PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82, 9/10, 15/29, 15/54, C12Q 1/68, A01H 5/00

A2

(11) International Publication Number: **WO 98/18949**

(43) International Publication Date: 7 May 1998 (07.05.98)

(21) International Application Number: PCT/US97/19529

(22) International Filing Date: 29 October 1997 (29.10.97)

(30) Priority Data:
60/029,987 29 October 1996 (29.10.96) US

(71) Applicant *(for all designated States except US)*: CALGENE, INC. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: STALKER, David, M. [US/US]; 870 W. Southwood, Woodland, CA 95695 (US). PEAR, Julie, R. [US/US]; 818 Douglass Avenue, Davis, CA 95616 (US). DELMER, Deborah, P. [US/US]; 29 West Mill Street, New Palestine, IN 46163 (US).

(74) Agents: SCHWEDLER, Carl, J. et al.; Calgene, Inc., 1920 Fifth Street, Davis, CA 95616 (US).

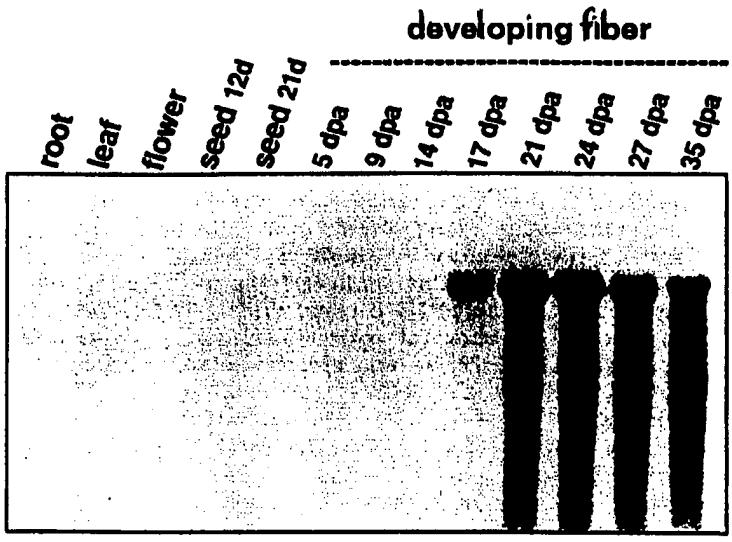
(81) Designated States: AU, BR, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PLANT CELLULOSE SYNTHASE AND PROMOTER SEQUENCES

(57) Abstract

Provided are two plant cDNA clones that are homologs of the bacterial CelA genes that encode the catalytic subunit of cellulose synthase, derived from cotton (*Gossypium hirsutum*). Also provided are genomic promoter regions to these encoding regions to cellulose synthase. Methods for using cellulose synthase in cotton fiber and wood quality modification are also provided.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

PLANT CELLULOSE SYNTHASE AND PROMOTER SEQUENCES

INTRODUCTION

Technical Field

This invention relates to plant cellulose synthase cDNA encoding sequences, and their use in modifying plant phenotypes. Methods are provided whereby the sequences can be used to control or limit the expression of endogenous cellulose synthase.

This invention also relates to methods of using *in vitro* constructed DNA transcription or expression cassettes capable of directing fiber-tissue transcription of a DNA sequence of interest in plants to produce fiber cells having an altered phenotype, and to methods of providing for or modifying various characteristics of cotton fiber. The invention is exemplified by methods of using cotton fiber promoters for altering the phenotype of cotton fiber, and cotton fibers produced by the method.

Background

In spite of much effort, no one has succeeded in isolating and characterizing the enzyme(s) responsible for synthesis of the major cell wall polymer of plants, cellulose.

Numerous efforts have been directed toward the study of synthesis of cellulose (1,4-β-D-glucan) in higher plants. However, hampered by low rates of activity *in vitro*, the cellulose synthase of plants has resisted purification and detailed characterization (for reviews, see 1,2). Aided by the discovery of cyclic-di-GMP as a specific activator, the cellulose synthase of the bacterium *Acetobacter xylinum* can be easily assayed *in vitro*, has been purified to homogeneity, and a catalytic subunit identified (for reviews, see 2,3). Furthermore, an operon of four genes involved in cellulose synthesis in *A. xylinum* has been cloned (4-7).

Characterization of these genes indicates that the first gene, termed either BcsA (7) or AcsAB (6) codes for the 83 kD

subunit of the cellulose synthase that binds the substrate UDP-glc and presumably catalyzes the polymerization of glucose residues to 1,4-β-D-glucan (8). The second gene (B) of the operon is believed to function as a regulatory subunit binding cyclic-di-GMP (9) while recent evidence suggests that the C and D genes may code for proteins that form a pore allowing secretion of the polymer and control the pattern of crystallization of the resulting microfibrils (6).

Recent studies with another gram-negative bacterium, *Agrobacterium tumefaciens*, have also led to cloning of genes involved in cellulose synthesis (10,11), although the proposed pathway of synthesis differs in some respects from that of *A. xylinum*. In *A. tumefaciens*, a CelA gene showing significant homology to the BcsA/AcsAB gene of *A. xylinum*, is proposed to transfer glc from UDP-glc to a lipid acceptor; other gene products may then build up a lipid oligosaccharide that is finally polymerized to cellulose by the action of an endo-glucanase functioning in a synthetic mode. In addition, homologs of the CelA, B, and C genes have been identified in *E. coli*, but, as this organism is not known to synthesize cellulose *in vivo*, the function of these genes is not clear (2).

These successes in bacterial systems opened the possibility that homologs of the bacterial genes might be identified in higher plants. However, experments in a number of laboratories utilizing the *A. xylinum* genes as probes for screening plant cDNA libraries have failed to identify similar plant genes. Such lack of success suggests that, if plants do contain homologs of the bacterial genes, their overall sequence homology is not very high. Recent studies analyzing the conserved motifs common to glycosyltransferases using either UDP-glc or UDP-GlcNAc as substrate suggest that there are specific conserved regions that might be expected to be found in any plant homolog of the catalytic subunit (referred to hereafter as CelA). In one of these studies, Delmer and Amor (2) identifed a motif common to many such glycosyltransferases including the bacterial CelA proteins.

An independent analysis (6) also concluded that this motif was highly conserved in a group of similar glycosyltransferases.

Extending these studies further, Saxena et al. (12) presented an elegant model for the mechanism of catalysis for enzymes such as cellulose synthase that have the unique problem of synthesizing consecutive residues that are rotated approximately rotated 180^{O} with respect to each other. The model invokes independent UDP-glc binding sites and, based upon hydrophobic cluster analysis of these enzymes, the authors concluded that 3 critical regions in all such processive glycosyltransferases each contain a conserved aspartate (D) residue, while a fourth region contained a conserved QXXRW motif. The first D residue resides in the motif as previously analyzed (2,6).

In general, genetic engineering techniques have been directed to modifying the phenotype of individual prokaryotic and eukaryotic cells, especially in culture. Plant cells have proven more intransigent than other eukaryotic cells, due not only to a lack of suitable vector systems but also as a result of the different goals involved. For many applications, it is desirable to be able to control gene expression at a particular stage in the growth of a plant or in a particular plant part. For this purpose, regulatory sequences are required which afford the desired initiation of transcription in the appropriate cell types and/or at the appropriate time in the plant's development without having serious detrimental effects on plant development and productivity. It is therefore of interest to be able to isolate sequences which can be used to provide the desired regulation of transcription in a plant cell during the growing cycle of the host plant.

One aspect of this interest is the ability to change the phenotype of particular cell types, such as differentiated epidermal cells that originate in fiber tissue, *i.e.* cotton fiber cells, so as to provide for altered or improved aspects of the mature cell type. Cotton is a plant of great commercial significance. In addition to the use of cotton fiber in the production of textiles, other uses of cotton

include food preparation with cotton seed oil and animal feed derived from cotton seed husks.

A related goal involving the control of cell wall and characteristics would be to affect valuable secondary tree characteristics of wood for paper forestry products. For instance, by altering the balance of cellulose and lignin, the quality of wood for paper production may be improved.

Finally, despite the importance of cotton as a crop, the breeding and genetic engineering of cotton fiber phenotypes has taken place at a relatively slow rate because of the absence of reliable promoters for use in selectively effecting changes in the phenotype of the fiber. In order to effect the desired phenotypic changes, transcription initiation regions capable of initiating transcription in fiber cells during development are desired. Thus, an important goal of cotton bioengineering research is the acquisition of a reliable promoter which would permit expression of a protein selectively in cotton fiber to affect such qualities as fiber strength, length, color and dyability.

Relevant Literature

Cotton fiber-specific promoters are discussed in PCT publications WO 94/12014 and WO 95/08914, and John and Crow, Proc. Natl. Acad. Sci. USA, 89:5769-5773, 1992. cDNA clones that are preferentially expressed in cotton fiber have been isolated. One of the clones isolated corresponds to mRNA and protein that are highest during the late primary cell wall and early secondary cell wall synthesis stages. John and Crow, *supra*.

In plants, control of cytoskeletal organization is poorly understood in spite of its importance for the regulation of patterns of cell division, expansion, and subsequent deposition of secondary cell wall polymers. The cotton fiber represents an excellent system for studying cytoskeletal organization. Cotton fibers are single cells in which cell elongation and secondary wall deposition can be studied as distinct events. These fibers develop synchronously within the boll following anthesis, and each fiber cell elongates for

about 3 weeks, depositing a thin primary wall (Meinert and Delmer, (1984) Plant Physiol. 59: 1088-1097; Basra and Malik, (1984) Int Rev of Cytol 89: 65-113). At the time of transition to secondary wall cellulose synthesis, the fiber cells undergo a synchronous shift in the pattern of cortical microtubule and cell wall microfibril alignments, events which may be regulated upstream by the organization of actin (Seagull, (1990) Protoplasma 159: 44-59; and (1992) In: Proceedings of the Cotton Fiber Cellulose Conference, National Cotton Council of America, Memphis RN, pp 171-192.

Agrobacterium-mediated cotton transformation is described in Umbeck, United States Patents Nos. 5,004,863 and 5,159,135 and cotton transformation by particle bombardment is reported in WO 92/15675, published September 17, 1992. Transformation of *Brassica* has been described by Radke *et al.* (Theor. Appl. Genet. (1988) 75;685-694; Plant Cell Reports (1992) 11:499-505.

Genes involved in lignin biosynthesis are described by Dwivedi, U.N., Campbell, W.H., Yu, J., Datla, R.S.S., Chiang, V.L., and Podila, G.K. (1994) ``Modification of lignin biosynthesis in transgenic Nicotiana through expression of an antisense O-methyltransferase gene from Populus'' Pl. Mol. Biol. 26: 61-71; and Tsai, C.J., Podila, G.K. and Chaing, V.L. (1995) ``Nucleotide sequence of Populus tremuloides gene for caffeic acid/5 hydroxyferulic acid O-methyltransferase'' Pl. Physiol. 107: 1459; and also U.S. Patent No. 5,451,514 (claiming the use of cinnamyl alcohol dehydrogenase gene in an antisense orientation such that the endogenous plant cinnamyl alcohol dehydrogenase gene is inhibited).

Other References Cited Throughout the Specification

1. Gibeaut, D.M., & Carpita, N.C. (1994) FASEB J. 8, 904-915.
2. Delmer, D.P., & Amor, Y. (1995) Plant Cell 7, 987-1000.
3. Ross, P., Mayer, R., & Benziman, M. (1991) Microbiol. Rev. 55, 35-58.
4. Saxena, I.M., Lin, F.C., & Brown, R.M., Jr. (1990) Plant Mol. Biol. 15, 673-683.

5. Saxena, I.M., Lin, F.C., & Brown, R.M., Jr. (1992) Plant Mol. Biol. 16, 947-954.

6. Saxena, I.M., Kudlicka, K., Okuda, K., & Brown, R.M., Jr. (1994) J. Bacteriol. 176, 5735-5752.

7. Wong, H.C., Fear, A.L., Calhoon,, R.D., Eidhinger, G.H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D.H., Meade, J.H., Emerick, A.W., Bruner, R., Ben-Basat, B.A., & Tal, R. (1990) Proc. Natl. Acad. Sci. USA 87, 8130-8134.

8. Lin, F.-C., Brown, R.M.Jr., Drake, R.R.Jr., & Haley, B.E. (1990) J. Biol. Chem. 265, 4782- 4784.

9. Mayer, R., Ross, P., Winhouse, H., Amikam, D., Volman, G., Ohana, P., Calhoon, R.D., Wong, H.C., Emerick, A.W., & Benziman, M. (1991) Proc. Natl. Acad. Sci. USA 88, 5472-5476.

10. Matthysse, A.G., White, S., & Lightfoot, R. (1995a) J. Bacteriol. 177, 1069-1075.

11. Matthysse, A.G., Thomas, D.O.L., & White, S. (1995b) J. Bacteriol. 177, 1076-1081.

12. Saxena, I.M., Brown, R.M.,Jr., Fevre, M., Geremia, R.A., & Henrissat, B. (1995) J. Bacteriol. 177, 1419-1424.

13. Meinert, M., & Delmer, D.P. (1977) Plant Physiol. 59, 1088-1097.

14. Delmer, D.P., Pear, J.R., Andrawis, A., & Stalker, D.M. (1995) Mol. Gen. Genet. 248, 43-51.

15. Delmer, D.P., Solomon, M., & Read, S.M. (1991) Plant Physiol. 95, 556-563.

16. Nagai, K., & Thogersen, H.C. (1987) Methods in Enzymol. 153, 461-481.

17. Laemmli, U.K. (1970) Nature 227, 680-685.

18. Kyte, J., & Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.

19. Oikonomakos, N.G., Acharya, K.R., Stuart, D.I., Melpidou, A.E., McLaughlin, P.J., & Johnson, L.N. (1988) Eur. J. Biochem. 173, 569-578.

20. Maltby, D., Carpita, N.C., Montezinos, D., Kulow, C., & Delmer, D.P. (1979) Plant Physiol. 63, 1158-1164.

21. Inoue, S.B., Takewaki, N., Takasuka, T., Mio, T., Adachi, M., Fujii, Y., Miyamoto, C., Arisawa, M., Furuichi, Y., & Watanabe, T. (1995) Eur. J. Biochem. 231, 845-854.

22. Jacob, S.R., & Northcote, D.H. (1985) J. Cell Sci. 2 (suppl.), 1-11.

23. Delmer, D.P. (1987) Annu. Rev. Plant Physiol. 38, 259-290.

24. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990) J. Mol. Biol. 215, 403-410

25. Milligan, G., Parenti, M., & Magee, A.I. (1995) TIBS 20, 183-186.

26. Amor, Y., Haigler, C.H., Johnson, S., Wainscott, M., & Delmer, D.P. (1995) Proc. Natl. Acad. Sci. USA 92, 9353-9357.

27. Amor, Y., Mayer, R., Benziman, M., & Delmer, D.P. (1991) Plant Cell 3, 989-995.

SUMMARY OF THE INVENTION

Two cotton genes, CelA1 and CelA2, have been shown to be highly expressed in developing fibers at the onset of secondary wall cellulose synthesis. Comparisons indicate that these genes and the rice CelA gene encode polypeptides that have three regions of reasonably high homology, both in terms of primary amino acid sequence and hydropathy, with bacterial CelA proteins. The fact that these homologous stretches are in the same sequential order as in the bacterial CelA proteins and also contain four sub-regions previously predicted to be critical for substrate binding and catalysis (12) argues that the plant genes encode true homologs of bacterial CelA proteins. Furthermore, the pattern of expression in fiber as well as our demonstration that at least one of these highly-conserved regions is critical for UDP-glc binding also supports this conclusion.

