Cloning and sequencing of the cellulose synthase catalytic subunit gene of *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)*

Abstract

The gene for the catalytic subunit of cellulose synthase from *Acetobacter xylinum* has been cloned by using an oligonucleotide probe designed from the N-terminal amino acid sequence of the catalytic subunit (an 83 kDa polypeptide) of the cellulose synthase purified from trypsin-treated membranes of *A. xylinum*. The gene was located on a 9.5 kb *Hind* III fragment of *A. xylinum* DNA that was cloned in the plasmid pUC18. DNA sequencing of approximately 3 kb of the *Hind* III fragment led to the identification of an open reading frame of 2169 base pairs coding for a polypeptide of 80 kDa. Fifteen amino acids in the N-terminal region (positions 6 to 20) of the amino acid sequence, deduced from the DNA sequence, match with the N-terminal amino acid sequence obtained for the 83 kDa polypeptide, confirming that the DNA sequence cloned codes for the catalytic subunit of cellulose synthase which transfers glucose from UDP-glucose to the growing glucan chain. Trypsin treatment of membranes during purification of the 83 kDa polypeptide cleaved the first 5 amino acids at the N-terminal end of this polypeptide as observed from the deduced amino acid sequence, and also from sequencing of the 83 kDa polypeptide purified from membranes that were not treated with trypsin. Sequence analysis suggests that the cellulose synthase catalytic subunit is an integral membrane protein with 6 transmembrane segments. There is no signal sequence and it is postulated that the protein is anchored in the membrane at the N-terminal end by a single hydrophobic helix. Two potential N-glycosylation sites are predicted from the sequence analysis, and this is in agreement with the earlier observations that the 83 kDa polypeptide is a glycoprotein [13]. The cloned gene is conserved among a number of *A. xylinum* strains, as determined by Southern hybridization.

Introduction

Cellulose, a β-1,4 linked polymer of glucose, is synthesized by all plants, algae, some fungi, and a few bacteria [3]. It is also synthesized in the animal kingdom by tunicates [25] and is reported to be present in humans under certain disease conditions, notably scleroderma [8]. Almost all

cellulose obtained commercially is produced by plants, where it serves a structural role in the cell wall. Efforts to understand cellulose biosynthesis in plants have been limited by the inability to purify the cellulose synthase and demonstrate *in vitro* synthesis of cellulose from purified membrane proteins [3]. Among the other organisms that synthesize cellulose, the Gram-negative bacterium *Acetobacter xylinum* is the most studied for its capacity to synthesize cellulose. Cells of *A. xylinum* can utilize a variety of substrates for synthesizing cellulose and convert as much as 50% of supplied carbon to cellulose. The cellulose ribbons intertwine to form a pellicle that floats at the air-liquid interface of stationary cultures. The cellulose is produced in a pure form without being attached to other polysaccharides or proteins. Chemically, *Acetobacter* cellulose is indistinguishable from cellulose obtained from plants, but certain additional features such as strength, hydrophilicity, and the capacity to mold the native pellicle into any desired shape makes it superior for certain applications [26].

Biochemical studies in *A. xylinum* established that UDP-glucose is the substrate for cellulose biosynthesis and that no intermediate is involved in the reaction catalyzed by the enzyme cellulose synthase (UDP-glucose: 1,4-β-D-glucosyltransferase; E.C.2.4.1.12). *In vitro* synthesis of cellulose from cell-free preparations of *A. xylinum* was first demonstrated by Glaser in 1958 [7]. More efficient synthesis of cellulose *in vitro* was obtained when an activator of cellulose synthase was discovered [21]. The demonstration of cellulose production *in vitro* from membrane preparations of *A. xylinum* provided a reliable assay for cellulose synthase activity [14]. Cellulose synthase was first purified by product entrapment, and the purified preparation showed two major polypeptides of 93 and 83 kDa on lithium dodecyl sulfate (LDS)-polyacrylamide gel following electrophoresis [13]. The 83 kDa polypeptide has been shown by photoaffinity labelling to be the catalytic subunit of cellulose synthase [15]. Nothing substantial is known about the subunit composition or the nature of the native enzyme, although different molecular weights have been suggested.

In order to understand the nature of the catalytic subunit of *A. xylinum* cellulose synthase and determine its relationship with cellulose synthases from other organisms, we have cloned and sequenced the gene for the cellulose synthase catalytic subunit. The gene was isolated using oligonucleotide probes designed from the N-terminal amino acid sequence of the 83 kDa polypeptide. The amino acid sequence and properties of the polypeptide predicted from the sequence of the cloned gene were found to be in agreement with the N-terminal amino acid sequence and biochemical data available for the cellulose synthase catalytic subunit.

Materials and methods

Materials

Polynucleotide kinase (10000 U/ml) was obtained from New England Biolabs, Beverly, MA. [γ-^{32}P]ATP (6000 Ci/mmol), [α-^{32}P]dCTP (3000 Ci/mmol), and [^{35}S]dATP (1000–1350 Ci/mmol) were obtained from NEN Research Products, Boston, MA. Restriction endonucleases were obtained from BRL, Gaithersburg, MD. T4 Ligase was from BRL and Promega, Madison, WI. Oligonucleotides were synthesized by Operon Technologies, San Pablo, CA.

