Plant Molecular Biology **16**: 947–954, 1991.

Identification of a new gene in an operon for cellulose biosynthesis in *Acetobacter xylinum*

Inder M. Saxena, Fong Chyr Lin and R. Malcolm Brown, Jr*
*Department of Botany, University of Texas at Austin, Austin, TX 78713-7640, USA (*author for correspondence)*

Received 10 December 1990; accepted 20 January 1991

Key words: *Acetobacter xylinum*, cellulose biosynthesis operon, cellulose synthase subunit gene, 93 kDa polypeptide, membrane protein

Abstract

DNA sequencing of the region downstream of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum* led to the identification of an open reading frame coding for a polypeptide of 86 kDa. The deduced amino acid sequence of this polypeptide matches from position 27 to 40 with the N-terminal amino acid sequence determined for a 93 kDa polypeptide that copurifies with the cellulose synthase catalytic subunit during purification of cellulose synthase. The cellulose synthase catalytic subunit gene and the gene encoding the 93 kDa polypeptide, along with other genes probably, are organized as an operon for cellulose biosynthesis in which the first gene is the catalytic subunit gene and the second gene codes for the 93 kDa polypeptide. The function of the 93 kDa polypeptide is not clear at present, however it appears to be tightly associated with the cellulose synthase catalytic subunit. Sequence analysis of the polypeptide shows that it is a membrane protein with a signal sequence at the N-terminal end and a transmembrane helix in the C-terminal region for anchoring it into the membrane.

Introduction

Cellulose biosynthesis in *Acetobacter xylinum* is believed to be a two-step process – polymerization of glucose molecules from the substrate UDP-glucose, catalyzed by the enzyme cellulose synthase, and crystallization of the nascent glucan chains into a ribbon of microfibrils. *In vivo*, these two steps are coupled resulting in the formation of native cellulose I [1]. The situation differs *in vitro*, where using solubilized membrane fraction or purified cellulose synthase, the cellulose obtained is of an altered crystalline form, namely cellulose II [4], suggesting that either the crystallization machinery has been disrupted or the intact cell has a role in the crystallization process. In order to identify the polypeptides involved in the synthesis of cellulose, cellulose synthase was purified by product entrapment [8]. Two major polypeptides of 83 and 93 kDa were identified in the purified enzyme preparation. The occurrence of these 2 polypeptides was always observed in the purified

---

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X54676.

cellulose synthase preparations, suggesting that they may be subunits of an oligomeric cellulose synthase complex. The 83 kDa polypeptide was conclusively identified to be the cellulose synthase catalytic subunit by photoaffinity labelling [9] and the gene for this polypeptide was cloned and sequenced using oligonucleotide probes designed from the N-terminal amino acid sequence of the polypeptide [14].

We now report on the identification of the gene for the 93 kDa polypeptide that is found downstream of the gene for the cellulose synthase catalytic subunit, and these two genes along with other genes possibly may be organized as a cellulose synthesizing operon with the first gene coding for the cellulose synthase catalytic subunit and the second gene coding for the 93 kDa polypeptide. The function of the 93 kDa polypeptide is not clear at present, but the very fact that it copurifies with the cellulose synthase catalytic subunit and that the two genes are linked suggests that it is a component of the cellulose synthase complex. A cellulose synthase operon comprising of 4 genes has recently been described for a different strain of *A. xylinum* in which the second gene in the operon (*bcsB*) is believed to encode the cellulose synthase catalytic subunit [16]. However from our sequence data it is clear that the first gene in the operon is the cellulose synthase catalytic subunit gene instead of the second gene as reported by Wong *et al.* [16].

Materials and methods

A. xylinum strain ATCC 53582 was used for purification of the cellulose synthase and isolation of the 93 kDa polypeptide. Plasmid pIS532, which carries the gene for the 93 kDa polypeptide has been described previously [14].

N-terminal amino acid sequence of the 93 kDa polypeptide was determined in the same manner as described for the cellulose synthase catalytic subunit [14].