Novel DNA promoter sequences are also supplied, and methods for their use are described for directing transcription of a gene of interest in cotton fiber.

The developing cotton fiber is an excellent system for studies on cellulose synthesis as these single cells develop synchronously in the boll and, at the end of elongation, initiate the synthesis of a nearly pure cellulosic cell wall. During this transition period, synthesis of other cell wall polymers ceases and the rate of cellulose synthesis is estimated to rise nearly 100-fold *in vivo* (13). In our

continuing efforts to identify genes critical to this phase of fiber development, we have initiated a program sequencing randomly selected cDNA clones derived from a library prepared from mRNA harvested from fibers at the stage in which secondary wall synthesis approaches its maximum rate (approximately 21 dpa).

We have characterized two cotton (*Gossypium hirsutum*) cDNA clones and identified one rice (*Oryza sativa*) cDNA that are homologs of the bacterial CelA genes that encode the catalytic subunit of cellulose synthase. Three regions in the deduced amino acid sequences of the plant CelA gene products are conserved with respect to the proteins encoded by bacterial CelA genes. Within these conserved regions are four highly conserved subdomains previously suggested to be critical for catalysis and/or binding of the substrate UDP-glc. An overexpressed DNA segment of the cotton CelA1 gene encodes a polypeptide fragment that spans these domains and effectively binds UDP-glc, while a similar fragment having one of these domains deleted does not. The plant CelA genes show little homology at the amino and carboxy terminal regions and also contain two internal insertions of sequence, one conserved and one hypervariable, that are not found in the bacterial gene sequences. Cotton CelA1 and CelA2 genes are expressed at high levels during active secondary wall cellulose synthesis in the developing fiber. Genomic Southern analyses in cotton demonstrate that CelA comprises a family of approximately four distinct genes.

We report here the discovery of two cotton genes that show highly-enhanced expression at the time of onset of secondary wall synthesis in the fiber. The sequences of these two cDNA clones, termed CelA1 and CelA2, while not identical, are highly homologous to each other and to a sequenced rice EST clone discovered in the dBEST databank. The deduced proteins also share signifigant regions of homology with the bacterial CelA proteins. Coupled with their high level and specificity of expression in fiber at the time of active cellulose synthesis, as well as the ability of an *E. coli* expressed fragment of the CelA1 gene product to bind UDP-glc,

these findings support the conclusion that these plant genes are true homologs of the bacterial CelA genes.

The methods of the present invention include transfecting a host plant cell of interest with a transcription or expression cassette comprising a cotton fiber promoter and generating a plant which is grown to produce fiber having the desired phenotype. Constructs and methods of the subject invention thus find use in modulation of endogenous fiber products, as well as production of exogenous products and in modifying the phenotype of fiber and fiber products. The constructs also find use as molecular probes. In particular, constructs and methods for use in gene expression in cotton embryo tissues are considered herein. By these methods, novel cotton plants and cotton plant parts, such as modified cotton fibers, may be obtained.

The sequences and constructs of this invention may also be used to isolate related cellulose synthase genes from forest tree species, for use in transforming and modifying wood quality. As and example, lignin, an undesirable by-product of the pulping process, by be reduced by over-expressing the cellulose synthase product and diverting production into cellulose.

Thus, the application provides constructs and methods of use relating to modification of cell and cell wall phenotype in cotton fiber and wood products.

DESCRIPTION OF THE DRAWINGS

Figure 1. Northern analysis of CelA1 gene in cotton tissues and developing fiber. Approximately 10μg total RNA from each tissue was loaded per lane. Blots were prepared and probe preparation and hybridization conditions were performed as described previously (14). The entire CelA1 cDNA insert was used as a probe in this experiment. Exposure time for the audoradiogram was seven hours at -70°.

Figure 2. Cotton genomic DNA analysis for both the CelA1 and CelA2 cDNAs. Approximately 10-12μg of DNA was digested with the designated restriction enzymes and electrophoresed 0.9% agarose gels. Probe preparation and hybridization conditions were as described previously (14).

The entire CelA1 and CelA2 cDNAs were utlized as probes. Exposure time for the audoradiograms was three days at -70°.

Figure 3. Multiple alignment of deduced amino acid sequences of plant and bacterial CelA proteins. Analyses were performed by Clustal Analysis using the Lasergene Multalign program (DNAStar, Madison, WI) with gap and gap-length penalties of 10 and a PAM250 weight table. Residues are boxed and shaded when they show chemical group similarity in 4 out of 7 proteins compared. H-1, H-2, H-3 regions are indicated where homology between plant and bacterial proteins is highest. The plant proteins show two insertions that are not present in the bacterial protein--one, P-CR, is conserved among the plant CelA genes, while a second insertion is hypervariable (HVR) between plant genes. The presence of the P-CR and HVR regions led to inaccurate alignments when the entire proteins were compared; the optimal alignments shown here were thus performed in five seperate blocks. Regions U-1 through U-4 are predicted to be critical for UDP-glc binding and catalysis in bacterial CelA proteins; the predicted critical D residues and QXXRW motif are boxed and starred respectively. Potential sites of N-glycosylation are indicate by -G-.

Figure 4. Kyte-Doolittle hydropathy plots of cotton CelA1 aligned with those of two bacterial CelA proteins. Alignments and designations are based upon those noted in Fig. 2. The hydropathy profiles shown were calculated using a window of 7, although a window of 19 was used for predictions of transmembrane helices that are indicated by the arrows.

Figure 5. An *E. coli* expressed GST cotton CelA-1 fusion protein binds the containing U1 through U4 binds UDP-glc *in vitro*. Panel A shows a hypothetical orientation of the cotton CelA1 protein in the plasma membrane and indicates the cytoplasmic region containing the sub-domains U-1 to U-4. GST-fusion constructs for CelA1 fragments spanning the region between the potential transmembrane helices (A through H) were prepared as described in Materials and Methods. The purified and blotted CelA1 fusion protein fragments were tested as described in Materials and Methods for their ability to bind

^{32}p-UDP-glc (panel B). M refers to the molecular weight markers while CS and ÆU1 to the full-length and deleted GST-CelA1 fusion polypeptides. The left panel shows proteins stained with Coomassie blue while the other three panels show representative autoradiograms under different binding conditions as described in Materials and Methods. Ph, BSA and Ova refer to the molecular weight standards phosphorylase b, bovine serum albumin and ovalbumin respectively.

Figure 6. Nucleic acid sequences to cDNA of CelA1 protein of cotton (*Gossypium hirsutum*).

Figure 7. Nucleic acid sequences to cDNA of CelA2 protein of cotton (*Gossypium hirsutum*), including approximately the last 3' two-thirds of the encoding region.

Figure 8. Genomic nucleic acid sequences of CelA1 protein of cotton (*Gossypium hirsutum*), including approximately 900 bases of the promoter region 5' to the encoding sequences.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, novel constructs and methods are described, which may be used provide for transcription of a nucleotide sequence of interest in cells of a plant host, preferentially in cotton fiber cells to produce cotton fiber having an altered color phenotype.

Cotton fiber is a differentiated single epidermal cell of the outer integument of the ovule. It has four distinct growth phases; initiation, elongation (primary cell wall synthesis), secondary cell wall synthesis, and maturation. Initiation of fiber development appears to be triggered by hormones. The primary cell wall is laid down during the elongation phase, lasting up to 25 days postanthesis (DPA). Synthesis of the secondary wall commences prior to the cessation of the elongation phase and continues to approximately 40 DPA, forming a wall of almost pure cellulose.

The constructs for use in such cells may include several forms, depending upon the intended use of the construct. Thus, the constructs include vectors, transcriptional cassettes, expression cassettes and plasmids. The

transcriptional and translational initiation region (also sometimes referred to as a "promoter,"), preferably comprises a transcriptional initiation regulatory region and a translational initiation regulatory region of untranslated 5' sequences, "ribosome binding sites," responsible for binding mRNA to ribosomes and translational initiation. It is preferred that all of the transcriptional and translational functional elements of the initiation control region are derived from or obtainable from the same gene. In some embodiments, the promoter will be modified by the addition of sequences, such as enhancers, or deletions of nonessential and/or undesired sequences. By "obtainable" is intended a promoter having a DNA sequence sufficiently similar to that of a native promoter to provide for the desired specificity of transcription of a DNA sequence of interest. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences.

Cotton fiber transcriptional initiation regions of cellulose synthase are used in cotton fiber modification.

A transcriptional cassette for transcription of a nucleotide sequence of interest in cotton fiber will include in the direction of transcription, the cotton fiber transcriptional initiation region, a DNA sequence of interest, and a transcriptional termination region functional in the plant cell. When the cassette provides for the transcription and translation of a DNA sequence of interest it is considered an expression cassette. One or more introns may be also be present.

Other sequences may also be present, including those encoding transit peptides and secretory leader sequences as desired.

Downstream from, and under the regulatory control of, the cellulose synthase transcriptional/translational initiation control region is a nucleotide sequence of interest which provides for modification of the phenotype of fiber. The nucleotide sequence may be any open reading frame encoding a polypeptide of interest, for example, an enzyme, or a sequence complementary to a genomic sequence, where the genomic sequence may be an open reading frame, an intron, a noncoding

leader sequence, or any other sequence where the complementary sequence inhibits transcription, messenger RNA processing, for example, splicing, or translation. The nucleotide sequences of this invention may be synthetic, naturally derived, or combinations thereof. Depending upon the nature of the DNA sequence of interest, it may be desirable to synthesize the sequence with plant preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest. Phenotypic modification can be achieved by modulating production either of an endogenous transcription or translation product, for example as to the amount, relative distribution, or the like, or an exogenous transcription or translation product, for example to provide for a novel function or products in a transgenic host cell or tissue. Of particular interest are DNA sequences encoding expression products associated with the development of plant fiber, including genes involved in metabolism of cytokinins, auxins, ethylene, abscissic acid, and the like. Methods and compositions for modulating cytokinin expression are described in United States Patent No. 5,177,307, which disclosure is hereby incorporated by reference. Alternatively, various genes, from sources including other eukaryotic or prokaryotic cells, including bacteria, such as those from *Agrobacterium tumefaciens* T-DNA auxin and cytokinin biosynthetic gene products, for example, and mammals, for example interferons, may be used.

Alternatively, the present invention provides the sequences to cotton cellulose synthase, which can be expressed, or down regulated by antisense or co-suppression with its own, or other cotton or other fiber promoters to modify fiber phenotyp.

In cotton, primary wall hemicellulose synthesis ceases as secondary wall synthesis initiates in the fiber, and there are only two possible β-glucans synthesized in fibers at the time these genes are highly-expressed; callose and cellulose (20). The following data strongly argue against the plant CelA genes coding for callose synthase: 1) callose synthase binds UDP-glc

and is activated in a Ca^{2+}-dependent manner (2), while the CelA1 polypeptide fragment containing the UDP-glc binding site preferentially binds UDP-glc in a Mg^{2+}-dependent manner, similar to bacterial cellulose synthase (9); 2) the timing of synthesis of callose *in vivo* in developing cotton fiber (20) does not match the expression of the cotton CelA genes (Fig. 1); 3) comparison of the CelA gene sequences with those of suspected 1,3-β-glucan synthase genes from yeast (21) indicated no significant homology.

It is still possibille that the CelA protein might encode both activities, as hypothesized some years ago (22-23), and the plant CelAs might be responsible for direct polymerization of glucan from UDP-glc as proposed for *A. xylinum*, although they may catalyze synthesis of a lipid-glc precursor as proposed for the CelA protein of *A. tumefaciens*.

In addition to their similarities, the plant CelA genes show several very interesting divergences from their bacterial ancestors, and these may account for the previous lack of success in using bacterial probes to detect these cDNA clones. However, a BLAST search of protein data banks (24) using the entire protein sequence of cotton CelA1 always shows highest homology with the bacterial cellulose synthases. Of particular interest is the insertion of two unique, plant-specific regions designated P-CR and HVR. These regions are clearly not artifacts of cloning as they are observed in both cotton genes as well as the rice CelA gene. The three plant proteins show a high degree of amino acid homology to each other throughout most of their length, diverging only at the N- and C-terminal ends and the very interesting HVR region. It is tempting to speculate that the HVR region may confer some specificity of function; the highly-charged and cysteine rich nature of the first portion of HVR could make this region a potential candidate for interaction with specific regulatory proteins, for cytoskeletal elements, or for redox regulation. In addition, we note the presence of several cysteine residues near the N- and C-terminal regions of the protein that might serve as

substrates for palmytolylation and also serve to help anchor the protein in the membrane (25).

In summary, the finding of these plant CelA homologs potentially opens up an exciting chapter in research on cellulose synthesis in higher plants. Their finding is of particular significance since biochemical approaches to identification of plant cellulose synthase have proven exceedingly difficult. One obvious challenge will be to gain definitive proof that these genes are truely functional in cellulose synthesis*in vivo*. Other promising goals will be to identify other components of a complex that might interact with CelA, such as that proposed for sucrose synthase (26), and/or a regulatory subunit that binds cyclic-di-GMP (9,27) or other glycosyltransferases (10,11).

Transcriptional cassettes may be used when the transcription of an anti-sense sequence is desired. When the expression of a polypeptide is desired, expression cassettes providing for transcription and translation of the DNA sequence of interest will be used. Various changes are of interest; these changes may include modulation (increase or decrease) of formation of particular saccharides, hormones, enzymes, or other biological parameters. These also include modifying the composition of the final fiber that is changing the ratio and/or amounts of water, solids, fiber or sugars. Other phenotypic properties of interest for modification include response to stress, organisms, herbicides, brushing, growth regulators, and the like. These results can be achieved by providing for reduction of expression of one or more endogenous products, particularly an enzyme or cofactor, either by producing a transcription product which is complementary (anti-sense) to the transcription product of a native gene, so as to inhibit the maturation and/or expression of the transcription product, or by providing for expression of a gene, either endogenous or exogenous, to be associated with the development of a plant fiber.

The termination region which is employed in the expression cassette will be primarily one of convenience, since the termination regions appear to be relatively

interchangeable. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, may be derived from another source. The termination region may be naturally occurring, or wholly or partially synthetic. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. In some embodiments, it may be desired to use the 3' termination region native to the cotton fiber transcription initiation region used in a particular construct.

As described herein, in some instances additional nucleotide sequences will be present in the constructs to provide for targeting of a particular gene product to specific cellular locations.

Similarly, other constitutive promoters may also be useful in certain applications, for example the mas, Mac or DoubleMac, promoters described in United States Patent No. 5,106,739 and by Comai *et al., Plant Mol. Biol.* (1990) *15:373-381)*. When plants comprising multiple gene constructs are desired, the plants may be obtained by co-transformation with both constructs, or by transformation with individual constructs followed by plant breeding methods to obtain plants expressing both of the desired genes.