Strains and plasmids

A. xylinum strain ATCC 53582 was used in the present study for obtaining the 83 kDa polypeptide for N-terminal amino-acid sequence and as a source of DNA for cloning the gene. *A. xylinum* strains ATCC 10245, ATCC 23769, and NQ-3 were used for determining DNA homology with the oligonucleotide probes and the cloned cellulose synthase catalytic subunit gene. *E. coli* HB101 was used as the host strain for DNA cloning and JM103 was used as a host for M13 vectors for DNA sequencing. Plasmid pUC18 was used for cloning the *A. xylinum* DNA.

Amino-acid sequencing

For determining the N-terminal amino-acid sequence of the 83 kDa polypeptide, purified cellulose synthase was loaded on a 8% LDS-polyacrylamide gel and the individual polypeptides separated by electrophoresis [13]. The polypeptides from the polyacrylamide gel were electrophoretically transferred [17] to Immobilon-P transfer Membrane (Millipore). The membrane was stained with Coomassie blue and the 83 kDa polypeptide band was cut out with a razor blade. The polypeptide present on the excised membrane band was sequenced using an Applied Biosystems Model 477A Pulse Liquid Phase Sequenator (at the Department of Zoology, University of Texas at Austin).

Southern blotting and hybridization

DNA fragments, obtained following digestion with restriction endonucleases, were separated by agarose gel electrophoresis and transferred to nylon membrane (GeneScreen) using a rapid transfer protocol [20]. For hybridization, oligonucleotides were labelled with $[\gamma^{-32}P]ATP$ and polynucleotide kinase according to standard procedures [16]. Hybridizations were done for 16–24 h at 45 °C in 5 × SSC, 50 mM sodium phosphate (pH 7.0), 10 × Denhardt's solution, 7% SDS, and 100 µg/ml sheared and denatured salmon sperm DNA. Membranes were washed for 1 h at 45 °C in 3 × SSC, 10 mM sodium phosphate (pH 7.0), and 5% SDS followed by a wash in 1 × SSC and 1% SDS. For higher stringency, membranes were washed in 1 × SSC and 1% SDS at higher temperatures for 5–10 min. Membranes were exposed to X-ray films (X-Omat AR, Kodak) at − 80 °C for autoradiography.

DNA manipulations

Total A. xylinum DNA was isolated essentially according to the procedure followed for isolation of DNA from E. coli [24]. Plasmid DNA from E. coli was isolated according to the alkaline lysis procedure [1]. Single-stranded M13 DNA was isolated according to the procedure described in the BRL instruction manual for M13 cloning/dideoxy sequencing. DNA was digested with restriction endonucleases following the manufacturer's recommendations and the digestion products were analysed by electrophoresis, usually on a 0.7% agarose gel. DNA ligation was done using T4 ligase in buffer supplied by the manufacturer, either for 4 h at room temperature or overnight at 15 °C. For transformation, competent cells of E. coli were prepared according to the procedure of Hanahan [9] and were stored frozen at − 80 °C.

DNA sequencing

DNA fragments were cloned in the multiple cloning site of M13mp18 and M13mp19 for sequencing. DNA sequencing of both the strands was performed according to the dideoxy chain-termination procedure [22]. The DNA was labelled with $[^{35}S]$-dATP using the Sequenase version 2.0 kit from U.S. Biochemical Corp., Cleveland, OH. For a large region, both dGTP and dITP reactions were performed to resolve regions of compression. Sequencing from the end was done using the primer provided in the Sequenase kit. For sequencing regions that were distant from the ends of some clones, synthetic oligonucleotide primers designed from the DNA sequence were used. The DNA sequence was analyzed by using the PC/GENE programs (IntelliGenetics, Inc.).

Results

N-terminal amino acid sequence of the 83 kDa polypeptide

During the purification of cellulose synthase from A. xylinum ATCC 53582, it was observed by LDS-polyacrylamide gel electrophoresis that

trypsin treatment of either the washed membranes or the detergent-solubilized enzyme preparation resulted in the degradation of polypeptides other than the 83 kDa polypeptide and at the same time maintained the cellulose synthase activity [13]. Since trypsin treatment proved to be an efficient method for obtaining relatively pure 83 kDa polypeptide, N-terminal amino acid sequencing of the 83 kDa polypeptide purified from membrane preparations that were treated with trypsin was done. The amino acid sequence obtained is shown below:

Ser-Ser-Thr-Gln-Ser-Glu-Ser-Gly-Met-Ser-Gln-Trp/Leu-Met-Gly-Lys-

Design of oligonucleotide probes

For designing oligonucleotide probes from the amino acid sequence, a stretch of 6 amino acids (-Met-Ser-Gln-Trp/Leu-Met-Gly-) was selected from the sequence of the 83 kDa polypeptide that was obtained from trypsin-treated membrane preparation. Eight 17-mer oligonucleotides, each with a unique sequence and having deoxyinosine at the ambiguous position of the serine codons were synthesized. The sequence of the oligonucleotides is given below:

```
83-1A  5'-ATGTCICAATGGATGGG-3'
83-1G  5'-ATGTCICAGTGGATGGG-3'
83-2A  5'-ATGAGICAATGGATGGG-3'
83-2G  5'-ATGAGICAGTGGATGGG-3'
83-3A  5'-ATGTCICAACTGATGGG-3'
83-3G  5'-ATGTCICAGCTGATGGG-3'
83-4A  5'-ATGAGICAACTGATGGG-3'
83-4G  5'-ATGAGICAGCTGATGGG-3'
         I = deoxyinosine
```