For DNA sequencing of the 93 kDa polypeptide coding region, a *Pst* I fragment of 2.7 kb from pIS532 was cloned in M13mp19 in both orientations and deletion derivatives were prepared according to the method of Henikoff [7] using the Erase-a-base kit from Promega, Madison, WI. Sequencing was done by the dideoxy chain-termination procedure [13] using the Sequenase version 2.0 kit from U.S. Biochemical Corp., Cleveland, OH.

DNA and amino acid sequences were analysed using the PC/GENE programs and the Intelligenetics suite.

Results

Identification of the gene for the 93 kDa polypeptide from the N-terminal amino acid sequence

Purification of cellulose synthase by the product entrapment procedure, using trypsin-treated membranes of *A. xylinum*, always showed two major polypeptide bands of 83 and 93 kDa. The N-terminal amino acid sequence of both these polypeptides was determined to help in the identification of these polypeptides and to design oligonucleotide probes for isolating the genes. The 83 kDa polypeptide was identified by photolabelling studies to be the catalytic subunit of cellulose synthase [9], and the gene for this polypeptide was isolated using oligonucleotide probes [14]. DNA sequence beyond the stop codon of the cellulose synthase catalytic subunit gene, from plasmid pIS532, showed no transcription termination signal. Instead, an open reading frame with ATG start codon was observed at position 2860, suggesting the presence of other gene(s) downstream of the catalytic subunit gene [14]. Sequencing of the downstream region led to the identification of an open reading frame starting at position 2860 and terminating at position 5289, coding for a polypeptide of 86 kDa (Fig. 1). A potential ribosome-binding site with the sequence GGAACC is present 6 base pairs upstream of the start codon. The deduced amino acid sequence of this polypeptide from amino acid position 27 to 40 is found to be identical to the sequence of the first 14 N-terminal amino acids determined for the 93 kDa polypeptide that is shown here:

```
---Cellulose synthase catalytic subunit---->|
CATCGCAGTTCCCCAACCAAGCCTTTGGCTGGCAATGCCCTGTCTGACGATACGAACAAC          2820
P  S  Q  F  P  N  Q  A  F  G  W  Q  C  P  V  STOP
                                  |-->93 Kd polypeptide
CCGTCACGCAAGGAGCGTGTGCTGAAGGGAACCGTGAAAATGGTTTCGCTTCTGGCGCTG          2880
                                  M  V  S  L  L  A  L          7
CTGACATTTGCTTCCTCGGCACAGGCGGCGTCAGCGCCCAGGGCCGTCGCGGCGAAGGCC          2940
 L  T  F  A  S  S  A  Q  A  A  S  A  P  R  A  V  A  A  K  A         27
CCGGCCCATCAGCCCGAAGCCTCTGACCTGCCGCCGTTGCCCGCGCTTCTGCCGGCCACC          3000
 P  A  H  Q  P  E  A  S  D  L  P  P  L  P  A  L  L  P  A  T         47
AGCGGCGCGGCGCAGGCGGGTGCGGGCGATGCCGGCGCCAATGGACCCGGCAGCCCCACG          3060