A variety of techniques are available and known to those skilled in the art for introduction of constructs into a plant cell host. These techniques include transfection with DNA employing *A. tumefaciens* or *A. rhizogenes* as the transfecting agent, protoplast fusion, injection, electroporation, particle acceleration, etc. For transformation with *Agrobacterium*, plasmids can be prepared in *E. coli* which contain DNA homologous with the Ti-plasmid, particularly T-DNA. The plasmid may or may not be capable of replication in *Agrobacterium*, that is, it may or may not have a broad spectrum prokaryotic replication system such as does, for example, pRK290, depending in part upon whether the transcription cassette is to be integrated into the Ti-plasmid or to be retained on an independent plasmid. The *Agrobacterium* host will contain a plasmid having the *vir* genes

necessary for transfer of the T-DNA to the plant cell and may or may not have the complete T-DNA. At least the right border and frequently both the right and left borders of the T-DNA of the Ti- or Ri-plasmids will be joined as flanking regions to the transcription construct. The use of T-DNA for transformation of plant cells has received extensive study and is amply described in EPA Serial No. 120,516, Hoekema, In: The Binary Plant Vector System Offset-drukkerij Kanters B.V., Alblasserdam, 1985, Chapter V, Knauf, *et al.*, Genetic Analysis of Host Range Expression by *Agrobacterium,* In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, NY, 1983, p. 245, and An, *et al.*, *EMBO J.* (1985) 4:277-284.

For infection, particle acceleration and electroporation, a disarmed Ti-plasmid lacking particularly the tumor genes found in the T-DNA region) may be introduced into the plant cell. By means of a helper plasmid, the construct may be transferred to the *A. tumefaciens* and the resulting transfected organism used for transfecting a plant cell; explants may be cultivated with transformed *A. tumefaciens* or *A. rhizogenes* to allow for transfer of the transcription cassette to the plant cells. Alternatively, to enhance integration into the plant genome, terminal repeats of transposons may be used as borders in conjunction with a transposase. In this situation, expression of the transposase should be inducible, so that once the transcription construct is integrated into the genome, it should be relatively stably integrated. Transgenic plant cells are then placed in an appropriate selective medium for selection of transgenic cells which are then grown to callus, shoots grown and plantlets generated from the shoot by growing in rooting medium.

To confirm the presence of the transgenes in transgenic cells and plants, a Southern blot analysis can be performed using methods known to those skilled in the art. Expression products of the transgenes can be detected in any of a variety of ways, depending upon the nature of the product, and include immune assay, enzyme assay or visual inspection, for example to detect pigment formation in the appropriate plant part or

cells. Once transgenic plants have been obtained, they may be grown to produce fiber having the desired phenotype. The fibers may be harvested, and/or the seed collected. The seed may serve as a source for growing additional plants having the desired characteristics. The terms transgenic plants and transgenic cells include plants and cells derived from either transgenic plants or transgenic cells.

The various sequences provided herein may be used as molecular probes for the isolation of other sequences which may be useful in the present invention, for example, to obtain related transcriptional initiation regions from the same or different plant sources. Related transcriptional initiation regions obtainable from the sequences provided in this invention will show at least about 60% homology, and more preferred regions will demonstrate an even greater percentage of homology with the probes.

Of particular importance is the ability to obtain related transcription initiation control regions having the timing and tissue parameters described herein. Thus, by employing the techniques described in this application, and other techniques known in the art (such as Maniatis, *et al.*, *Molecular Cloning,- A Laboratory Manual* (Cold Spring Harbor, New York) 1982), other encoding regions or transcription initiation regions of cellulose synthase as described in this invention may be determined. The constructs can also be used in conjunction with plant regeneration systems to obtain plant cells and plants; thus, the constructs may be used to modify the phenotype of fiber cells, to provide cotton fibers which are colored as the result of genetic engineering to heretofor unavailable hues and/or intensities.

Various varieties and lines of cotton may find use in the described methods. Cultivated cotton species include *Gossypium hirsutum* and *G. babadense* (extra-long stable, or Pima cotton), which evolved in the New World, and the Old World crops G. *herbaceum* and *G. arboreum*.

By using encoding sequences to enzymes which control wood quality and wood product characteristics, i.e., cellulose synthase and O-methyltransferase (a key enzyme in lignin

biosynthesis) the relative synthesis of cellulose and lignin by plants may be controlled. Transformation of the plant genome with a recombinant gene construct which contains the gene specifying an enzyme critical to the synthesis of cellulose or lignin or a lignin precursor, in either a sense or in an antisense orientation. If an antisense orientation, the gene will transcribed so mRNA having a sequence complementary to the equivalent mRNA transcribed from the endogenous gene is expressed, leading to suppression of the synthesis of lignin or cellulose.

If the recombinant gene has the lignin enzyme gene in normal, or "sense" orientation, increased production of the enzyme may occur when the insert is the full length DNA but suppression may occur if only a partial sequence is employed.

Furthermore, the expression of one may be increased in this manner while the other is reduced. For instance, the production of cellulose may by increased through the overexpression of cellulose synthase, while lignin production is reduced. By thus reducing the relative lignin content, the quality of wood for paper production would be improved.

EXAMPLES

The following examples are offered by way of illustration and not by limitation.

Example 1

cDNA libraries

An unamplified cDNA library was used to prepare the Lambda Uni-Zap vector (Stratagene, LaJolla, CA) using cDNA derived from polyA+ mRNA prepared from fibers of *Gossypium hirsutum* Acala SJ-2 harvested at 21 DPA, the time at which secondary wall cellulose synthesis is approaching a maximal rate (13). Approximately 250 plaques were randomly selected from the cDNA library, phages purified and plasmids excised from the phage vector and transformed.

The resulting clones/inserts were size screened on 0.8% agarose gels (DNA inserts below 600bp were excluded).

Example 2

Isolation and Sequencing of cDNA Clones

Plasmid DNA inserts were randomly sequenced using an Applied Biosystems (Foster City, CA) Model 373A DNA sequencer. A search of the GenBank EST databank revealed that there were at least 23 rice and 8 *Arabidopsis* EST clones that contain sequences similar to the cotton CelA1 DNA sequence. EST clone S14965 was obtained from Y. Nagamura (Rice Genome Research Program, Tsukuba). A series of deletion mutants were generated and used for DNA sequencing analysis at the Weizmann Institute of Science (Rehovot).

Example 3

Northern and Southern Analyses.

Cotton plants (*G. hirsutum* cv. Coker 130) were grown in the greenhouse and tissues harvested at the appropriate times indicated and frozen in liquid N_2. Total cotton RNA and cotton genomic DNA was prepared and subjected to Northern and Southern analyses as described previously (14).

Example 4

UDP-Glc Binding Studies

To construct a GST-CelA1 protein fusion, a 1.6kb DNA CelA1 DNA fragment containing a putative cytoplasmic domain between the second and third transmembrane helices was PCR amplified with the primers ATTGAATTCCTGGGTGTTGGATCAGTT and ATTCTCGAGTGGAAGGGATTGAAA in a reaction containing 1 ng plasmid DNA (clone 213) as template. The amplified fragment was unidirectionally cloned into the *Eco*RI and *Xho*I sites of the GST expression vector pGEX4T-3 (Pharmacia), generating a fusion protein GST-CS containing the amino acids Ser215 to Leu759 of the cotton CelA1 protein. Two CelA1 gene internal *Pst*I sites within the plasmid pGST-CS were used to generate the deletion mutant pGST-CSΔU1, which lacks 196 amino acids (and the U1 binding region) from Val252 to Ala447.

For the UDGP binding assays, α-^{32}P-labeled UDP-glc was prepared as described (15). The two fusion proteins GST-CS and GST-CSÆU1 were expressed in *E. coli* and purified from inclusion bodies (16). Proteins were suspended in sample

buffer, heated to 100_ C for 5 min and approximately 50ng of the two fusion protein products and molecular weight standards (Bio-Rad) subjected to SDS-PAGE using 4.5% and 7.5% acrylamide in the stacking and separating gels, respectively (17). After electrophoresis, protein transfer to nitrocellulose filters was carried out in transfer buffer (25mM Tris, 192mM glycine and 20% (v/v) methanol). The filter was briefly rinsed in deionized H_2O and incubated in PBS buffer for 15 min, then stained with Ponceau-S in PBS buffer. After washing in deionized H_2O, protein was further renatured on the filter by incubation in PBS buffer for 30 min and used directly for binding assays. All binding buffers contained 50mM HEPES/KOH (pH 7.3), 50mM NaCl and 1mMDTT. In addition, binding buffers contained either 5mM $MgCl_2$ and 5mM EGTA (Buffer Mg/EGTA), 5mM EDTA (Buffer EDTA) or 1mM $CaCl_2$ and 20mM cellobiose (Buffer Ca/CB). Binding reaction was carried out in 7ml containing ^{32}P-labeled UDP-glc (1x 10^7 cpm) at room temperature for 3 hours with constant shaking. Filters were washed separately three times in 20ml washing buffer consisting of 50mM HEPES/KOH (pH 7.3) and 50mM NaCl for 5min each, briefly dried and analyzed on a Bio-imaging analyzer BAS1000 (Fugi).

Example 5

Identification, Differential Expression and Genomic Analysis of Cotton CelA Genes

During the course of screening and sequencing random cDNA clones from a cotton fiber specific cDNA library prepared from RNA collected approximately 21 dpa, it was discovered that two cDNA clones that initially exhibited small blocks of amino acid homology to the proteins encoded by the bacterial CelA genes. Clone 213 appeared to be full-length cDNA while another distinct clone, 207, appeared to be a partial clone relative to the length of 213. These two clones were partially homologous at the nucleotide and amino acid levels and designated CelA1 and CelA2 respectively.

These clones were then utilized as probes for Northern blot analysis to determine their differential expression in cotton tissues and developing cotton fiber. Figure 1

indicates the expression pattern for the CelA1 gene. The CelA1 gene encodes a mRNA of approximately 3.2kb in length and is expressed at extremely high levels in developing fiber, beginning at approximately 17 dpa, the time at which secondary wall cellulose synthesis is initiated(13). The gene is also expressed at low levels in all other cotton tissues, most notably in root, flower and developing seeds. Since regions of these genes are somewhat homologous at the nucleotide level, gene specific probes were designed (using the hypervariable regions described in Fig 3) to distinguish the specific expression patterns of CelA1 and CelA2. These gene specific probes generated expression patterns (data not shown) for the two genes identical to that shown in Figure 1, except that a very low mRNA level was also detected in the primary wall phase of fiber development (5-14dpa) for the CelA2 gene when the blots were overexposed. The CelA2 gene specific probe also encoded a 3.2kb mRNA, analogous in size to the mRNA specified by the gene for CelA1. Messenger RNAs for both genes exhibit a characteristic degradation pattern similar to other mRNAs specifically expressed late in fiber development (J. Pear, unpublished observations) and this degradation is not a result of the integrity of the mRNA preparations (14). We estimate that both cotton CelA genes are expressed in developing fiber approximately 500 times their level of expression in other cotton tissues and that they constitute approximately 1-2% of the 24dpa fiber mRNA.

In order to estimate the number of CelA genes in the cotton genome, Southern analysis was performed utilizing both CelA cDNAs independently as probes (Fig 2). Although the two cotton genes are fairly non-homologous at the nucleotide level over their entire length, there are regions of homology (the H1, H2 and H3 regions described below) and it was thought these regions could be useful in identifying other cotton CelA genes. Figure 2 indicates that the CelA1 cDNA probe will hybridize, albeit weakly, to the CelA2 genomic equivalent and vise versa. The *Hin*dIII pattern for both genes and cDNA probes is particularly discriminating. There are also a number of other weakly hybridzing bands in these digests and

from these data we estimate that the cotton CelA genes constitute a small family of approximately four genes.

Homology of Plant and Bacterial CelA Gene Products.

In addition to the two similar cotton CelA genes, a homologous cDNA clone was discovered in the dBest databank* of rice and Arabidopsis ESTs. Accession No. D48636, the rice clone having the longest insert was obtained and sequenced, and the homology comparisons with bacterial proteins reported here also include results with the rice CelA. Figure 3 shows the results of a multiple alignment of the deduced amino acid sequences from the three plant CelA genes and four bacterial CelA genes from *A. xylinum* (AcsAB and BcsA), *E. coli* , and *A. tumefaciens*. Figure 4 shows hydropathy plots (18) of cotton CelA1 similarly aligned with two bacterial CelA proteins and serves as a more general summary of the overall homologies.

Of the plant genes, only the cotton CelA1 appears to be a full-length clone of 3.2kb exhibiting an open reading frame that could potentially code for a polypeptide of 109,586 kD, a pI of 6.4, and four potential sites of N-glycosylation. Comparison of the N-terminal region of cotton CelA1 with bacterial genes indicates that the plant protein has an extended N-terminal similar in length and hydropathy profile, but with only poor amino acid sequence homology to the *A. tumefaciens* CelA protein. In general, sequence homology of plant and bacterial genes in both the N-terminal and C-terminal regions is poor. However, although overall similarity comparing plant to bacterial proteins is less than 25%, three homologous regions were identified, called H-1, H-2, and H-3, where the sequence similarity rises to 50-60% at the amino acid level. Interspersed between these regions of homology are two plant-specific regions not found at all in the bacterial proteins. Sequences in the first of these

* The following accession numbers were identified as showing homology with cotton CelA-1. For rice: D48636, D41261, D40691, D46824, D47622, D47175, D41766, D41986, D24655, D23732, D24375, D47732, D47821, D47850, D47494, D24964, D24862, D24860, D24711, D23841, D48053, D48612, D40673; for *Arabidopsis*: T45303, T45414, H76149, H36985, Z30729, H36425, T45311, A35212.

insertions are highly conserved in the plant genes (P-CR), while the second interspersed region seems to be a hypervariable regions (HVR) for there is considerable sequence divergence among the plant proteins analyzed.

None of the plant or bacterial CelA proteins contains obvious signal sequences even though they are presumably transmembrane proteins (4). However, the overall profiles suggest two potential transmembrane helices in the N-terminal and six in the C-terminal region of the cotton CelA1 that could anchor the protein in the membrane (see arrows Fig.3 and also panel A of Fig.5). The amino acid sequence positions for these predicted transmembrane helices are: A (169-187), B (200-218), C (759-777), D (783-801), E (819-837), F (870-888), G (903-921), H (933-951). The central portions of the proteins are more hydrophilic and are predicted to reside in the cytoplasm and contain the site(s) of catalysis. More detailed inspection of these hydrophilic stretches reveals four particularly conserved sub-regions (marked U-1 through U-4 on Figs. 3-4) that contain the conserved asp (D) residues (in U-1-3) and the motif QXXRW (in U-4) that have been proposed (12) to be involved in substrate binding and/or catalysis .