Identification of A. xylinum DNA fragments homologous to oligonucleotide probes

To determine if the oligonucleotides showed any homology with A. xylinum DNA, individual oligonucleotides were used as probes in Southern
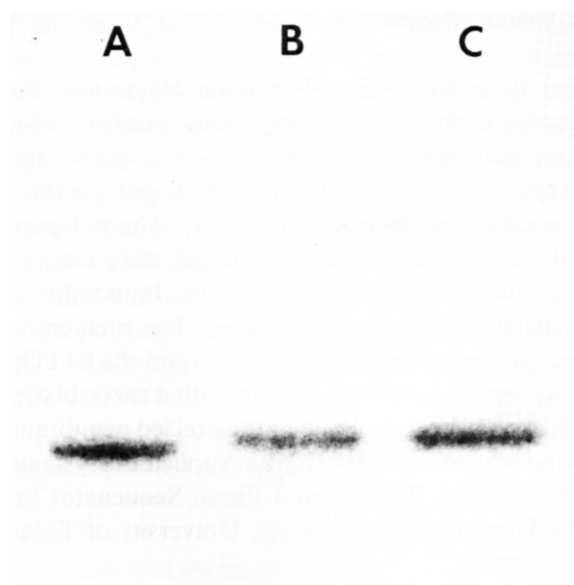

Fig. 1. Autoradiograph showing hybridization of ^{32}P-labelled oligonucleotide 83-1G with HindIII-digested DNA from A. xylinum strains, A, ATCC 23769; B, ATCC 53582; and C, ATCC 53582-2, a mutant of ATCC 53582 which produces cellulose II.

hybridization with A. xylinum ATCC 53582 DNA digested with HindIII. Under conditions of low stringency washing, a single fragment of 9.5 kb showed an intense signal of hybridization with oligonucleotide probes 83-1A and 83-1G, and a faint signal with 83-2G. However, with more stringent washings no signal was observed in membranes which had been probed with 83-1A and 83-2G. Similar washings did not remove the signal from membranes probed with 83-1G, suggesting that a DNA sequence highly homologous to the oligonucleotide 83-1G sequence was present in A. xylinum. Probing of HindIII-cleaved DNA from A. xylinum ATCC 23769 and a mutant of A. xylinum ATCC 53582 defective in cellulose production [23] with oligonucleotide 83-1G similarly showed a single intense band of hybridization (Fig. 1). Less intense hybridization signals were observed when DNA from A. xylinum strains NQ-3 and ATCC 10245 were probed with 83-1G, suggesting that the 83-1G sequence may be partly conserved in different strains of A. xylinum. DNA from A. xylinum strains ATCC

53582 and ATCC 23769 cleaved with restriction enzymes *Bgl*II, *Cla*I, *Nar*I, *Nru*I, and *Sma*I and probed with 83-1G showed a single band of hybridization in all cases, suggesting that the 83-1G homologous sequence is present in a single copy in the *A. xylinum* genome.

Cloning of the cellulose synthase catalytic subunit gene from A. xylinum

For cloning the cellulose synthase catalytic subunit gene from *A. xylinum*, DNA from *A. xylinum* ATCC 53582 was digested with *Hind*III and the fragments separated on an agarose gel. DNA fragments in the size range of 9.5 kb were extracted from the gel slice and ligated to *Hind*III-cleaved plasmid pUC18. The ligation mix was used to transform *E. coli* HB101, and plasmid pools from ampicillin-resistant transformants were screened by hybridization to 83-1G. A plasmid, pIS532, was identified from this screening to carry a DNA fragment homologous to the oligonucleotide 83-1G. pIS532 carried a 9.5 kb *Hind*III fragment of *A. xylinum* ATCC 53582. Southern hybridization of pIS532 restriction fragments with 83-1G led to the identification of the region on the 9.5 kb *Hind*III fragment that carried the gene for the cellulose synthase catalytic subunit.