 S  G  A  A  Q  A  G  A  G  D  A  G  A  N  G  P  G  S  P  T         67
GGCCAGCCCCTGGCGGCCGACAGTGCCGATGCGCTGGTGGAAAATGCGGAAAATACGTCC          3120
 G  Q  P  L  A  A  D  S  A  D  A  L  V  E  N  A  E  N  T  S         87
GATACGGCGACTGTCCATAATTATACCCTCAAGGATCTTGGCGCCGCAGGGTCCATCACG          3180
 D  T  A  T  V  H  N  Y  T  L  K  D  L  G  A  A  G  S  I  T        107
ATGCGTGGCCTTGCCCCGTTGCAGGGGATCGAGTTCGGGATTCCCTCTGACCAGCTTGTG          3240
 M  R  G  L  A  P  L  Q  G  I  E  F  G  I  P  S  D  Q  L  V        127
ACGTCCGCGCGCCTTGTGCTGTCGGGTTCGATGTCGCCCAACCTGCGTCCGGAAACTAAT          3300
 T  S  A  R  L  V  L  S  G  S  M  S  P  N  L  R  P  E  T  N        147
TCGGTCACAATGACGCTGAACGAGCAGTATATCGGCACGTTGCGTCCCGACCCGGCGCAC          3360
 S  V  T  M  T  L  N  E  Q  Y  I  G  T  L  R  P  D  P  A  H        167
CCGACATTCGGCCCCATGTCGTTCGAGATCAACCCGATCTTCTTCGTCAGCGGAAACCGT          3420
 P  T  F  G  P  M  S  F  E  I  N  P  I  F  F  V  S  G  N  R        187
CTGAACTTCAACTTCGCCTCCGGGTCGAAGGGATGTTCGGACATCACGAACGATACGCTG          3480
 L  N  F  N  F  A  S  G  S  K  G  C  S  D  I  T  N  D  T  L        207
TGGGCCACGATCTCGCAGAACTCGCAGCTTCAGATCACGACAATCGCGCTGCCGCCGCGT          3540
 W  A  T  I  S  Q  N  S  Q  L  Q  I  T  T  I  A  L  P  P  R        227
CGCCTGCTGTCACGTCTGCCGCAGCCCTTCTATGACAAGAACGTGCGCCAGCATGTTACG          3600
 R  L  L  S  R  L  P  Q  P  F  Y  D  K  N  V  R  Q  H  V  T        247
GTCCCGATGGTGCTGGCGCAGACCTATGACCCGCAGATACTCAAGTCCGCGGGGATTCTC          3660
 V  P  M  V  L  A  Q  T  Y  D  P  Q  I  L  K  S  A  G  I  L        267
GCTTCGTGGTTTGGCAAGCAGACGGACTTCCTAGGCGTGACGTTCCCGGTGTCGTCCACC          3720
 A  S  W  F  G  K  Q  T  D  F  L  G  V  T  F  P  V  S  S  T        287
ATCCCGCAGAGTGGCAACGCCATCCTGATCGGCGTGGCCGATGAACTGCCGACCAGCCTC          3780
 I  P  Q  S  G  N  A  I  L  I  G  V  A  D  E  L  P  T  S  L        307
GGGCGGCCGCAGGTCAATGGCCCGGCGGTTCTGGAACTGCCGAACCCGTCGGATGCAAAC          3840
 G  R  P  Q  V  N  G  P  A  V  L  E  L  P  N  P  S  D  A  N        327
GCCACGATCCTGGTGGTGACGGGGCGTGACCGTGATGAGGTCATTACCGCGAGCAAGGGC          3900
 A  T  I  L  V  V  T  G  R  D  R  D  E  V  I  T  A  S  K  G        347
ATCGCCTTCGCGTCTGCTCCCCTGCCGACCGACAGCCATATGGATGTCGCGCCGGTCGAT          3960
 I  A  F  A  S  A  P  L  P  T  D  S  H  M  D  V  A  P  V  D        367
ATCGCCCCGCGCAAGCCCAATGACGCGCCATCTTTTATCGCGATGGACCATCCGGTGCGC          4020
 I  A  P  R  K  P  N  D  A  P  S  F  I  A  M  D  H  P  V  R        387
TTTGGCGACCTTGTAACGGCCAGCAAGCTACAGGGAACCGGCTTTACGTCCGGTGTGCTG          4080
 F  G  D  L  V  T  A  S  K  L  Q  G  T  G  F  T  S  G  V  L        407
TCGGTTCCGTTCCGCATTCCGCCCGATCTTTATACGTGGCGTAACCGCCCGTACAAGATG          4140
 S  V  P  F  R  I  P  P  D  L  Y  T  W  R  N  R  P  Y  K  M        427
CAGGTACGTTTCCGTTCCCCCGCAGGGGAGGCGAAGGATGTCGAAAAGTCACGTCTCGAT          4200
 Q  V  R  F  R  S  P  A  G  E  A  K  D  V  E  K  S  R  L  D        447