Binding of UDP-glucose. Further evidence that the proteins encoded by these plant genes are CelA homologs comes from our demonstration that a DNA segment encoding the central region of the cotton CelA1 protein, over-expressed in *E. coli*, binds UDP-glc. We subcloned a 1.6 kb fragment of the cotton CelA1 clone to create a hybrid gene that encodes GST fused to the CelA1 sequence encoding amino acid residues 215-759 of the CelA1 protein (Fig. 5a). This region spans U-1 through U-4 that are suspected to be critical for UDP-glc binding. As a control, another GST fusion was created using a 1.0 kb *Pst*I fragment that had the U-1 region deleted and might not be predicted to bind UDP-glc. The fusion proteins were overexpressed in *E. coli* , purifed, and shown to have the predicted sizes of approximately 87 and 64 kD, respectively (Fig. 5b). The purified proteins were then subjected to SDS-PAGE, and blotted to nitrocellulose. Blotted proteins

were renatured, and incubated with ^{32}P-UDP-glc in order to test for binding (Fig. 5b). As predicted, the 87 kD GST-CelA1 fusion does indeed bind UDP-glc in a Mg^{2+} dependent manner, while the shorter fusion with the U-1 domain deleted did not show any binding (Although not observed in the experiment shown, in some experiments very weak labeling in the presence of Ca^{2+} could be observed). As further controls, note that the molecular weight standards BSA and ovalbumin, proteins lacking UDP-glc binding sites, show no interaction with UDP-glc, while phosphorylase b, an enzyme inhibited by UDP-glc (19), binds this substrate.

Figure 6 provides the encoding sequence to the cDNA to celA1 (start ATG at ~ base 179), while Figure 7 provides the encoding sequence to the approximately two-thirds 3' of the cDNA to celA2.

Example 6

Genomic DNA

cDNA for the cellulose synthase clones was used to probe for genomic clones. For both, full length genomic DNA was obtained from a library made using the lambda dash 2 vector from Stratagene®, which was used to construct a genomic DNA library from cotton variety Coker 130 (*Gossypium hirsutum* cv. coker 130), using DNA obtained from germinating seedlings.

The cotton genomic library was probed with a cellulose synthase probe and genomic phage candidates were identified and purified. Figure 8 provides an approximately 1 kb sequence of the cellulose synthase promoter region which is immediately 5' to the celA1 encoding region. The start of the cellulose synthase enzyme encoding region is at the ATG at base number 954.

Example 7

Cotton Transformation

Explant Preparation

Promoter constructs comprising the cellulose synthase promoter sequences of celA1 can be cotton prepared. Coker 315 seeds are surface disinfected by placing in 50% Clorox (2.5%

sodium hypochlorite solution) for 20 minutes and rinsing 3 times in sterile distilled water. Following surface sterilization, seeds are germinated in 25 x 150 sterile tubes containing 25 mls 1/2 x MS salts: 1/2 x B5 vitamins: 1.5% glucose: 0.3% gelrite. Seedlings are germinated in the dark at 28°C for 7 days. On the seventh day seedlings are placed in the light at 28±2°C.

Cocultivation and Plant Regeneration

Single colonies of *A. tumefaciens* strain 2760 containing binary plasmids pCGN2917 and pCGN2926 are transferred to 5 ml of MG/L broth and grown overnight at 30°C. Bacteria cultures are diluted to 1 x 10^8 cells/ml with MG/L just prior to cocultivation. Hypocotyls are excised from eight day old seedlings, cut into 0.5-0.7 cm sections and placed onto tobacco feeder plates (Horsch *et al.* 1985). Feeder plates are prepared one day before use by plating 1.0 ml tobacco suspension culture onto a petri plate containing Callus Initiation Medium CIM without antibiotics (MS salts: B5 vitamins: 3 % glucose: 0.1 mg/L 2,4-D: 0.1 mg/L kinetin: 0.3% gelrite, pH adjusted to 5.8 prior to autoclaving). A sterile filter paper disc (Whatman #1) was placed on top of the feeder cells prior to use. After all sections are prepared, each section was dipped into an *A. tumefaciens* culture, blotted on sterile paper towels and returned to the tobacco feeder plates.

Following two days of cocultivation on the feeder plates, hypocotyl sections are placed on fresh Callus Initiation Medium containing 75 mg/L kanamycin and 500 mg/L carbenicillin. Tissue is incubated at 28±2°C, 30uE 16:8 light:dark period for 4 weeks. At four weeks the entire explant is transferred to fresh callus initiation medium containing antibiotics. After two weeks on the second pass, the callus is removed from the explants and split between Callus Initiation Medium and Regeneration Medium (MS salts: 40mM KNO_3: 10 mM NH_4Cl:B5 vitamins:3% glucose:0.3% gelrite:400 mg/L carb:75 mg/L kanamycin).

Embryogenic callus is identified 2-6 months following initiation and was subcultured onto fresh regeneration medium. Embryos are selected for germination, placed in static liquid Embryo Pulsing Medium (Stewart and Hsu medium: 0.01 mg/l NAA: 0.01 mg/L kinetin: 0.2 mg/L GA3) and incubated overnight at 30°C. The embryos are blotted on paper towels and placed into Magenta boxes containing 40 mls of Stewart and Hsu medium solidified with Gelrite. Germinating embryos are maintained at 28±2°C 50 uE $m^{-2}s^{-1}$ 16:8 photoperiod. Rooted plantlets are transferred to soil and established in the greenhouse.

Cotton growth conditions in growth chambers are as follows: 16 hour photoperiod, temperature of approximately 80-85°, light intensity of approximately 500μEinsteins. Cotton growth conditions in greenhouses are as follows: 14-16 hour photoperiod with light intensity of at least 400μEinsteins, day temperature 90-95°F, night temperature 70-75°F, relative humidity to approximately 80%.

Plant Analysis

Flowers from greenhouse grown T1 plants are tagged at anthesis in the greenhouse. Squares (cotton flower buds), flowers, bolls etc. are harvested from these plants at various stages of development and assayed for observable phenotype or tested for enzyme activity.

Example 7
Transformation of Tree Species

Numerous methods are known to the art for transforming forest tree species, for example U.S. Patent No. 5,654,190 discloses a process for producing transgenic plant belonging to the genus Populus, the section Leuce.

The above results demonstrate how the cellulose synthase cDNA may be used to alter the phenotype of a transgenic plant cell, and how the promoter may be used to modify transgenic cotton fiber cells.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application are specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail, by way of illustration and example for purposes of clarity and understanding, it will be readily apparent to those of ordinary skill in the art that certain changes and modifications may be made thereto, without departing from the spirit or scope of the appended claims.

CLAIMS

What is claimed is:

1. An isolated DNA encoding sequence to a plant cellulose synthesis enzyme.

2. The DNA encoding sequence of Claim 1 wherein said cellulose synthesis enzyme is cellulose synthase.

3. The DNA encoding sequence of Claim 2 wherein said cellulose synthase is from cotton.

4. The DNA encoding sequence of Claim 3 wherein said cotton cellulose synthase is celA1.

5. The DNA encoding sequence of Claim 4 wherein said celA1 is encoded by the sequence of Figure 6.

6. The DNA encoding sequence of Claim 3 wherein said cotton cellulose synthase is celA2.

7. The DNA encoding sequence of Claim 6 wherein said celA2 is encoded by the sequence of Figure 7.

8. An isolated DNA encoding sequence to a plant cellulose synthesis promoter region.

9. The promoter encoding sequence of Claim 8 wherein said cellulose synthesis promoter region is to cellulose synthase.

10. The promoter sequence of Claim 9 wherein said cellulose synthase promoter region is from cotton.

11. The promoter sequence of Claim 10 wherein said cotton cellulose synthase promoter region is from celA1.

12. The promoter sequence of Claim 11 wherein said cotton cellulose synthase promoter region is the from sequence of Figure 8.

13. A recombinant DNA construct comprising any of the DNA encoding sequences of Claims 1-10.

14. The DNA construct of Claim 13 comprising as operably joined components in the direction of transcription, a cotton fiber transcriptional factor and the sequence of any of Claims 1-7.

15. A plant cell comprising a DNA construct of Claims 13 or 14.

16. A plant comprising a cell of Claim 15.

17. A method of modifying fiber phenotype in a cotton plant, said method comprising:

transforming a plant cell with DNA comprising a construct of Claims 13 or 14.

18. A method of modifying the wood quality phenotype in a forest tree species, said method comprising:

transforming a plant cell of said species with DNA comprising a construct of Claim 13.

19. A method according to Claim 18 wherein said cellulose sythesis enzyme is cellulose synthase and wherein the encoding sequence is in an antisense orientation, wherein transcribed mRNA from said sequence is complementary to the equivalent mRNA transcribed from the endogenous gene, whereby the synthesis of cellulose in said plant cell is suppressed.

20. A method according to Claim 18, wherein said cellulose sythesis enzyme is cellulose synthase and wherein the encoding sequence is in a sense orientation, and wherein the synthesis of cellulose in said plant cell is increased.

21. A method according to Claim 20 wherein said plant cell additionally comprises a construct encoding a sequence to an enzyme involved in the synthesis of lignin or a lignin precursor.

22. A method according to Claim 20 wherein said lignin encoding sequence is in an antisense orientation, wherein transcribed mRNA from said sequence is complementary to the equivalent mRNA transcribed from the endogenous gene, whereby the synthesis of lignin is suppressed.

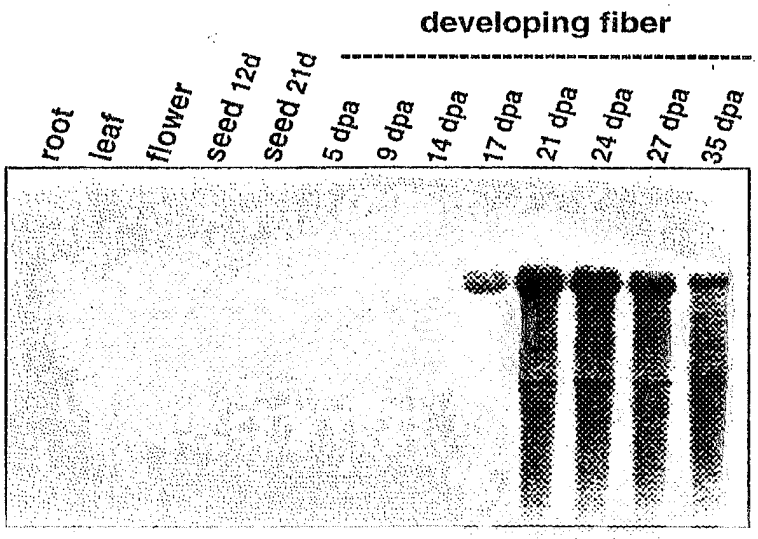

developing fiber
root
leaf
flower
seed 12d
seed 21d
5 dpa
9 dpa
14 dpa
17 dpa
21 dpa
24 dpa
27 dpa
35 dpa

FIG. 1

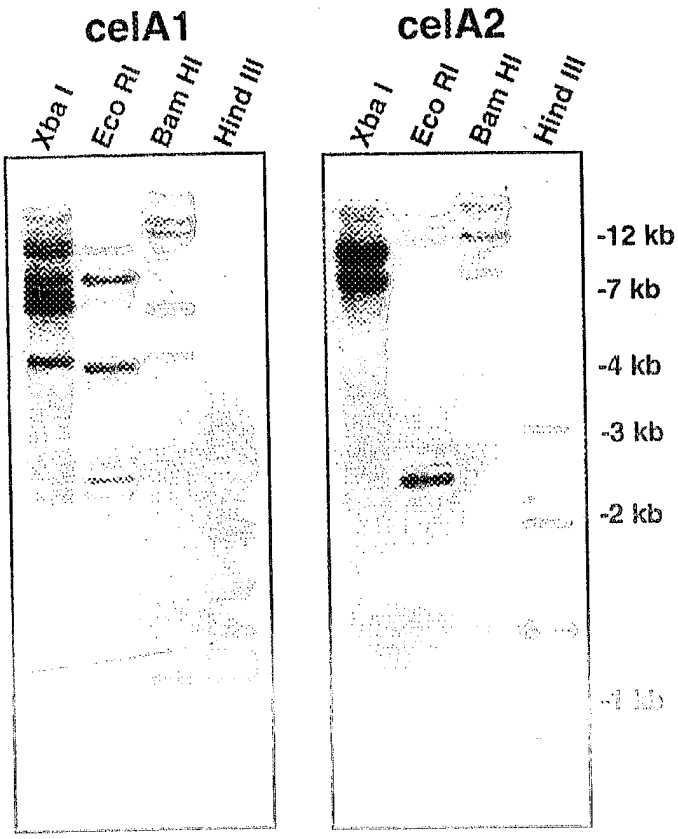

celA1
celA2
Xba I
Eco RI
Bam HI
Hind III
Xba I
Eco RI
Bam HI
Hind III
-12 kb
-7 kb
-4 kb
-3 kb
-2 kb

FIG. 2


G HIRSUTUM CELA-1
G HIRSUTUM CELA-2
O SATIVA CELA
A XYLINUM ACSA(I)
A XYLINUM BCSA
E COLI ORF F692
A TUMEFACIENS CELA
-G-
H-I
-U-I
P-CR