N-terminal amino acid sequence analysis

A region of approximately 3 kb from the plasmid pIS532 was sequenced to obtain the DNA sequence of the cellulose synthase catalytic subunit gene. From the sequence obtained, an ORF of 2169 bp was identified that coded for a polypeptide of 80218 Da (Fig. 2). The N-terminal amino acid sequence of the cellulose synthase catalytic subunit polypeptide that had been obtained from trypsin-treated membrane preparation was identical to the amino acid sequence from position 6 to position 20 of the amino acid sequence deduced from the DNA sequence. In the N-terminal sequence of the cellulose synthase catalytic subunit polypeptide it was unclear if position 12 was occupied by tryptophan or leucine, and since this amino acid position was used in the design of the oligonucleotides, the single codon of tryptophan (TGG) was used in 4 of the oligonucleotides while the other 4 oligonucleotides had the sequence for the most frequently used leucine codon (CTG). From the DNA sequence and the deduced amino acid sequence it is clear that this position is occupied by tryptophan. Since no N-terminal methionine residue was observed in the amino acid sequence of the cellulose synthase catalytic subunit polypeptide that was purified from trypsin-treated membrane preparations, it was uncertain whether this amino acid sequence was the N-terminal end of a processed polypeptide from which a signal sequence had been cleaved, or whether the N-terminal region of the native cellulose synthase catalytic subunit polypeptide had been cleaved by trypsin. Amino acid sequencing of the cellulose synthase catalytic subunit polypeptide obtained from membrane preparations that had not been treated with trypsin was also performed. Even though the sequence obtained showed ambiguities at a number of positions, it unambiguously showed methionine at the N-terminus and the sequence -Glu-Val-Arg-X-X-Thr-Gln- which matches with positions 3 to 9 of the amino acid sequence deduced from the DNA sequence shown in Fig. 2. From the N-terminal amino acid sequence of the cellulose synthase catalytic subunit purified from trypsin-treated membranes and from the amino acid sequence deduced from the DNA sequence, it appears that trypsin cleaved five amino acids from the N-terminal end without in any way affecting the enzyme activity. Such a cleavage would cause no detectable change in the mobility of the polypeptide during polyacrylamide gel electrophoresis, and no such change was indeed observed.