GTCGGGATCAACGAGGTTTACCTGCATTCCTATCCGCTGCGGGAAACGCATGGCCTGGTT          4260
 V  G  I  N  E  V  Y  L  H  S  Y  P  L  R  E  T  H  G  L  V        467
GGCGCGGTTTTGCAGGGTGTCGGCCTTGCCCGCCCTGCAAGCGGCATGCAGGTGCATGAT          4320
 G  A  V  L  Q  G  V  G  L  A  R  P  A  S  G  M  Q  V  H  D        487
CTCGACGTACCGCCGTGGACCGTGTTCGGGCAGGATCAGTTGAACTTCTACTTTGACGCG          4380
 L  D  V  P  P  W  T  V  F  G  Q  D  Q  L  N  F  Y  F  D  A        507
ATGCCCCTTGCGCGCGGAATCTGTCAAAGTGGCGCGGCAAACAATGCGTTCCATCTTGGG          4440
 M  P  L  A  R  G  I  C  Q  S  G  A  A  N  N  A  F  H  L  G        527
```

```
CTTGACCCGGATTCCACCATTGATTTTTCCCGTGCCCATCACATCGCCCAGATGCCCAAC          4500
 L  D  P  D  S  T  I  D  F  S  R  A  H  H  I  A  Q  M  P  N    547
CTTGCCTATATGGCGACGGTCGGTTTTCCTTTCACCACCTATGCCGATCTGTCGCAGACG          4560
 L  A  Y  M  A  T  V  G  F  P  F  T  T  Y  A  D  L  S  Q  T    567
GCGGTTGTTCTGCCTGAACACCCGAATGCCGCGACTGTTGGCGCCTATCTTGACCTGATG          4620
 A  V  V  L  P  E  H  P  N  A  A  T  V  G  A  Y  L  D  L  M    587
GGGTTCATGGGCGCGGCGACGTGGTATCCGGTTGCGGGCGTGGACATTGTGTCGGCCGAT          4680
 G  F  M  G  A  A  T  W  Y  P  V  A  G  V  D  I  V  S  A  D    607
CATGTCAGTGACGTTGCGGACCGTAACCTGCTGGTGATTTCGACGCTGGCGACAAGTGGC          4740
 H  V  S  D  V  A  D  R  N  L  L  V  I  S  T  L  A  T  S  G    627
GAGATCGCGCCGCTGCTGTCACGTTCATCCTACGAAGTGGCGGATGGTCATTTGCGCACG          4800
 E  I  A  P  L  L  S  R  S  S  Y  E  V  A  D  G  H  L  R  T    647
GTGTCGCATGCGTCCGCGCTGGATAACGCGATCAAGGCGGTCGACGATCCGCTGACGGCT          4860
 V  S  H  A  S  A  L  D  N  A  I  K  A  V  D  D  P  L  T  A    667
TTCCGCGACCGTGACAGCAAGCCGCAGGATGTCGATACCCCCCTGACGGGTGGTGTCGGC          4920
 F  R  D  R  D  S  K  P  Q  D  V  D  T  P  L  T  G  G  V  G    687
GCGATGATCGAGGCGGAATCCCCCCTGACGGCGGGCCGCACCGTTCTGGCGCTGCTGTCG          4980
 A  M  I  E  A  E  S  P  L  T  A  G  R  T  V  L  A  L  L  S    707
TCTGACGGTGCGGGGCTGAACAACCTCCTTCAGATGTTGGGGGAGCGCAAGAAGCAGGCG          5040
 S  D  G  A  G  L  N  N  L  L  Q  M  L  G  E  R  K  K  Q  A    727
AATATCCAGGGGGACCTGGTTGTTGCCCATGGCGAGGACCTGTCTTCGTACCGCACTTCG          5100
 N  I  Q  G  D  L  V  V  A  H  G  E  D  L  S  S  Y  R  T  S    747
CCGGTCTATACGATCGGCACCCTGCCGCTCTGGCTGTGGCCGGACTGGTATATGCACAAC          5160
 P  V  Y  T  I  G  T  L  P  L  W  L  W  P  D  W  Y  M  H  N    767
AGACCGGTTCGTGTGCTGCTTGTCGGCCTGTTGGGATGTATTCTGATCGTCAGTGTTCTT          5220
 R  P  V  R  V  L  L  V  G  L  L  G  C  I  L  I  V  S  V  L    787
GCACGGGCTCTGGCGCGGCATGCGACCCGGCGTTTCAAGCAGCTTGAGGATGAGAGGCGC          5280
 A  R  A  L  A  R  H  A  T  R  R  F  K  Q  L  E  D  E  R  R    807
AAGTCGTGACCCATAAACGATATGCTTCGTCCCTGTCCGCCGGTCTTCTCGCAACGACCT          5340
 K  S  STOP                                                     809
GCGTCGCAGGTCTGTTGCTGCAG                                        5363
```