FIG. 3A

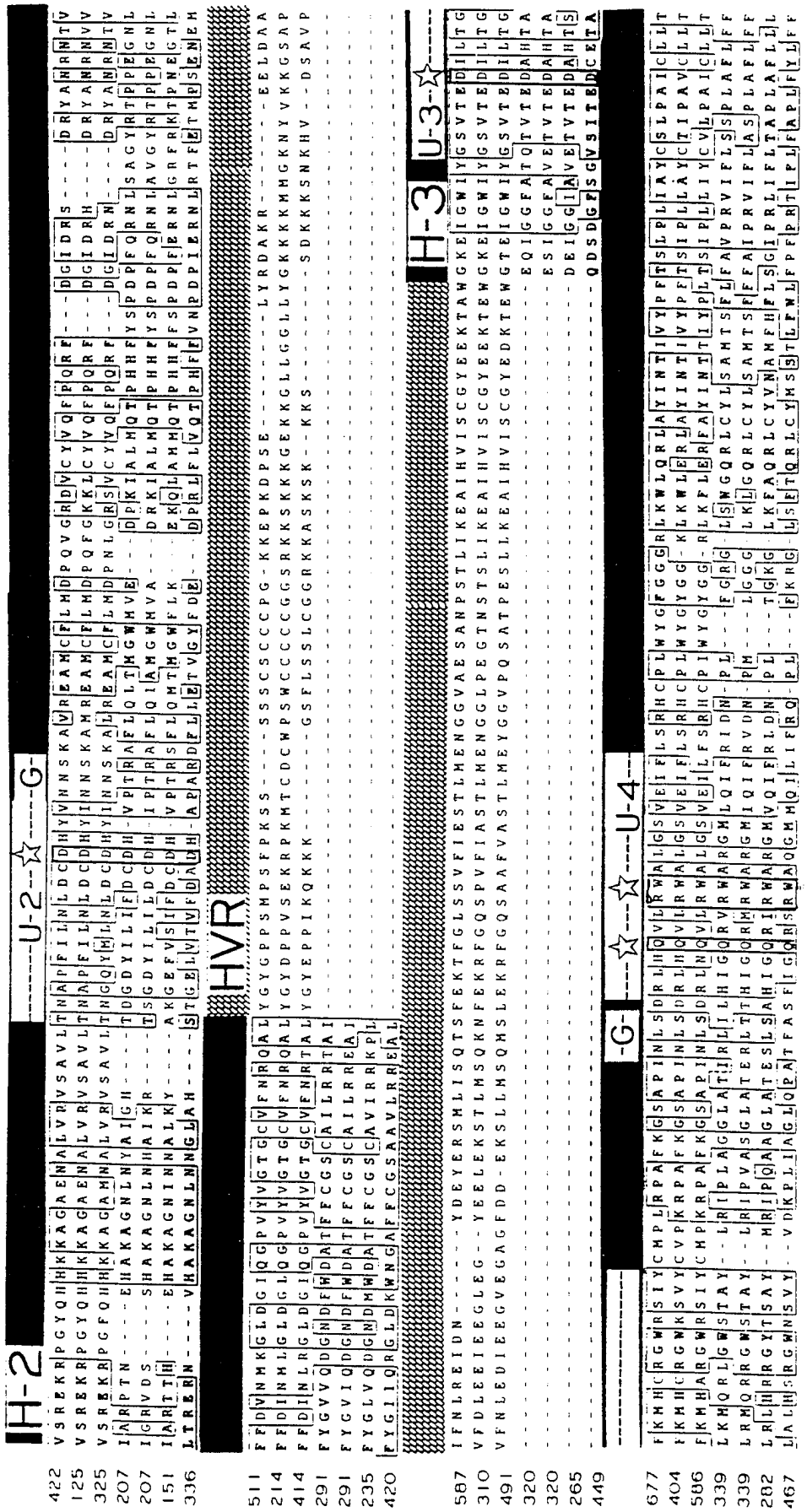


FIG. 3B

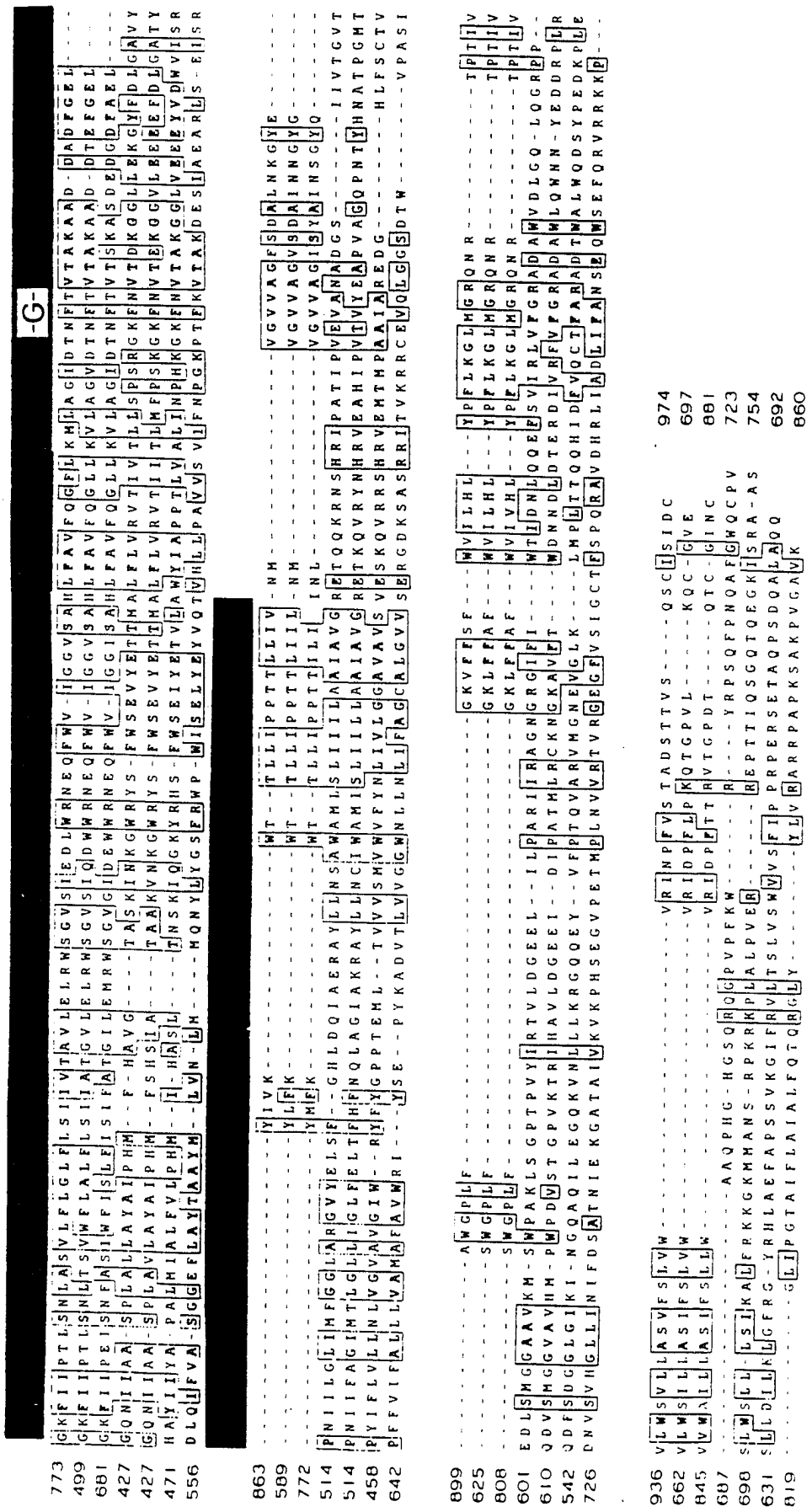


FIG. 3C

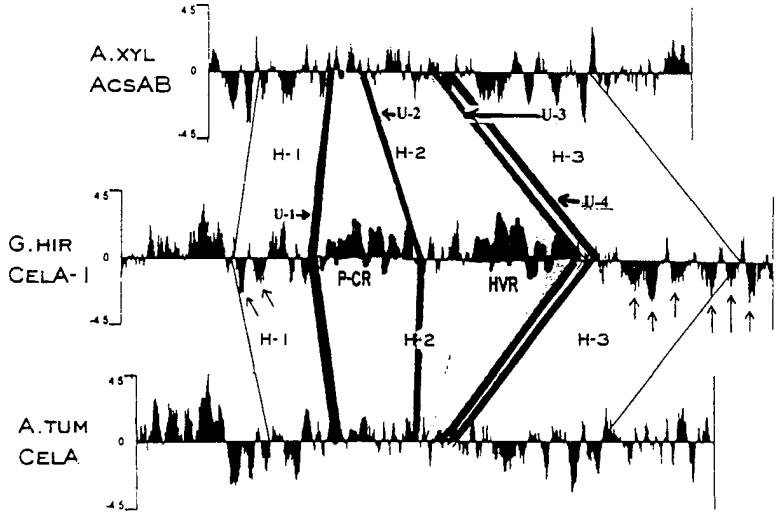

A.XYL
AcsAB
U-2
U-3
H-1
H-2
H-3
U-1
U-4
G.HIR
CELA-1
P-CR
HVR
H-1
H-2
H-3
A.TUM
CELA

FIG. 4

6/18

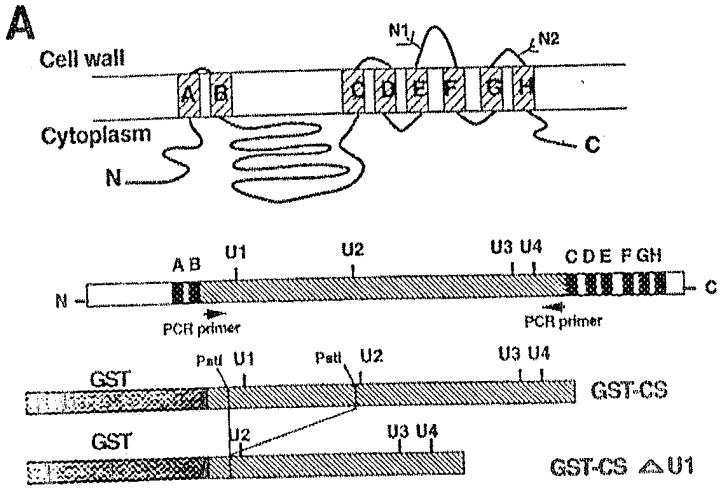

A
Cell wall
Cytoplasm
N
C
N1
N2
A
B
C
D
E
F
G
H
AB
U1
U2
U3 U4
CDE FGH
PCR primer
PCR primer
PstI
GST
GST-CS
GST-CS △U1

FIG. 5A

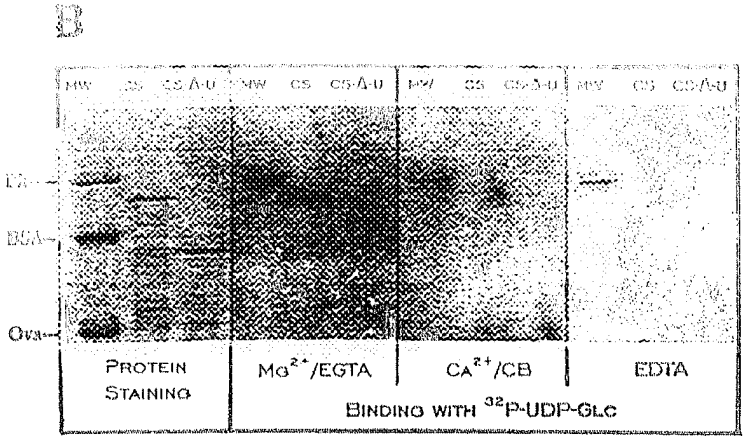

B
PROTEIN STAINING
Mg2+/EGTA
Ca2+/CB
EDTA
BINDING WITH 32P-UDP-GLC

FIG. 5B

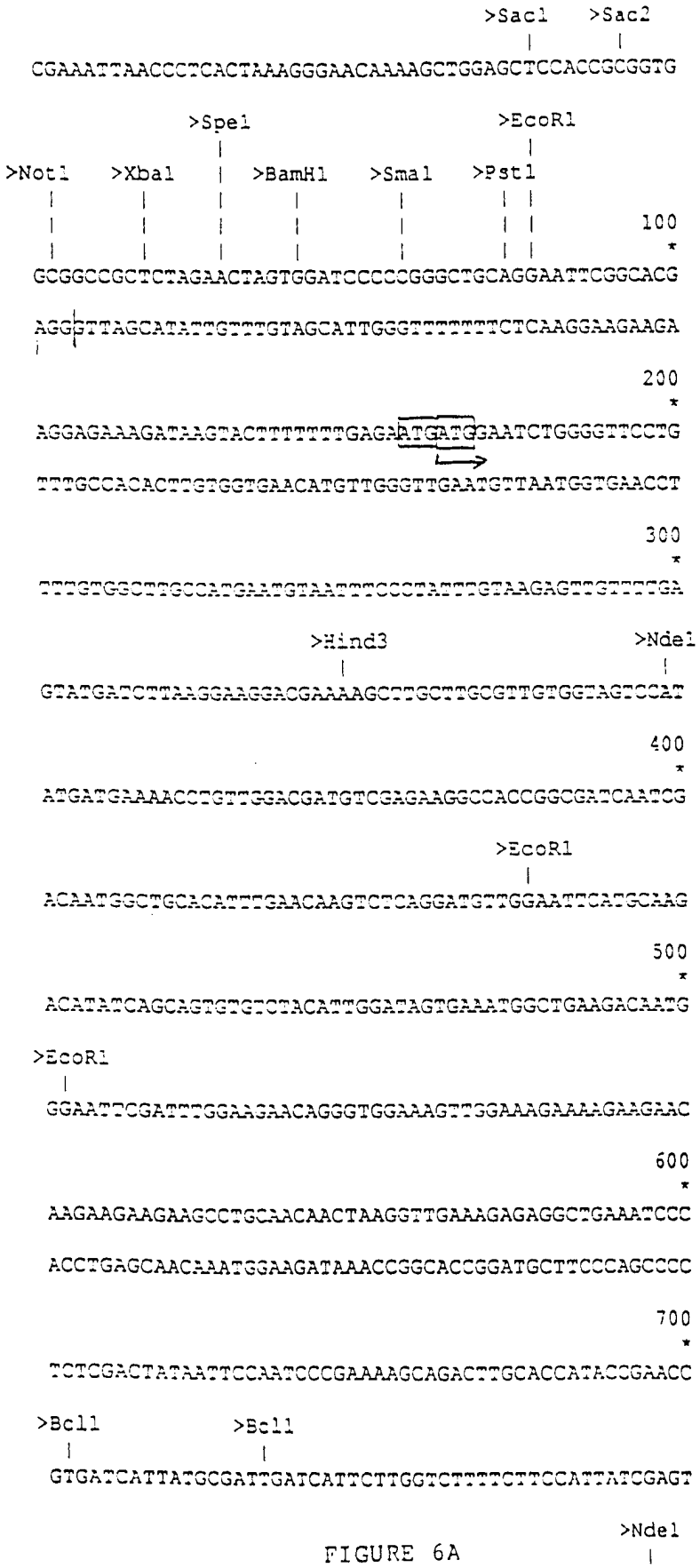

>Sac1
>Sac2
CGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTG
>Spe1
>EcoR1
>Not1
>Xba1
>BamH1
>Sma1
>Pst1
100
GCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACG
AGGGTTAGCATATTGTTTGTAGCATTGGGTTTTTTTCTCAAGGAAGAAGA
200
AGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAATCTGGGGTTCCTG
TTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTTAATGGTGAACCT
300
TTTGTGGCTTGCCATGAATGTAATTTCCCTATTTGTAAGAGTTGTTTTGA
>Hind3
>Nde1
GTATGATCTTAAGGAAGGACGAAAAGCTTGCTTGCGTTGTGGTAGTCCAT
400
ATGATGAAAACCTGTTGGACGATGTCGAGAAGGCCACCGGCGATCAATCG
>EcoR1
ACAATGGCTGCACATTTGAACAAGTCTCAGGATGTTGGAATTCATGCAAG
500
ACATATCAGCAGTGTGTCTACATTGGATAGTGAAATGGCTGAAGACAATG
>EcoR1
GGAATTCGATTTGGAAGAACAGGGTGGAAAGTTGGAAAGAAAAGAAGAAC
600
AAGAAGAAGAAGCCTGCAACAACTAAGGTTGAAAGAGAGGCTGAAATCCC
ACCTGAGCAACAAATGGAAGATAAACCGGCACCGGATGCTTCCCAGCCCC
700
TCTCGACTATAATTCCAATCCCGAAAAGCAGACTTGCACCATACCGAACC
>Bcl1
>Bcl1
GTGATCATTATGCGATTGATCATTCTTGGTCTTTTCTTCCATTATCGAGT
>Nde1