Sequence analysis of the cellulose synthase catalytic subunit gene

The overall G + C content of the gene for the cellulose synthase catalytic subunit was found to

```
CGCTTCCCGCACATCAGCATCGTCCATAAGTGAGCGTCCGGTGACAGGGTGTTGCCGATG          60
AATGGAGTCCTGTTCCGAAGCACGCCTCAGCCGCGGGACGTCCGCGTCCCGGCGATCTGA         120
GCTTTTTCTTTCAGGGGATGCGCGACACCCGTGATGAAAAGAAGTTCTTTCCCGTGGCGT         180
CCACGCGATCAGTTCGTTCTAATGTTTCCAGGATGACCAGCATGACCAAGACAGACACGA         240
ATTCCTCTCAGGCTTCTCGTCCCGGCAGCCCCGTCGCCTCGCCTGATGGGTCGCCCACAA         300
TGGCCGAAGTGTTCATGACGCTGGGTGGTCGTGCGACGGAACTCCTCAGCCCCCGTCCTT         360
CGCTGCGGGAGGCGCTGTTGCGTCGTCGTGAAAACGAAGAAGAATCCTAAGGCCCTATAT         420
TCAGGCGACCGCCTGTCCCTGTGTCCTGCCTGGTGAAATGGCGGGGGCAGGGGCAGACAC         480
GCTGCACTGGCTCGGCGATCTATTCCGTCTTTGGCCTGGGTGATGTTTCCGCAGCCCAGG         540
CGGGCCCGCTTTGCCCCGATGTGACCGCAGATGGTCGGGGTCAGGTTTTTTTTGACGTAA         600
TTTTTCTGTTTTACAGCATTTCGGACGAGTTGTTTATGCCAGAGGTTCGGTCGTCAACGC         660
                              M P E V R S S T          8
AGTCAGAGTCAGGAATGTCACAGTGGATGGGGAAAATTCTTTCCATTCGCGGTGCTGGGC         720
Q S E S G M S Q W M G K I L S I R G A G                28
TGACTATTGGTGTTTTTGGCCTGTGTGCGCTGATTGCGGCTACGTCCGTGACCCTGCCGC         780
L T I G V F G L C A L I A A T S V T L P                48
CAGAACAGCAGTTGATTGTGGCATTTGTATGTGTCGTGATCTTTTTTATTGTCGGTCATA         840
P E Q Q L I V A F V C V V I F F I V G H                68
AGCCCAGCCGTCGGTCCCAGATTTTCCTTGAAGTGCTGTCAGGGCTGGTTTCGCTGCGCT         900
K P S R R S Q I F L E V L S G L V S L R                88
ATCTGACATGGCGCCTGACGGAAACGCTTTCATTCGATACATGGTTGCAGGGTCTGCTTG         960
Y L T W R L T E T L S F D T W L Q G L L                108
GGACAATGCTTCTGGTGGCGGAACTTTACGCCCTGATGATGCTGTTCCTCAGCTATTTCC        1020
G T M L L V A E L Y A L M M F L S Y F                 128
AGACGATCGCGCCATTGCATCGTGCGCCTCTGCCGCTGCCGCCGAACCCTGACGAATGGC        1080
Q T I A P L H R A P L P L P P N P D E W                148
CCACGGTCGATATCTTCGTCCCGACCTACAACGAAGAACTGAGCATTGTCCGCCTGACGG        1140
P T V D I F V P T Y N E E L S I V R L T                168
TGCTGGGATCACTGGGGATTGACTGGCCACCGGAAAAGGTGCGGGTTCATATCCTTGATG        1200
V L G S L G I D W P P E K V R V H I L D                188
ACGGTCGTCGTCCTGAATTCGCCGCCTTTGCCGCTGAATGTGGCGCGAATTATATCGCCC        1260
D G R R P E F A A F A A E C G A N Y I A                208
GCCCGACGAACGAACATGCAAAGGCCGGTAATCTTAACTATGCCATTGGTCATACCGATG        1320
R P T N E H A K A G N L N Y A I G H T D                228
GTGATTACATCCTGATCTTTGACTGCGACCACGTCCCGACCCGCGCCTTCCTGCAGTTGA        1380
G D Y I L I F D C D H V P T R A F L Q L                248
CAATGGGCTGGATGGTCGAAGACCCGAAGATCGCGCTGATGCAGACCCCGCATCACTTCT        1440
T M G W M V E D P K I A L M Q T P H H F                268
ATTCCCCCGACCCGTTCCAGCGGAACCTGTCGGCTGGTTATCGCACCCCGCCCGAAGGCA        1500
Y S P D P F Q R N L S A G Y R T P P E G                288
ACCTGTTTTATGGCGTGGTGCAGGATGGCAACGATTTCTGGGATGCGACCTTCTTTTGCG        1560
N L F Y G V V Q D G N D F W D A T F F C                308
GGTCATGTGCAATCCTGCGTCGCACGGCGATTGAGCAGATCGGCGGCTTTGCGACCCAGA        1620
G S C A I L R R T A I E Q I G G F A T Q                328
CCGTGACCGAAGACGCGCATACCGCACTCAAGATGCAGCGTCTGGGCTGGTCCACGGCCT        1680
T V T E D A H T A L K M Q R L G W S T A                348
ATCTGCGTATCCCGCTTGCCGGTGGTCTCGCGACGGAACGCCTGATCCTGCATATCGGAC        1740
Y L R I P L A G G L A T E R L I L H I G                368
```

```
AGCGCGTGCGCTGGGCGCGTGGGATGCTGCAGATCTTCCGCATCGACAATCCTCTGTTCG          1800
Q   R   V   R   W   A   R   G   M   L   Q   I   F   R   I   D   N   P   L   F     388
GGCGTGGCCTGTCATGGGGGCAGCGGCTTTGTTACCTGTCGGCCATGACGTCGTTCCTGT          1860
G   R   G   L   S   W   G   Q   R   L   C   Y   L   S   A   M   T   S   F   L     408
TCGCTGTCCCGCGCGTCATCTTCCTGAGCTCCCCGCTGGCGTTCCTGTTCTTTGGGCAGA          1920
F   A   V   P   R   V   I   F   L   S   S   P   L   A   F   L   F   F   G   Q     428
ACATCATTGCCGCGTCGCCGCTCGCGCTGCTGGCCTATGCCATCCCGCACATGTTCCACG          1980
N   I   I   A   A   S   P   L   A   L   L   A   Y   A   I   P   H   M   F   H     448
CCGTCGGCACGGCGTCGAAGATCAACAAGGGCTGGCGCTACTCCTTCTGGAGTGAGGTCT          2040
A   V   G   T   A   S   K   I   N   K   G   W   R   Y   S   F   W   S   E   V     468
ATGAAACCACGATGGCGCTGTTCCTGGTGCGCGTGACGATTGTCACCCTGCTCAGCCCTT          2100
Y   E   T   T   M   A   L   F   L   V   R   V   T   I   V   T   L   L   S   P     488
CACGTGGGAAGTTCAACGTGACGGACAAGGGCGGGGTTGCTTGAAAAAGGTTATTTCGACC          2160
S   R   G   K   F   N   V   T   D   K   G   G   L   L   E   K   G   Y   F   D     508
TTGGCGCTGTCTACCCGAACATCATCCTTGGCCTGATCATGTTCGGCGGCCTGGCGCGTG          2220
L   G   A   V   Y   P   N   I   I   L   G   L   I   M   F   G   G   L   A   R     528
GTGTCTATGAACTGTCTTTCGGCCATCTCGACCAGATCGCCGAACGTGCCTACCTGCTGA          2280
G   V   Y   E   L   S   F   G   H   L   D   Q   I   A   E   R   A   Y   L   L     548
ACTCCGCCTGGGCAATGCTCAGCCTCATCATCATCCTTGCGGCCATCGCCGTGGGGCGTG          2340
N   S   A   W   A   M   L   S   L   I   I   I   L   A   A   I   A   V   G   R     568
AAACACAGCAGAAACGCAACAGTCATCGCATCCCCGCAACCATCCCGGTGGAAGTGGCGA          2400
E   T   Q   Q   K   R   N   S   H   R   I   P   A   T   I   P   V   E   V   A     588
ATGCCGATGGGTCCATCATCGTGACGGGCGTGACCGAGGACCTGTCCATGGGTGGGGCCG          2460
N   A   D   G   S   I   I   V   T   G   V   T   E   D   L   S   M   G   G   A     608
CGGTGAAGATGTCATGGCCTGCGAAGCTGTCGGGGCCGACGCCGGTTTATATCCGTACTG          2520
A   V   K   M   S   W   P   A   K   L   S   G   P   T   P   V   Y   I   R   T     628
TCCTTGACGGGGAGGAACTGATCCTTCCCGCCAGGATCATCCGTGCTGGCAACGGACGGG          2580
V   L   D   G   E   E   L   I   L   P   A   R   I   I   R   A   G   N   G   R     648
GGATTTTCATCTGGACGATTGATAACCTGCAGCAGGAATTCTCGGTTATCCGTCTGGTGT          2640
G   I   F   I   W   T   I   D   N   L   Q   Q   E   F   S   V   I   R   L   V     668
TCGGCCGTGCCGACGCATGGGTTGACTTGGGGCAATTACAAGGCCGACCGCCCGCTGCTC          2700
F   G   R   A   D   A   W   V   D   L   G   Q   L   Q   G   R   P   P   A   A     688
AGCCTCATGGACATGGTTCTCAGCGTCAAGGGCCTGTTCCGTTCAAGTGGCGATATCGTC          2760
Q   P   H   G   H   G   S   Q   R   Q   G   P   V   P   F   K   W   R   Y   R     708
CATCGCAGTTCCCCAACCAAGCCTTTGGCTGGCAATGCCCTGTCTGACGATACGAACAAC          2820
P   S   Q   F   P   N   Q   A   F   G   W   Q   C   P   V   STOP             723
CCGTCACGCAAGGAGCGTGTGCTGAAGGGAACCGTGAAAATGGTTTCGCTTCTGGCGCTG          2880
CTGACATTTGCTTCCTCGGCACAGGCGGCGTC                                      2912
```