Fig. 1. DNA sequence and deduced amino acid sequence of the 93 kDa polypeptide of *A. xylinum* cellulose synthase. The DNA sequence is numbered such that it is in continuation of the sequence reported earlier for the cellulose synthase catalytic subunit gene [14]. Numbers to the right of each line in the first column are the amino acid positions and in the second column are the nucleotide positions. The sequence coding for the cellulose synthase catalytic subunit stops at nucleotide position 2804 and the sequence for the 93 kDa polypeptide starts from position 2860. A potential ribosome-binding site for the 93 kDa polypeptide is marked by double lines under the sequence. The signal sequence (amino acid positions 1–6) is shown in italics. N-terminal amino acid sequence determined for the 93 kDa polypeptide and found to be matching with the deduced amino acid sequence (positions 27–40) is underlined. The transmembrane segment (amino acid positions 586–606) is indicated by dotted underline.

Ala-Pro-Ala-Trp/His-Gln-Pro-Glu-Ala-Thr/Ser-Asp-Leu-Pro-Pro-Leu-Arg/Leu. The amino acid sequence identity between that determined for the 93 kDa polypeptide and one deduced from the DNA sequence confirms that this sequence codes for the 93 kDa polypeptide.

Signal sequence at the N-terminal end of the 93 kDa polypeptide

Since no N-terminal methionine is observed in the amino acid sequence determined for the 93 kDa polypeptide, the N-terminal sequence obtained could be either following cleavage of a signal peptide or following cleavage of a N-terminal fragment by trypsin which was used in the purification of cellulose synthase. In this instance, no N-terminal sequencing of the polypeptide, purified from

non-trypsin-treated membranes, was done to determine if trypsin had cleaved an N-terminal fragment as was shown for the cellulose synthase catalytic subunit [14]. The amino acid sequence derived from the DNA sequence shows occurrence of a number of potential signal peptide cleavage sites. Potential cleavage sites at the N-terminal end of the polypeptide are present at positions 14,15; 16,17; 18,19; 19,20; and 24,25. The cleavage site at position 16,17 showing the highest score conforms to the (−3, −1) rule [11, 15] and could be the actual cleavage site for the signal peptide. The amino acid sequence at the N-terminal end of the processed polypeptide still differs from the amino acid sequence determined for the 93 kDa polypeptide, suggesting that a few more amino acids were removed following cleavage of the signal peptide. Since the 93 kDa polypeptide was purified from trypsin-treated membranes, a short fragment of the processed polypeptide appears to be cleaved by trypsin at the N-terminal end. This is suggested by the amino acid sequence obtained for the 93 kDa polypeptide which is found to start after lysine at position 27 in the deduced amino acid sequence.

Analysis of the 93 kDa polypeptide from the deduced amino acid sequence

The 93 kDa polypeptide is a membrane protein, localized to the cytoplasmic membrane [2]. Analysis of the amino acid sequence deduced from the DNA sequence confirms this observation, and predicts 5 membrane-associated helices based on the average hydrophobicity of a 21 residue segment [5]. Tentative characterization of the predicted membrane-associated helices using hydrophobic moments plot coordinates classifies a single region (amino acid residues 586–606) as transmembrane multimeric, suggesting that this region may be involved in anchoring the polypeptide into the membrane and also in interacting with other helices in the membrane. A major region of this polypeptide is predicted to be projecting out of the cytoplasmic membrane, with the C-terminal region being present in the cytoplasm.

