanding cystic fibrosis at the molecular level, and also provides insights into bio-physiological elements of chloride channel function and regulation. Full length CFTR cDNA from only a limited number of species has been cloned into vectors, including human (Gregory R. J. et al., Nature 347:382-386, 1990), *Xenopus* (Price M. P. et al., J. Biol. Chem. 271:25184-25191, 1996) and killifish (Singer T. D. et al., Am. J. Physiol. 43:C715-C723, 1998), but not shark. While several groups have attempted to clone shark CFTR (sCFTR) cDNA over the past decade, a full length expression clone has not yet been isolated due to difficulties associated with the shark sequence. Cloning of shark CFTR would be useful for mutagenesis studies, protein-protein interactions and expression assays to examine structure-function relationships and regulation of this channel protein.

A long-standing puzzle is why shark CFTR cDNA is so difficult to clone. Based on other groups work in the cloning of human and *Xenopus* CFTR, we hypothesized that there are bacterial promoter-like sequences embedding in the shark CFTR cDNA that are toxic to the host bacteria by turning on aberrant expression of shark CFTR. One commonly used approach to identify a promoter is the functional assay of a reporter fused to the candidate promoter sequences in a heterologous expression system (Fei Y. et al., J. Biochem. 125:1189-1199, 1999). However, this approach is very time-consuming and costly. In order to clone the full-length sCFTR cDNA, we undertook alternative strategies to identify these potential toxic sequences.

We initially attempted to clone sCFTR with high and low copy number vectors and host bacterial strains, but these attempts were unsuccessful. As shown in Figure 1, the 4.55kb full-length sCFTR cDNA was amplified from shark rectal gland-derived cDNA using Expand High Fidelity PCR System with a forward primer containing the Kozak sequence, transcription start site and a Not I site and a backward primer containing the stop codon, 3'untranslated sequence and a Sma I site. Then it was either directly cloned into pCR II -TOPO vector with a TA-cloning kit or digested with Not I/Sma I, gel purified and cloned into pBluescript II KS (+) that was used successfully for cloning human CFTR. TOP 10F' and low copy bacterial strains ABLE C and ABLE A were used as host strains for transformation. Plasmid DNA was digested with Not I/Sma I. No clones contain a correct insert. sCFTR cDNA clones constructed in a low copy number plasmid and host bacterial strain still did not have an insert of expected size. sCFTR was PCR-amplified, phosphorylated, gel-purified, and ligated to pSMART vector that has translation terminators. Low copy number host strain E. Coli 10G was used for transformation. Plasmid DNA was digested with Hind III. Fewer colonies were grown.

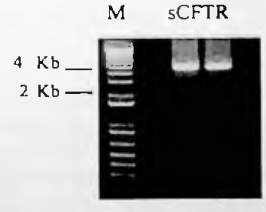

Figure 1. PCR-amplified full-length sCFTR

To identify the toxic bacterial promoter sequences, we compared shark CFTR cDNA sequences with consensus promoter sequences in bacteria (Harley CB, Reynolds RP, Nucleic Acids Res 11;15(5):2343-61, 1987), as well as the identified bacterial promoters in human and *Xenopus* CFTR cDNA (Gregory R. J. et al., Nature 347:382-386, 1990. Price M. P. et al., J. Biol. Chem. 271:25184-25191, 1996). As shown in Figure 2, such sequence comparison revealed one highly conserved (promoter 1) and one less conserved bacterial promoter (promoter 2).

Bacterial consensus promoter sequence:

TTGACANNNNNNNNNNNNNNNNTATAAT

	Conserved bacterial promoters in CFTR	Less conserved bacterial promoter in CFTR
Human	776TTGTGATTACCTCAGAAATGATTGAAAAT804	1159TTAACGACTACAGAAGTAGTGATGGAGAAT1188
Xenopus	779TTGTGATAACATCTCAAATCATCGAAAAC807	1162TTGACAACCACAGAAGTTGCCATGGAAAAT1191
Shark	779 TTGCAATCACCTCTGAAATTATTGACAAT807	1162CTAACAACAAAAGAGGTTGAGATGGTAAAT1191
Killifish	779TGGCTCTTACTTCTGAGATTGTGGAGAAC807	1162CTGACCACAGTTGGACTAGAGCTAATCAAT1191

	Third putative bacterial promoter in sCFTR
Human	1240CAAAACAATAACAATAGAAAA1260
Xenopus	1243CTAGAAGTAAATGGAGGGAAT1263
Shark	1243CAGAATGACTCCGAGCGCAAG1263
Killifish	1243CGGGAAAACAAAGCCAATGGA1263

Figure 2. Sequence comparison of CFTR cDNAs among species identifying conserved sequence motifs that are homologous to consensus bacterial promoter sequences. Numbers indicate the positions of the nucleotides relative to the A of the translation start codon. N represents any of the A, T, C and G.