Fig. 2. DNA sequence and deduced amino acid sequence of the *A. xylinum* cellulose synthase catalytic subunit. Numbers to the right of each line refer to the amino acid position in the first column and to the nucleotide position in the second column. Double line under the nucleotide sequence indicates potential ribosome-binding site. N-terminal amino acid sequence determined by sequencing of the polypeptide purified from trypsin-treated membranes is underlined (amino acid positions 6–20) and inner boundaries of predicted transmembrane segments are indicated by wavy underlines (amino acid positions 29–45, 52–68, 107–123, 430–446, 511–527, 551–567).

be 58 mol%, with values for codon positions 1, 2, and 3 to be 60, 44, and 69 mol% respectively. That the ATG at position 636 of the DNA sequence is the start codon is suggested from the open reading frame analysis of the complete sequence, since the polypeptide coded by the ORF starting at this position is close to the molecular weight determined for the cellulose synthase catalytic subunit. In this reading frame no other ATG was observed between the stop codon at position 408 and the ATG at position 636. The N-terminal amino acid sequence of the cellulose synthase catalytic subunit from trypsin-treated membranes matched with the deduced

amino acid sequence of this ORF. The presence of sequences corresponding to the *E. coli* promoter have not been observed upstream of the start codon of the cellulose synthase catalytic subunit gene, nor have they been observed anywhere else in the complete sequence. Preliminary data suggest that the cellulose synthase catalytic subunit gene is not expressed in *E. coli* when it is under the control of its own promoter. Genes cloned from other *Acetobacter* species similarly have been shown to carry promoters that do not function in *E. coli* [5, 6]. A potential ribosome-binding site GGACGAGT is present 5 base pairs upstream of the initiating ATG codon.

Discussion

Cellulose biosynthesis in *A. xylinum* has been studied in considerable detail at the ultrastructural level. More recently the final step of the pathway and its regulation has become better understood at the biochemical level. Cellulose biosynthesis is believed to be controlled *in vivo* by a complex multi-component regulatory system, and *in vitro* by the allosteric activator c-di-GMP [21]. This suggests that cellulose synthase has at least 2 sites, a catalytic site and a regulatory site, to function efficiently. Whether these sites are present on a single polypeptide chain or on separate subunits of an enzyme complex [18] is not very clear at present. The native enzyme has however been suggested to be oligomeric in nature with a M_r in the range of 400–500 kDa [13, 18].

From the DNA sequence, the cellulose synthase catalytic subunit has been predicted to have 723 amino acids and a M_r of 80 kDa. The amino acid composition of the polypeptide (Table 1) shows that it is hydrophobic in nature (Leu + Ala + Ile + Pro + Val + Phe + Met + Trp = 50.2 mol%), in agreement with its observed location in the cytoplasmic membrane [2]. The hydropathicity plot determined by the method of Kyte and Doolittle [11] shows that the polypeptide is composed of a number of hydrophobic regions (Fig. 3). Analysis of the amino acid sequence by the method of Eisenberg *et al.*

Table 1. Amino acid composition of *A. xylinum* cellulose synthase catalytic subunit deduced from the DNA sequence.