Discussion

DNA sequencing of region downstream of the cellulose synthase catalytic subunit gene led to the identification of an open reading frame coding for a polypeptide of 86 kDa, the N-terminal region of which matches with the N-terminal amino acid sequence of a 93 kDa polypeptide from purified cellulose synthase. The start codon for the 93 kDa polypeptide is present 54 bp after the stop codon of the 83 kDa gene suggesting that the 2 genes are part of an operon. This confirms the earlier view that the 93 kDa is a component of the cellulose synthase complex, though at present no specific function can be assigned to this polypeptide.

Whereas no signal peptide sequence was observed for the cellulose synthase catalytic subunit, a signal sequence is found for the 93 kDa polypeptide localizing it to the cytoplasmic membrane. However, for both the cellulose synthase catalytic subunit and the 93 kDa polypeptide, a short fragment at the N-terminal was found to be cleaved as observed from comparison of amino acid sequences determined by sequencing of these polypeptides with those deduced from the DNA sequence. In the case of the cellulose synthase catalytic subunit, the cleavage was due to trypsin treatment of membranes used for enzyme purification, and this was confirmed by sequencing of the 83 kDa polypeptide purified from membranes that were treated with trypsin and those that were not. However from the precursor 93 kDa polypeptide, a N-terminal signal sequence is cleaved by signal peptidase. A short N-terminal region from the mature 93 kDa polypeptide is further cleaved by trypsin during purification of cellulose synthase from trypsin-treated membranes as observed from the amino acid sequence.

Following our initial identification by cloning and sequencing of the cellulose synthase catalytic subunit gene from *A. xylinum* [14] and presence of the gene for the 93 kDa polypeptide downstream of the catalytic subunit gene, an operon for bacterial cellulose synthase was described in *A. xylinum* [16]. This operon was shown to consist of 4 genes in the order *bcsA*, *bcsB*, *bcsC* and *bcsD* transcribed as a polycistronic mRNA. The

second gene in the operon coding for polypeptide BcsB has been suggested to be the cellulose synthase catalytic subunit gene, based on amino acid sequence of a 91 kDa polypeptide purified by product entrapment from trypsin treated membranes and supported by complementation analysis of mutants. Comparison of the cellulose synthase catalytic subunit DNA and amino acid sequence obtained from our studies [14] with the sequence of the *bcsA*, *bcsB*, *bcsC* and *bcsD* genes and gene products shows that the cellulose synthase catalytic subunit gene is homologous to the *bcsA* gene. A 72% identity is observed for the DNA sequences of the cellulose synthase catalytic subunit gene [14] and the *bcsA* gene [16] and a 67% identity and 8% similarity is observed for the amino acid sequences deduced from the DNA sequences of these two genes, suggesting that *bcsA* is the catalytic subunit gene in place of *bcsB* as reported [16]. The DNA sequence for the 93 kDa polypeptide shows homology (68% identity) to the *bcsB* gene, and the amino acid sequences deduced from these two genes show 63% identity and 15% similarity. The organization of the genes for the cellulose synthase catalytic subunit and the 93 kDa polypeptide is therefore identical to the organization of the *bcsA* and the *bcsB* genes in the operon for cellulose biosynthesis. A search of the PIR and Swiss-Prot data base for homology to the 93 kDa polypeptide sequence showed no significant homologies with previously reported proteins.