FIGURE 6A

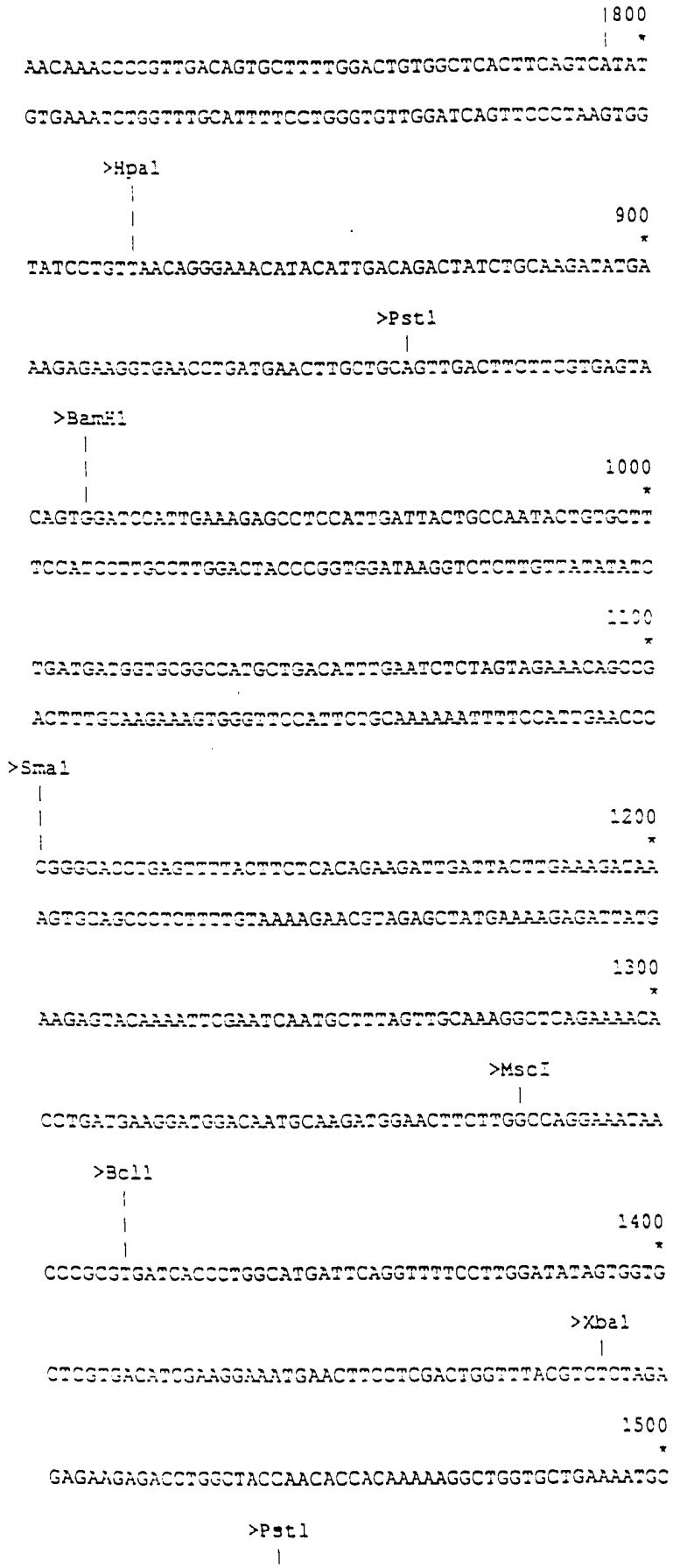

1800
AACAAACCCCGTTGACAGTGCTTTTGGACTGTGGCTCACTTCAGTCATAT
GTGAAATCTGGTTTGCATTTTCCTGGGTGTTGGATCAGTTCCCTAAGTGG
>HpaI
900
TATCCTGTTAACAGGGAAACATACATTGACAGACTATCTGCAAGATATGA
>PstI
AAGAGAAGGTGAACCTGATGAACTTGCTGCAGTTGACTTCTTCGTGAGTA
>BamHI
1000
CAGTGGATCCATTGAAAGAGCCTCCATTGATTACTGCCAATACTGTGCTT
TCCATCCTTGCCTTGGACTACCCGGTGGATAAGGTCTCTTGTTATATATC
1100
TGATGATGGTGCGGCCATGCTGACATTTGAATCTCTAGTAGAAACAGCCG
ACTTTGCAAGAAAGTGGGTTCCATTCTGCAAAAAATTTTCCATTGAACCC
>SmaI
1200
CGGGCACCTGAGTTTTACTTCTCACAGAAGATTGATTACTTGAAAGATAA
AGTGCAGCCCTCTTTTGTAAAAGAACGTAGAGCTATGAAAAGAGATTATG
1300
AAGAGTACAAAATTCGAATCAATGCTTTAGTTGCAAAGGCTCAGAAAACA
>MscI
CCTGATGAAGGATGGACAATGCAAGATGGAACTTCTTGGCCAGGAAATAA
>BclI
1400
CCCGCGTGATCACCCTGGCATGATTCAGGTTTTCCTTGGATATAGTGGTG
>XbaI
CTCGTGACATCGAAGGAAATGAACTTCCTCGACTGGTTTACGTCTCTAGA
1500
GAGAAGAGACCTGGCTACCAACACCACAAAAAGGCTGGTGCTGAAAATGC
>PstI

FIGURE 6B

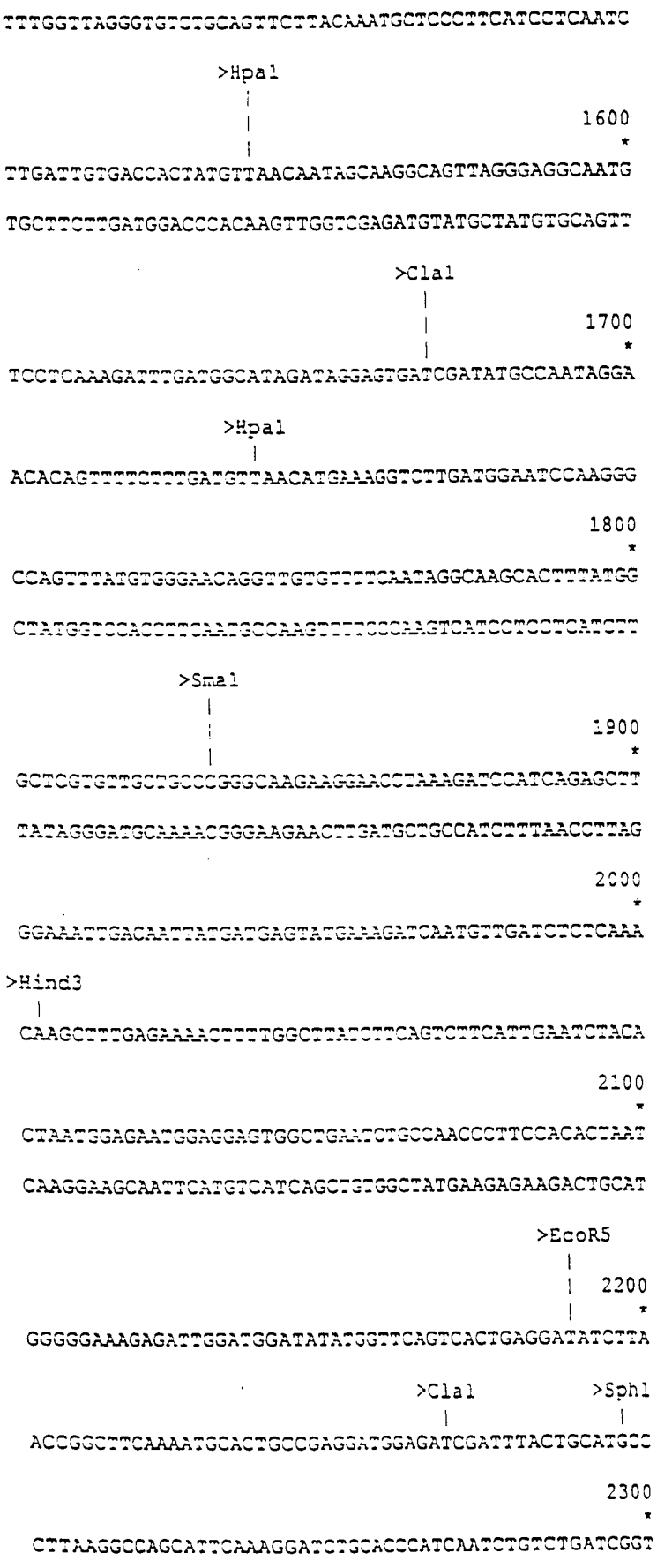

TTTGGTTAGGGTGTCTGCAGTTCTTACAAATGCTCCCTTCATCCTCAATC
>Hpa1
1600
TTGATTGTGACCACTATGTTAACAATAGCAAGGCAGTTAGGGAGGCAATG
TGCTTCTTGATGGACCCACAAGTTGGTCGAGATGTATGCTATGTGCAGTT
>Cla1
1700
TCCTCAAAGATTTGATGGCATAGATAGGAGTGATCGATATGCCAATAGGA
>Hpa1
ACACAGTTTTCTTTGATGTTAACATGAAAGGTCTTGATGGAATCCAAGGG
1800
CCAGTTTATGTGGGAACAGGTTGTGTTTTCAATAGGCAAGCACTTTATGG
CTATGGTCCACCTTCAATGCCAAGTTTTCCCAAGTCATCCTCCTCATCTT
>Sma1
1900
GCTCGTGTTGCTGCCCGGGCAAGAAGGAACCTAAAGATCCATCAGAGCTT
TATAGGGATGCAAAACGGGAAGAACTTGATGCTGCCATCTTTAACCTTAG
2000
GGAAATTGACAATTATGATGAGTATGAAAGATCAATGTTGATCTCTCAAA
>Hind3
CAAGCTTTGAGAAAACTTTTGGCTTATCTTCAGTCTTCATTGAATCTACA
2100
CTAATGGAGAATGGAGGAGTGGCTGAATCTGCCAACCCTTCCACACTAAT
CAAGGAAGCAATTCATGTCATCAGCTGTGGCTATGAAGAGAAGACTGCAT
>EcoR5
2200
GGGGGAAAGAGATTGGATGGATATATGGTTCAGTCACTGAGGATATCTTA
>Cla1
>Sph1
ACCGGCTTCAAAATGCACTGCCGAGGATGGAGATCGATTTACTGCATGCC
2300
CTTAAGGCCAGCATTCAAAGGATCTGCACCCATCAATCTGTCTGATCGGT

FIGURE 6C

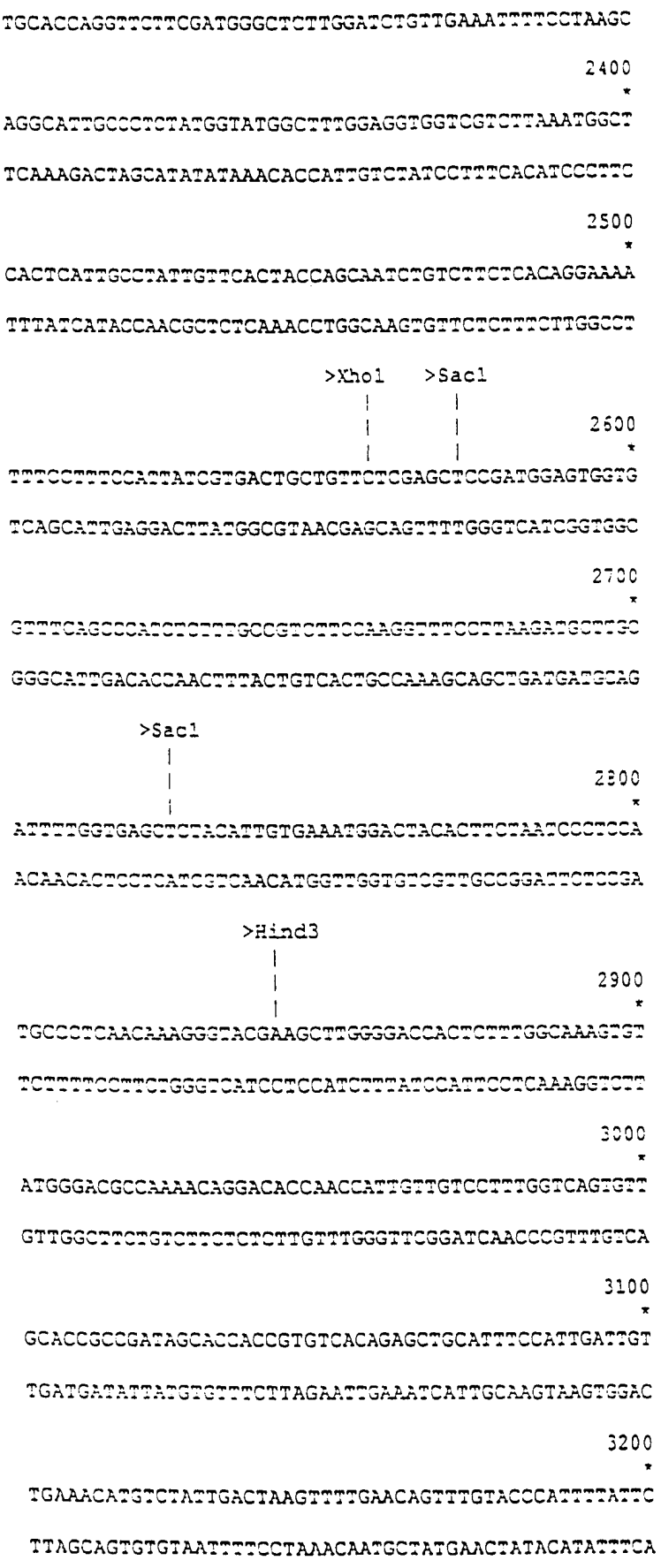

TGCACCAGGTTCTTCGATGGGCTCTTGGATCTGTTGAAATTTTCCTAAGC
2400
AGGCATTGCCCTCTATGGTATGGCTTTGGAGGTGGTCGTCTTAAATGGCT
TCAAAGACTAGCATATATAAACACCATTGTCTATCCTTTCACATCCCTTC
2500
CACTCATTGCCTATTGTTCACTACCAGCAATCTGTCTTCTCACAGGAAAA
TTTATCATACCAACGCTCTCAAACCTGGCAAGTGTTCTCTTTCTTGGCCT
>Xho1
>Sac1
2600
TTTCCTTTCCATTATCGTGACTGCTGTTCTCGAGCTCCGATGGAGTGGTG
TCAGCATTGAGGACTTATGGCGTAACGAGCAGTTTTGGGTCATCGGTGGC
2700
GTTTCAGCCCATCTCTTTGCCGTCTTCCAAGGTTTCCTTAAGATGCTTGC
GGGCATTGACACCAACTTTACTGTCACTGCCAAAGCAGCTGATGATGCAG
>Sac1
2800
ATTTTGGTGAGCTCTACATTGTGAAATGGACTACACTTCTAATCCCTCCA
ACAACACTCCTCATCGTCAACATGGTTGGTGTCGTTGCCGGATTCTCCGA
>Hind3
2900
TGCCCTCAACAAAGGGTACGAAGCTTGGGGACCACTCTTTGGCAAAGTGT
TCTTTTCCTTCTGGGTCATCCTCCATCTTTATCCATTCCTCAAAGGTCTT
3000
ATGGGACGCCAAAACAGGACACCAACCATTGTTGTCCTTTGGTCAGTGTT
GTTGGCTTCTGTCTTCTCTCTTGTTTGGGTTCGGATCAACCCGTTTGTCA
3100
GCACCGCCGATAGCACCACCGTGTCACAGAGCTGCATTTCCATTGATTGT
TGATGATATTATGTGTTTCTTAGAATTGAAATCATTGCAAGTAAGTGGAC
3200
TGAAACATGTCTATTGACTAAGTTTTGAACAGTTTGTACCCATTTTATTC
TTAGCAGTGTGTAATTTTCCTAAACAATGCTATGAACTATACATATTTCA