Amino acid	%
Leucine	11.4
Alanine	8.8
Glycine	8.7
Isoleucine	7.1
Proline	6.2
Valine	6.3
Arginine	6.2
Threonine	5.6
Phenylalanine	5.6
Serine	5.6
Glutamine	4.2
Glutamic acid	4.0
Aspartic acid	3.5
Tyrosine	2.9
Asparagine	2.7
Methionine	2.4
Tryptophan	2.4
Histidine	2.2
Lysine	2.0
Cysteine	1.1

[4] predicts that the polypeptide has 10 membrane-associated helices based on the average hydrophobicity of a 21 residue segment. The positions of the membrane-associated helices were classified with the aid of the hydrophobic moment plot by the method of Eisenberg *et al.* [4] and using the PC/GENE program Helixmem into monomeric transmembrane, multimeric transmembrane, globular and surface. Of the 6 transmembrane helices in the cellulose synthase catalytic subunit, the first helix was determined to be monomeric transmembrane based on highest hydrophobicity and lowest hydrophobic moment. Such monomeric transmembrane regions are believed to be membrane anchors [4]. From such an analysis, it appears that the cellulose synthase catalytic subunit is anchored in the membrane at its N-terminal end by a single hydrophobic helix. The remaining 5 transmembrane segments have the potential to interact with other helices. The inner boundaries of the 6 transmembrane segments predicted by the method of Eisenberg *et al.* [4] and determined by the method of Klein *et al.* [10] are shown in Fig. 2. No signal sequence has
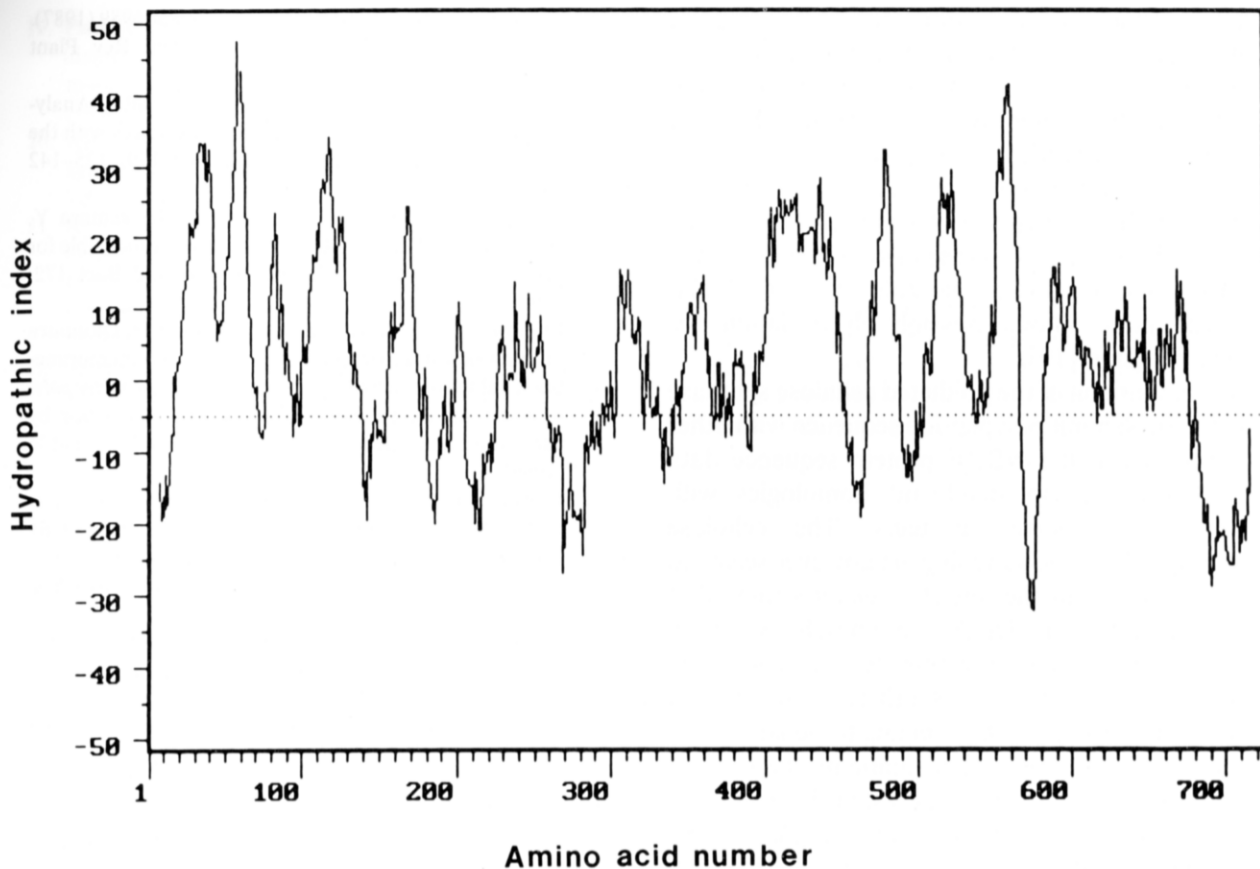
Fig. 3. Hydropathic profile of the predicted cellulose synthase catalytic subunit polypeptide calculated according to Kyte and Doolittle [11] using a window of 15 amino acids.

been observed at the N-terminus of the deduced amino acid sequence; therefore, it is very likely that the cellulose synthase catalytic subunit is anchored in the membrane by a hydrophobic helix. An understanding of the interaction of the cellulose synthase catalytic subunits among themselves and with other putative subunits of the cellulose synthase complex will be important in determining the organization of cellulose synthase in the cytoplasmic membrane and the way this organization influences cellulose production. Such an understanding will be made possible by using gene fusions and mutational analysis with the cloned cellulose synthase catalytic subunit gene for identifying regions of the polypeptide chain involved in interaction with various components of the cellulose synthase structural and

regulatory components as well as the cytoplasmic membrane.