Even though the nucleotide sequences that we have obtained for the cellulose synthase catalytic subunit gene [14] and the 93 kDa polypeptide gene are homologous to the sequences of the *bcsA* and *bcsB* genes respectively [16], they are not identical. This divergence could be due to the different strains of *A. xylinum* used in the two studies. Variability for cellulose production is commonly observed between different strains of *A. xylinum*, and this variability might very well be due to the differences in the DNA sequences coding for the various polypeptides involved in cellulose synthesis. Southern hybridization of DNA coding for the cellulose synthase catalytic subunit from *A. xylinum* ATCC 53582 with DNA from other strains of *A. xylinum* shows homology, but the size of DNA fragments showing hybridization and the intensity of the hybridization signal is found to vary with the strains (unpublished data) suggesting that there is a difference in the DNA sequences of these strains. Whereas some strains of *A. xylinum* seem closely related to one another, other strains are distantly related. Such variation of nucleotide sequences between different strains of bacteria is commonly observed, as in the case of the *trp* region in *Escherichia coli* [6]. A study of the extent of variation of nucleotide sequences involved in cellulose biosynthesis will be useful in understanding the evolutionary changes that have taken place in different strains of *A. xylinum* and other cellulose-producing organisms.

From our data it was conclusively shown that the 83 kDa polypeptide, obtained from purified cellulose synthase and identified by photolabelling with [^{32}P]5-azido UDP-glucose is the cellulose synthase catalytic subunit [9]. Moreover antibodies raised against the 93 kDa polypeptide, when used in cellulose synthase enzyme assays, showed no effect on enzyme activity suggesting that the 93 kDa polypeptide was not the catalytic subunit [2]. This is in disagreement with the data of Wong *et al.* [16], where the 91 kDa polypeptide which is homologous to the 93 kDa polypeptide of the present report, has been suggested to be the catalytic subunit. However there is no biochemical evidence in their report to suggest that the 91 kDa polypeptide is the catalytic subunit. In an earlier report, Mayer *et al.* [10] identified three polypeptide bands of 90, 67 and 57 kDa from purified cellulose synthase, and based on photoaffinity labelling using [^{32}P]c-di-GMP or [^{32}P]UDP-glucose, these investigators suggested that the activator- and substrate-specific binding sites are associated with the 67 kDa and 57 kDa polypeptide bands respectively. The 67 kDa polypeptide band has now been identified to be a tryptic fragment of the 91 kDa polypeptide [16], suggesting that the 91 kDa polypeptide may be an activator-binding subunit instead of being the catalytic subunit. It appears that no intact catalytic subunit was observed by Wong *et al.* [16] pos-

sibly due to cleavage by trypsin during purification of cellulose synthase. To corraborate that the first gene in the cellulose synthesizing operon is the cellulose synthase catalytic subunit gene, Wong *et al.* [16] have provided genetic complementation data where cellulose synthesis is restored in cellulose deficient mutants both by the *bcsA* and the *bcsB* genes. Moreover, normal levels of the BcsB protein (the 91 kDa polypeptide) were present in the mutants that were complemented by the *bcsA* gene making it obvious that the *bcsA* gene product missing in these mutants was the catalytic subunit. The *bcsB* gene apparently codes for the activator-binding subunit, and mutants which are defective in this gene product may appear cellulose-deficient even though they have the catalytic subunit.

Whether the cellulose synthase catalytic subunit can bind UDP-glucose on its own or requires some other polypeptides for the binding is not known at present. However we do know that the activator c-di-GMP enhances the binding of UDP-glucose to the catalytic subunit [9] suggesting that the activator-binding site is either present on the catalytic subunit or on another polypeptide that is tightly associated with it. The expression of the *A. xylinum* cellulose synthase catalytic subunit gene and the 93 kDa polypeptide gene from the *E. coli* lac promoter is being studied in an *E. coli in vitro* transcription and translation system for synthesis of these proteins to be used in UDP-glucose binding assays and later in the identification of the UDP-glucose binding site.