FIGURE 6D

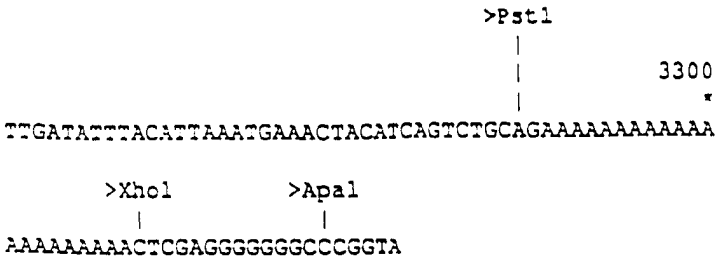

>Pst1
3300
TTGATATTTACATTAAATGAAACTACATCAGTCTGCAGAAAAAAAAAAAA
>Xho1
>Apa1
AAAAAAAAACTCGAGGGGGGGCCCGGTA

FIGURE 6E

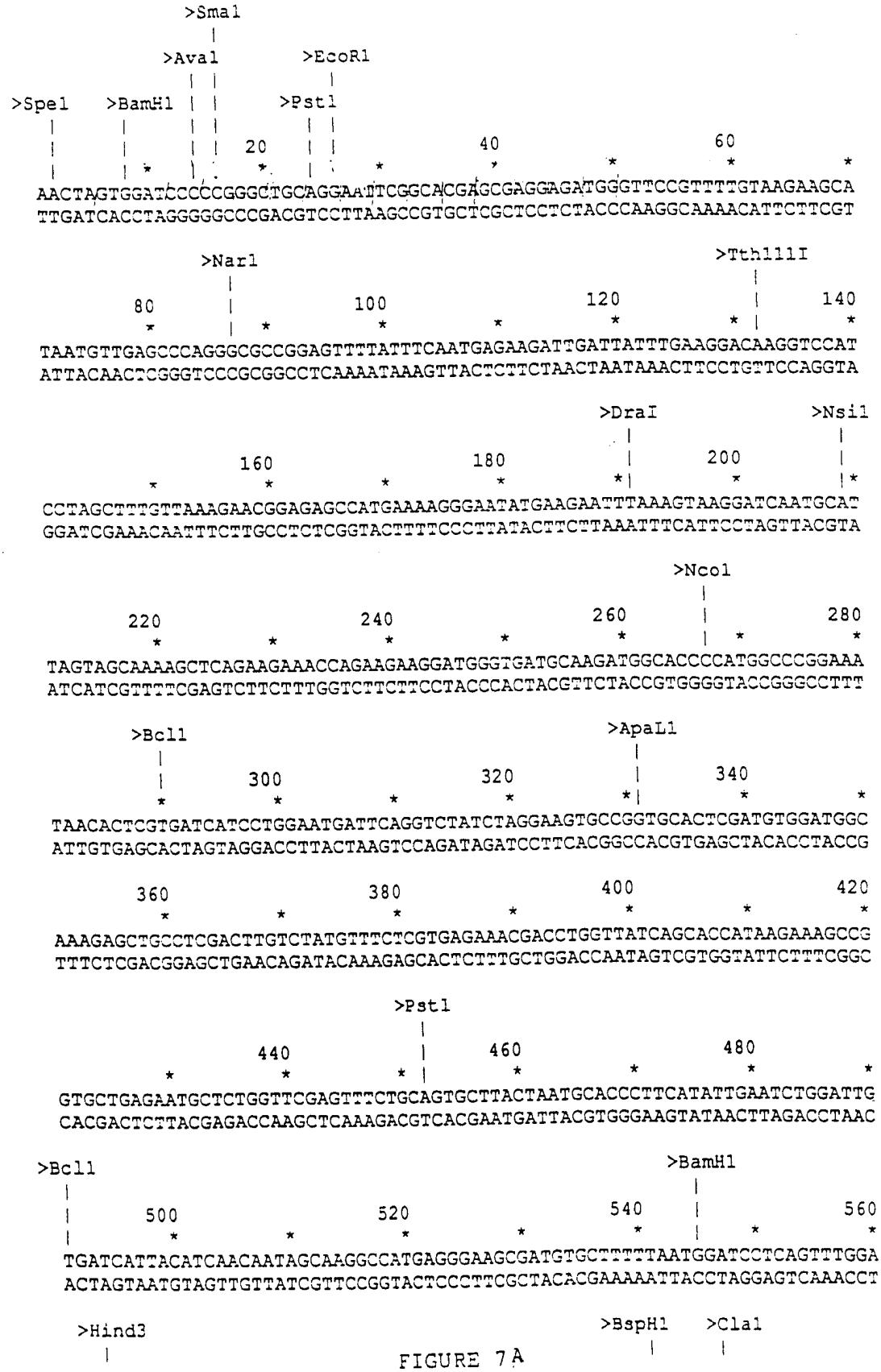

>Sma1
>Aval
>EcoR1
>Spe1
>BamH1
>Pst1
20
40
60
AACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACGAGCGAGGAGATGGGTTCCGTTTTGTAAGAAGCA
TTGATCACCTAGGGGGCCCGACGTCCTTAAGCCGTGCTCGCTCCTCTACCCAAGGCAAAACATTCTTCGT
>Nar1
>Tth111I
80
100
120
140
TAATGTTGAGCCCAGGGCGCCGGAGTTTTATTTCAATGAGAAGATTGATTATTTGAAGGACAAGGTCCAT
ATTACAACTCGGGTCCCGCGGCCTCAAAATAAAGTTACTCTTCTAACTAATAAACTTCCTGTTCCAGGTA
>DraI
>Nsi1
160
180
200
CCTAGCTTTGTTAAAGAACGGAGAGCCATGAAAAGGGAATATGAAGAATTTAAAGTAAGGATCAATGCAT
GGATCGAAACAATTTCTTGCCTCTCGGTACTTTTCCCTTATACTTCTTAAATTTCATTCCTAGTTACGTA
>Nco1
220
240
260
280
TAGTAGCAAAAGCTCAGAAGAAACCAGAAGAAGGATGGGTGATGCAAGATGGCACCCCATGGCCCGGAAA
ATCATCGTTTTCGAGTCTTCTTTGGTCTTCTTCCTACCCACTACGTTCTACCGTGGGGTACCGGGCCTTT
>Bcl1
>ApaL1
300
320
340
TAACACTCGTGATCATCCTGGAATGATTCAGGTCTATCTAGGAAGTGCCGGTGCACTCGATGTGGATGGC
ATTGTGAGCACTAGTAGGACCTTACTAAGTCCAGATAGATCCTTCACGGCCACGTGAGCTACACCTACCG
360
380
400
420
AAAGAGCTGCCTCGACTTGTCTATGTTTCTCGTGAGAAACGACCTGGTTATCAGCACCATAAGAAAGCCG
TTTCTCGACGGAGCTGAACAGATACAAAGAGCACTCTTTGCTGGACCAATAGTCGTGGTATTCTTTCGGC
>Pst1
440
460
480
GTGCTGAGAATGCTCTGGTTCGAGTTTCTGCAGTGCTTACTAATGCACCCTTCATATTGAATCTGGATTG
CACGACTCTTACGAGACCAAGCTCAAAGACGTCACGAATGATTACGTGGGAAGTATAACTTAGACCTAAC
>Bcl1
>BamH1
500
520
540
560
TGATCATTACATCAACAATAGCAAGGCCATGAGGGAAGCGATGTGCTTTTTAATGGATCCTCAGTTTGGA
ACTAGTAATGTAGTTGTTATCGTTCCGGTACTCCCTTCGCTACACGAAAAATTACCTAGGAGTCAAACCT
>Hind3
>BspH1
>Cla1

FIGURE 7A

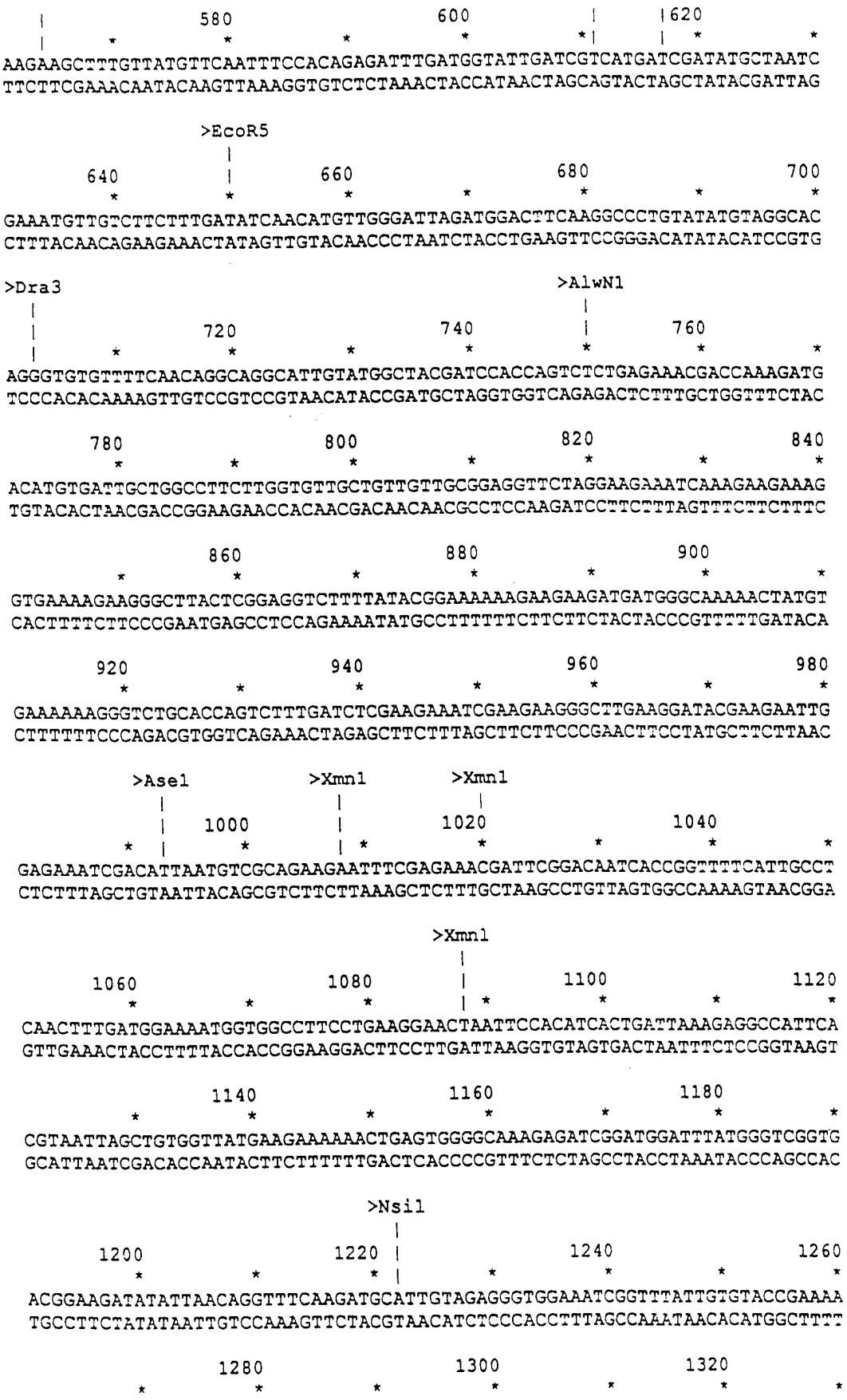

580
600
620
AAGAAGCTTTGTTATGTTCAATTTCCACAGAGATTTGATGGTATTGATCGTCATGATCGATATGCTAATC
TTCTTCGAAACAATACAAGTTAAAGGTGTCTCTAAACTACCATAACTAGCAGTACTAGCTATACGATTAG
>EcoR5
640
660
680
700
GAAATGTTGTCTTCTTTGATATCAACATGTTGGGATTAGATGGACTTCAAGGCCCTGTATATGTAGGCAC
CTTTACAACAGAAGAAACTATAGTTGTACAACCCTAATCTACCTGAAGTTCCGGGACATATACATCCGTG
>Dra3
>AlwN1
720
740
760
AGGGTGTGTTTTCAACAGGCAGGCATTGTATGGCTACGATCCACCAGTCTCTGAGAAACGACCAAAGATG
TCCCACACAAAAGTTGTCCGTCCGTAACATACCGATGCTAGGTGGTCAGAGACTCTTTGCTGGTTTCTAC
780
800
820
840
ACATGTGATTGCTGGCCTTCTTGGTGTTGCTGTTGTTGCGGAGGTTCTAGGAAGAAATCAAAGAAGAAAG
TGTACACTAACGACCGGAAGAACCACAACGACAACAACGCCTCCAAGATCCTTCTTTAGTTTCTTCTTTC
860
880
900
GTGAAAAGAAGGGCTTACTCGGAGGTCTTTTATACGGAAAAAAGAAGAAGATGATGGGCAAAAACTATGT
CACTTTTCTTCCCGAATGAGCCTCCAGAAAATATGCCTTTTTTCTTCTTCTACTACCCGTTTTTGATACA
920
940
960
980
GAAAAAAGGGTCTGCACCAGTCTTTGATCTCGAAGAAATCGAAGAAGGGCTTGAAGGATACGAAGAATTG
CTTTTTTCCCAGACGTGGTCAGAAACTAGAGCTTCTTTAGCTTCTTCCCGAACTTCCTATGCTTCTTAAC
>Ase1
>Xmn1
>Xmn1
1000
1020
1040
GAGAAATCGACATTAATGTCGCAGAAGAATTTCGAGAAACGATTCGGACAATCACCGGTTTTCATTGCCT
CTCTTTAGCTGTAATTACAGCGTCTTCTTAAAGCTCTTTGCTAAGCCTGTTAGTGGCCAAAAGTAACGGA
>Xmn1
1060
1080
1100
1120
CAACTTTGATGGAAAATGGTGGCCTTCCTGAAGGAACTAATTCCACATCACTGATTAAAGAGGCCATTCA
GTTGAAACTACCTTTTACCACCGGAAGGACTTCCTTGATTAAGGTGTAGTGACTAATTTCTCCGGTAAGT
1140
1160
1180
CGTAATTAGCTGTGGTTATGAAGAAAAAACTGAGTGGGGCAAAGAGATCGGATGGATTTATGGGTCGGTG
GCATTAATCGACACCAATACTTCTTTTTTGACTCACCCCGTTTCTCTAGCCTACCTAAATACCCAGCCAC
>Nsi1
1200
1220
1240
1260
ACGGAAGATATATTAACAGGTTTCAAGATGCATTGTAGAGGGTGGAAATCGGTTTATTGTGTACCGAAAA
TGCCTTCTATATAATTGTCCAAAGTTCTACGTAACATCTCCCACCTTTAGCCAAATAACACATGGCTTTT
1280
1300
1320