The cellulose synthase catalytic subunit was previously characterized as a glycoprotein by lectin affinity chromatography, and Schiff-periodate and fluorescein isothiocyanate-concanavalin A staining analysis [13]. From sequence analysis of the cellulose synthase catalytic subunit polypeptide, 2 N-glycosylation sites (amino acid positions 277 and 294) were predicted based on the consensus sequence Asn-X-Ser/Thr. Even though this consensus sequence is derived from studies with eukaryotic glycoproteins, it has also been shown to be the acceptor sequence for glycosylation in bacteria [12]. The specific role, if any, that glycosylation may play in the organization or function of cellu-

lose synthase activity is not clear. In this respect, it is interesting that a set of integral membrane proteins related to gliding motility in *Cytophaga johnsonae* have been suggested to be glycoproteins [19]. Furthermore, these protein bands were found to shift to lower positions on SDS-polyacrylamide gels in some of the non-motile mutants and return to the wild-type position in some motile revertants, suggesting the role of post-translational modifications, possibly glycosylation, for this variation [19].

Comparison of the predicted cellulose synthase catalytic subunit polypeptide sequence with other sequences in the NBRF protein sequence data base revealed no significant homologies with previously reported proteins. The cellulose synthase catalytic subunit gene however seems to be conserved in the few *A. xylinum* strains that were analyzed for DNA homology by Southern hybridization. Whether this gene is conserved among the other cellulose synthesizing organisms, including higher plants, remains to be studied. If there exists a common enzymatic mechanism for cellulose synthesis, as suggested by immuno-chemical analysis [18], then the *A. xylinum* gene will be a suitable probe for isolating and identifying the corresponding cellulose synthase gene from higher plants.

Acknowledgements

References

1. Birnboim HC: A rapid alkaline extraction method for the isolation of plasmid DNA. Meth Enzymol 100: 243–255 (1983).
2. 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).
3. Delmer DP: Cellulose biosynthesis. Ann Rev Plant Physiol 38: 259–290 (1987).
4. 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).
5. Fukaya M, Takemura H, Okumura H, Kawamura Y, Horinouchi S, Beppu T: Cloning of genes responsible for acetic acid resistance in *Acetobacter aceti*. J Bact 172: 2096–2104 (1990).
6. Fukaya M, Tayama K, Tamaki T, Tagami H, Okumura H, Kawamura Y, Beppu T: Cloning of the membrane-bound aldehyde dehydrogenase gene of *Acetobacter polyoxogenes* and improvement of acetic acid production by use of the cloned gene. Appl Envir Microbiol 55: 171–176 (1989).
7. Glaser L: The synthesis of cellulose in cell-free extracts of *Acetobacter xylinum*. J Biol Chem 232: 627–636 (1958).
8. Hall DA, Happey F, Lloyd PF, Saxl H: Oriented cellulose as a component of mammalian tisue. Proc Roy Soc London B151: 497–516 (1960).
9. Hanahan D: Techniques for transformation of *E. coli*. In: Glover DM (ed) DNA Cloning Vol 1: a practical approach, pp 109–135. IRL Press, Oxford (1985).
10. Klein P, Kanehisa M, DeLisi C: The detection and classification of membrane-spanning proteins. Biochim Biophys Acta 815: 468–476 (1985).
11. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
12. Lechner J, Wieland F: Structure and biosynthesis of prokaryotic glycoproteins. Ann Rev Biochem 58: 173–194 (1989).
13. 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 and Sons, New York (1989).
14. Lin FC, Brown Jr RM, Cooper JB, Delmer DP: Synthesis of fibrils *in vitro* by a solubilized cellulose synthase from *Acetobacter xylinum*. Science 230: 822–825 (1985).
15. 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).
16. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
17. Matsudaira P: Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem 262: 10035–10038 (1987).
18. 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 Fifth Cell Wall meeting (1989).

19. Pate JL, De Jong DM: Use of nonmotile mutants to identify a set of membrane proteins related to gliding motility in *Cytophaga johnsonae*. J Bact 172: 3117–3124 (1990).

20. Reed KC, Mann DA: Rapid transfer of DNA from agarose gels to nylon membranes. Nucl Acids Res 13: 7207–7221 (1985).

21. Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, Benziman M: Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. Nature 325: 279–281 (1987).

22. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).

23. Saxena IM, Brown Jr RM: Cellulose biosynthesis in *Acetobacter xylinum*: a genetic approach. In: Schuerch C (ed) Cellulose and Wood – Chemistry and Technology, pp 537–557. John Wiley and Sons, New York (1989).

24. Silhavy TJ, Berman ML, Enquist LW: Experiments with Gene Fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984).

25. Wardrop AB: The structure and formation of the test of *Pyura stolonifera* (*Tunicata*). Protoplasma 70: 73–86 (1970).

26. White DG, Brown Jr RM: Prospects for the commercialization of the biosynthesis of microbial cellulose. In: Schuerch C (ed) Cellulose and Wood – Chemistry and Technology, pp 573–590. John Wiley and Sons, New York (1989).