Four silent mutations were introduced into the highly conserved bacterial promoter 1 sequence in sCFTR using 2-step PCR-based mutagenesis with an inner pair of primers containing the mutations and a flanking pair of primers harboring unique restriction sites (Fei Y. et al., Invest. Ophthalmol. Vis. Sci. 41: 2849-2856, 2000). The cDNA was either directly cloned into PCR II-TOPO vector using TA-cloning or into pBluescript KS (+) at Not I/Sma I sites. More colonies were grown. However, screening the sCFTR cDNA clones bearing the mutant bacterial promoter was unable to reveal any full-length clones. Most clones had inserts with reduced sizes representing deleted sCFTR sequences. Thus disruption of this promoter sequence failed to allow cloning of the full-length sCFTR suggesting that this highly conserved promoter alone is not sufficient to confer the sequence toxicity of sCFTR in host bacteria.

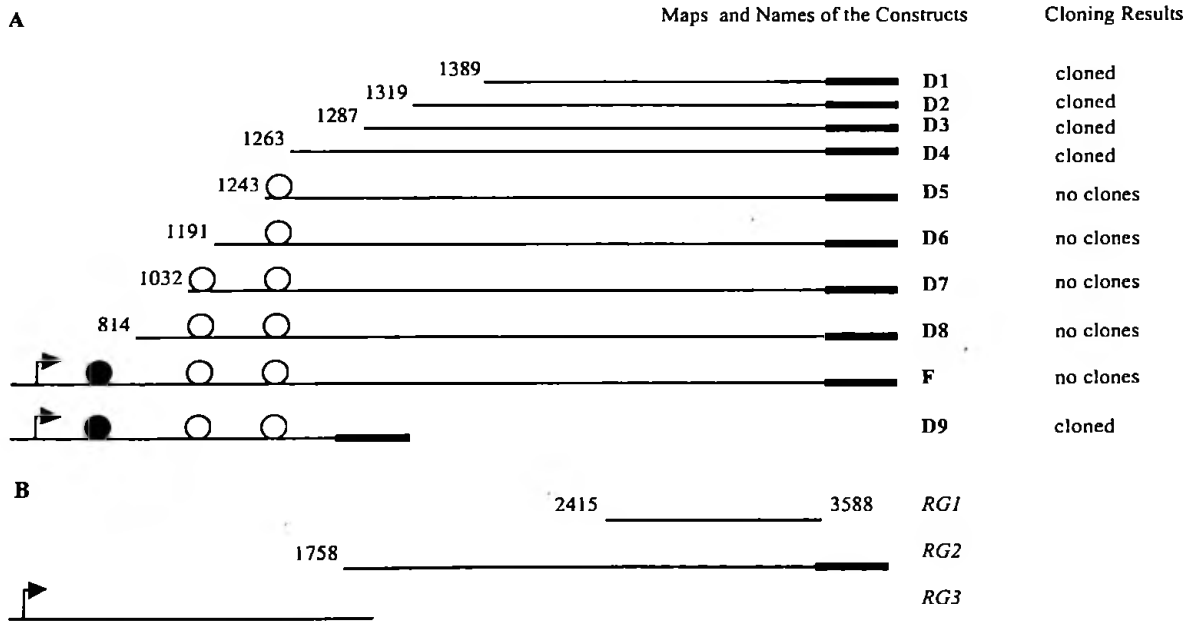

Figure 3. A: Mapping the additional toxic sequences in sCFTR cDNA by systematic cloning of the 5' progressively deleted sCFTR cDNA sequences. Lines: sCFTR. Bars: vector. D1-D9: deletion constructs. F: full-length construct. Numbers: positions of the 5' nucleotides. Arrows: translation start site. Darkest circle: highly conserved bacterial promoter; Darker circle: less conserved bacterial promoter; Open circle: putative promoter identified by deletion mapping. B: Clones identified by screening a shark rectal gland cDNA library (Marshal et al., Journal of Biological Chemistry 25;266(33):22749-54, 1991). Scales in A and B are different and not linear (see nucleotide numbers).