The finding that genes for cellulose biosynthesis are organized in an operon in *A. xylinum* and the identification of two major polypeptides, coded by genes in the operon, to be components of the cellulose synthase will be useful in understanding the arrangement of the various polypeptides into a structure that directs the synthesis of cellulose microfibrils. The proper arrangement of the different polypeptides into structures (pores?) which may be further organized (for example, as a linear row) appears to be necessary for the crystallization of nascent glucan chains into cellulose I [1]. Any change that would disturb the proper arrangement of these polypeptides could affect the crystallization process resulting in the formation of the thermodynamically favourable cellulose II as observed during *in vitro* synthesis of cellulose from membrane fractions [4] and *in vivo* in mutants of *A. xylinum* [12]. We believe that the cellulose synthase catalytic subunit and the 93 kDa polypeptide are tightly associated and are involved in the polymerization reaction. Other polypeptide(s), that probably are coded for by gene(s) in the cellulose synthesizing operon, may be loosely associated with the cellulose synthase complex and could therefore not be copurified with the cellulose synthase activity. These polypeptides may be involved in the crystallization process either directly by organizing the glucan chains for efficient, directed export or indirectly by organizing the various polypeptides into a pore-like structure [3] through which the glucan chain aggregates are guided facilitating the assembly of cellulose I microfibrils and ribbons.

Acknowledgements

We thank Richard Santos for technical help during DNA sequencing and sequence analysis. This work was supported by grant 90-37262-5520 from the USDA.

References

1. Benziman M, Haigler CH, Brown Jr RM, White AR, Cooper KM: Cellulose biogenesis: Polymerization and crystallization are coupled processes in *Acetobacter xylinum*. Proc Natl Acad Sci USA 77: 6678–6682 (1980).
2. Bokros CL: Immunochemical studies on partially purified cellulose synthase of *Acetobacter xylinum*. Thesis, University of Texas at Austin (1990).
3. Brown Jr RM, Willison JHM, Richardson CL: Cellulose biosynthesis in *Acetobacter xylinum*: Visualization of the site of synthesis and direct measurement of the in vivo process. Proc Natl Acad Sci USA 73: 4565–4569 (1976).
4. Bureau TE, Brown Jr RM: In vitro synthesis of cellulose II from a cytoplasmic membrane fraction of *Acetobacter xylinum*. Proc Natl Acad Sci USA 84: 6985–6989 (1987).
5. Eisenberg D, Schwarz E, Komaromy M, Wall R: Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179: 125–142 (1984).

6. Harshman L, Riley M: Conservation and variation of nucleotide sequences in *Escherichia coli* strains isolated from nature. J Bact 144: 560–568 (1980).
7. Henikoff S: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359 (1984).
8. Lin FC, Brown Jr RM: Purification of cellulose synthase from *Acetobacter xylinum*. In: Schuerch C (ed) Cellulose and Wood – Chemistry and Technology, pp. 473–492. John Wiley, New York (1989).
9. Lin FC, Brown Jr RM, Drake Jr RR, Haley BE: Identification of the uridine 5′-diphosphoglucose (UDP-Glc) binding subunit of cellulose synthase in *Acetobacter xylinum* using the photoaffinity probe 5-azido-UDP-Glc. J Biol Chem 265: 4782–4784 (1990).
10. Mayer R, Ross P, Weinhouse H, Amikam D, Volman G, Ohana P, Benziman M: The polypeptide substructure of bacterial cellulose synthase and its occurrence in higher plants. In: Abstracts from the Fifth Cell Wall Meeting (1989).
11. Perlman D, Halvorson HO: A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J Mol Biol 167: 391–409 (1983).
12. Roberts EM, Saxena IM, Brown Jr RM: Biosynthesis of cellulose II. In: Schuerch C (ed) Cellulose and Wood – Chemistry and Technology, pp 689–704. John Wiley, New York (1989).
13. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
14. Saxena IM, Lin FC, Brown Jr RM: Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. Plant Mol Biol 15: 673–683 (1990).
15. Von Heijne G: Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133: 17–21 (1983).
16. Wong HC, Fear AL, Calhoon RD, Eichinger GH, Mayer R, Amikam D, Benziman M, Gelfand DH, Meade JH, Emerick AW, Bruner R, Ben-Bassat A, Tal R: Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. Proc Natl Acad Sci USA 87: 8130–8134 (1990).