To map the potential bacterial promoter-like toxic sequences, we used another strategy by creating a series of constructs that contain the 5' progressively deleted sCFTR cDNA (Figure 3) and testing whether these partial sequences could be cloned. Rapid PCR-based TA cloning cloning was used for cloning these PCR amplified individual sCFTR cDNA fragments. Representative positive clones that were verified by DNA sequencing are illustrated in Figure 4. Positive clones from sCFTR D1 to D4 could be readily isolated (Figure 4A), and confirmed by sequencing. However, the sCFTR D7 and D8 constructs that deleted the highly conserved first bacterial promoter and the D5 and D6 constructs that deleted both the first and the second bacterial promoters could not be readily cloned. Instead, deleted or rearranged inserts were observed in these clones (Figure 4B). On the other hand, a 1.6 kb 5' sCFTR sequence can be readily cloned (Figure 4C) and confirmed by sequencing. These results suggest that these two putative bacterial promoters alone are not sufficient to confer the sequence toxicity. Interestingly, construct D5 that could not be cloned contains only 20 additional nucleotides of the 5' sCFTR cDNA compared with D4 that could be readily cloned. Thus these 20 nucleotides may contain another toxic promoter sequence that alone or most likely, in cooperation with the other two promoters confer the toxicity of sCFTR cDNA in the host bacteria. Sequence analysis of this 20 bp fragment revealed a sequence motif that is similar to the half site of the conserved bacterial promoter sequences.

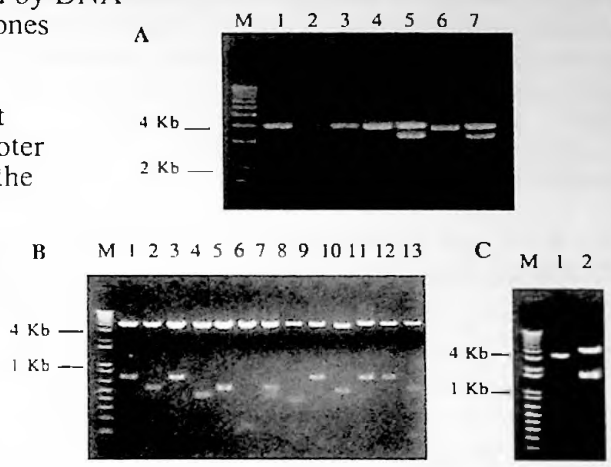

Figure 4. Restriction digestion screening of the systematically deleted sCFTR cDNA clones with. A: 1 to 5 are D3 clones; 6, 7 are D4 clones. 5 and 7 are right clones with 2 correct bands. B: 1-13 are individual clones of D5. C: the 5' 1.6 kb clone, 1 is undigested plasmid, 2 is digested with 2 correct bands.

Silent mutations were subsequently made in the third putative promoter. As shown in Figure 5, screening of full length sCFTR cDNA clones that contain multiple silent mutations introduced into both the highly conserved bacterial promoter and the one identified by systematic cloning of the 5' deleted sCFTR restriction digestion with *Nsi* I revealed a more regular digestion pattern and some promising clones that have an an insert similar to full-length sCFTR. More bacterial colonies were also observed. These results suggest that the mutations that disrupt the combined bacterial promoter sequences appear to significantly reduce the sCFTR sequence toxicity.

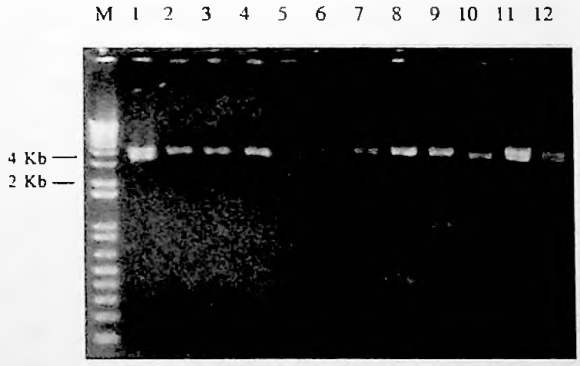

Figure 5. Restriction digestion, with *Nsi* I, of the full-length sCFTR cDNA clones with mutant promoters. M: 1 kb ladder. 1-12: individual clones. 1, 5, 11 and 12 are promising clones showing double bands similar to the expected ones.

In summary, several conclusions can be drawn from this work. First, it was demonstrated that the shark CFTR full length cDNA sequence can not be cloned with the common standard cloning procedures including the use of low copy number vectors and host strains. Second, three

putative toxic bacterial promoter-like sequence motifs were identified in the sCFTR cDNA by sequence comparison across species and by systematic cloning of the progressively deleted sCFTR cDNA fragments. The presence of these toxic bacterial promoters in sCFTR sequences is a major obstacle preventing the isolation of the full-length sCFTR clone. Third, taking together with the mutation analysis of the potential bacterial promoters, these results suggest that these bacterial promoter-like sequence motifs may function in concert to confer the sequence toxicity of the sCFTR cDNA in host bacteria, and that the full-length sCFTR cDNA may be cloned with mutagenesis of the toxic bacterial promoters. We are currently screening and sequencing the full-length sCFTR cDNA clones bearing multiple mutations in the bacterial promoters. Characterization of these potential promoter sequences with functional assays is also under consideration.

This work was supported by NIH grants DK 34208 and NIEHS P30-ES 3828 (Center for Membrane Toxicology Studies).