FIGURE 7B

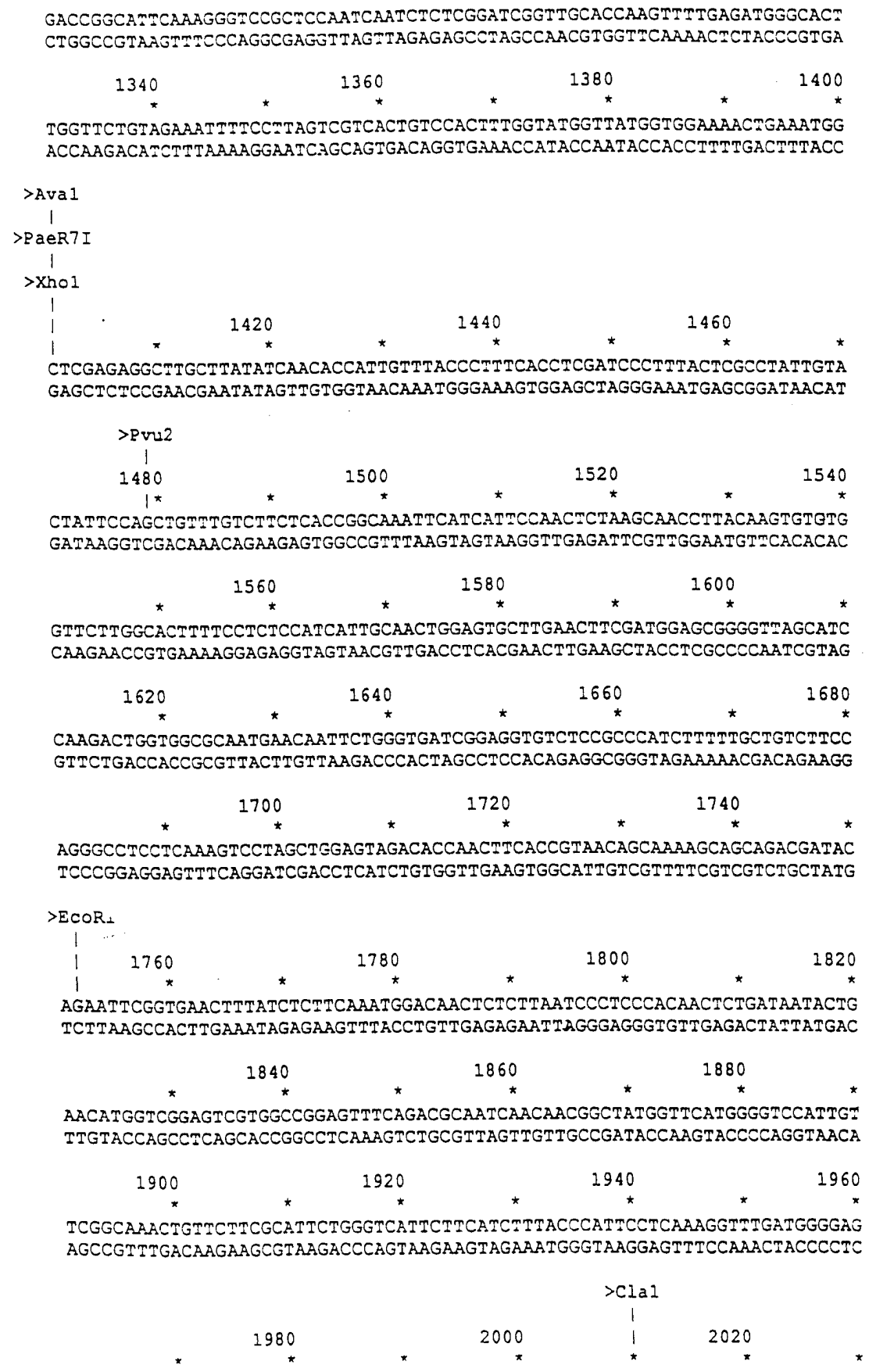

GACCGGCATTCAAAGGGTCCGCTCCAATCAATCTCTCGGATCGGTTGCACCAAGTTTTGAGATGGGCACT
CTGGCCGTAAGTTTCCCAGGCGAGGTTAGTTAGAGAGCCTAGCCAACGTGGTTCAAAACTCTACCCGTGA
1340
1360
1380
1400
TGGTTCTGTAGAAATTTTCCTTAGTCGTCACTGTCCACTTTGGTATGGTTATGGTGGAAAACTGAAATGG
ACCAAGACATCTTTAAAAGGAATCAGCAGTGACAGGTGAAACCATACCAATACCACCTTTTGACTTTACC
>Ava1
>PaeR7I
>Xho1
1420
1440
1460
CTCGAGAGGCTTGCTTATATCAACACCATTGTTTACCCTTTCACCTCGATCCCTTTACTCGCCTATTGTA
GAGCTCTCCGAACGAATATAGTTGTGGTAACAAATGGGAAAGTGGAGCTAGGGAAATGAGCGGATAACAT
>Pvu2
1480
1500
1520
1540
CTATTCCAGCTGTTTGTCTTCTCACCGGCAAATTCATCATTCCAACTCTAAGCAACCTTACAAGTGTGTG
GATAAGGTCGACAAACAGAAGAGTGGCCGTTTAAGTAGTAAGGTTGAGATTCGTTGGAATGTTCACACAC
1560
1580
1600
GTTCTTGGCACTTTTCCTCTCCATCATTGCAACTGGAGTGCTTGAACTTCGATGGAGCGGGGTTAGCATC
CAAGAACCGTGAAAAGGAGAGGTAGTAACGTTGACCTCACGAACTTGAAGCTACCTCGCCCCAATCGTAG
1620
1640
1660
1680
CAAGACTGGTGGCGCAATGAACAATTCTGGGTGATCGGAGGTGTCTCCGCCCATCTTTTTGCTGTCTTCC
GTTCTGACCACCGCGTTACTTGTTAAGACCCACTAGCCTCCACAGAGGCGGGTAGAAAAACGACAGAAGG
1700
1720
1740
AGGGCCTCCTCAAAGTCCTAGCTGGAGTAGACACCAACTTCACCGTAACAGCAAAAGCAGCAGACGATAC
TCCCGGAGGAGTTTCAGGATCGACCTCATCTGTGGTTGAAGTGGCATTGTCGTTTTCGTCGTCTGCTATG
>EcoR1
1760
1780
1800
1820
AGAATTCGGTGAACTTTATCTCTTCAAATGGACAACTCTCTTAATCCCTCCCACAACTCTGATAATACTG
TCTTAAGCCACTTGAAATAGAGAAGTTTACCTGTTGAGAGAATTAGGGAGGGTGTTGAGACTATTATGAC
1840
1860
1880
AACATGGTCGGAGTCGTGGCCGGAGTTTCAGACGCAATCAACAACGGCTATGGTTCATGGGGTCCATTGT
TTGTACCAGCCTCAGCACCGGCCTCAAAGTCTGCGTTAGTTGTTGCCGATACCAAGTACCCCAGGTAACA
1900
1920
1940
1960
TCGGCAAACTGTTCTTCGCATTCTGGGTCATTCTTCATCTTTACCCATTCCTCAAAGGTTTGATGGGGAG
AGCCGTTTGACAAGAAGCGTAAGACCCAGTAAGAAGTAGAAATGGGTAAGGAGTTTCCAAACTACCCCTC
>Cla1
1980
2000
2020

FIGURE 7C

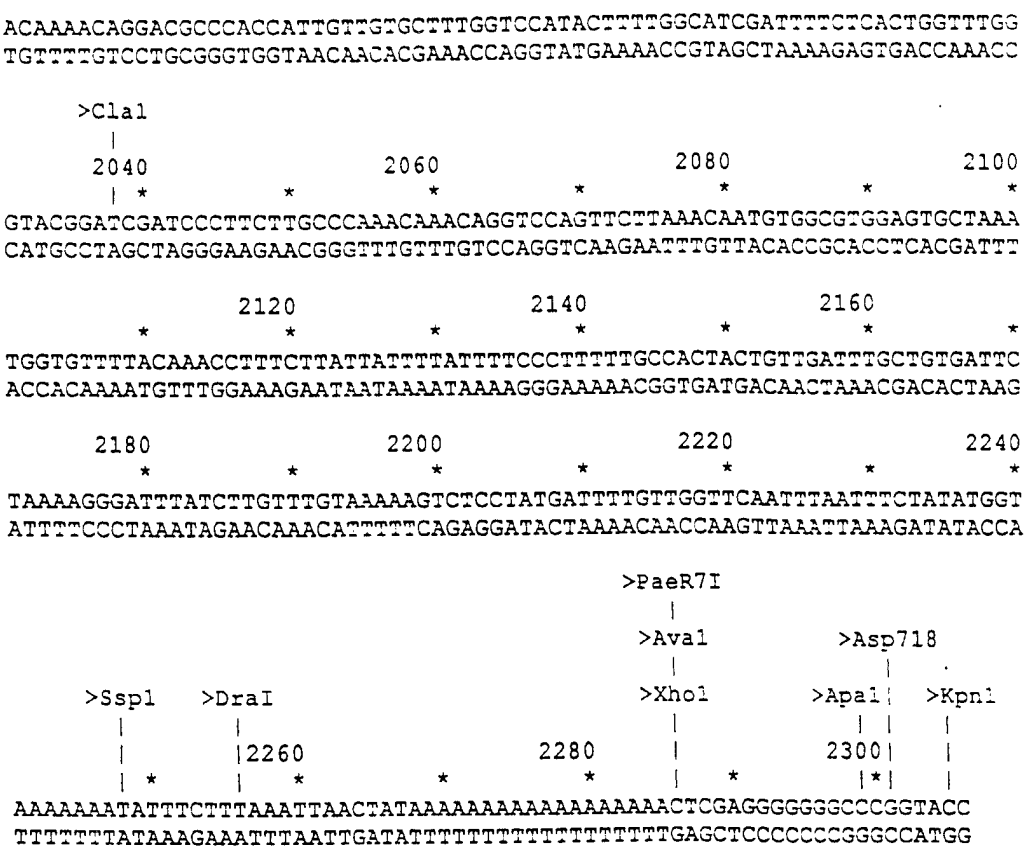

ACAAAACAGGACGCCCACCATTGTTGTGCTTTGGTCCATACTTTTGGCATCGATTTTCTCACTGGTTTGG
TGTTTTGTCCTGCGGGTGGTAACAACACGAAACCAGGTATGAAAACCGTAGCTAAAAGAGTGACCAAACC
>ClaI
2040
2060
2080
2100
GTACGGATCGATCCCTTCTTGCCCAAACAAACAGGTCCAGTTCTTAAACAATGTGGCGTGGAGTGCTAAA
CATGCCTAGCTAGGGAAGAACGGGTTTGTTTGTCCAGGTCAAGAATTTGTTACACCGCACCTCACGATTT
2120
2140
2160
TGGTGTTTTACAAACCTTTCTTATTATTTTATTTTCCCTTTTTGCCACTACTGTTGATTTGCTGTGATTC
ACCACAAAATGTTTGGAAAGAATAATAAAATAAAAGGGAAAAACGGTGATGACAACTAAACGACACTAAG
2180
2200
2220
2240
TAAAAGGGATTTATCTTGTTTGTAAAAAGTCTCCTATGATTTTGTTGGTTCAATTTAATTTCTATATGGT
ATTTTCCCTAAATAGAACAAACATTTTTCAGAGGATACTAAAACAACCAAGTTAAATTAAAGATATACCA
>PaeR7I
>Aval
>Asp718
>SspI
>DraI
>XhoI
>ApaI
>KpnI
2260
2280
2300
AAAAAAATATTTCTTTAAATTAACTATAAAAAAAAAAAAAAAAAACTCGAGGGGGGGCCCGGTACC
TTTTTTTATAAAGAAATTTAATTGATATTTTTTTTTTTTTTTTTGAGCTCCCCCCCGGGCCATGG

FIGURE 7D

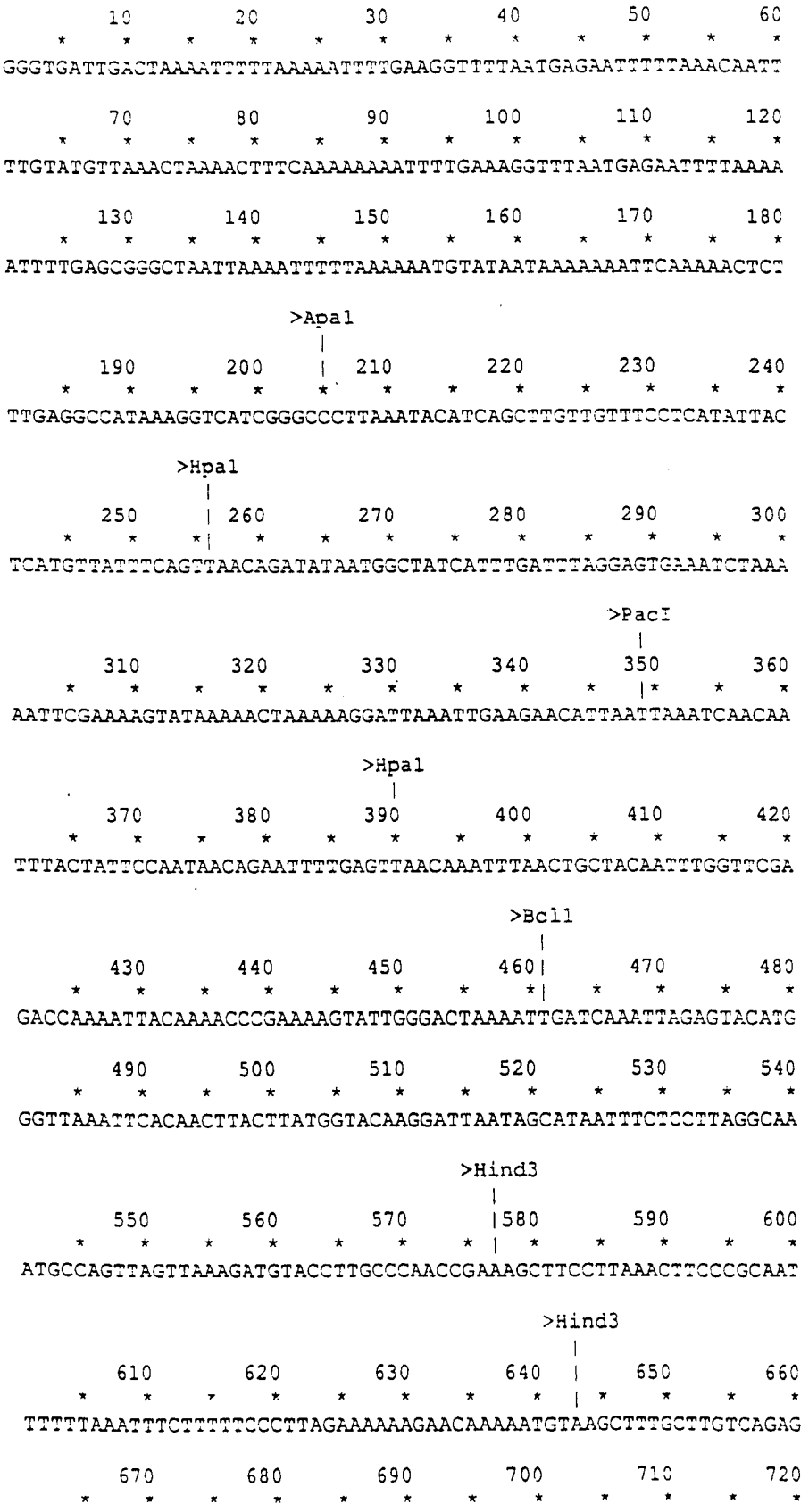

10 20 30 40 50 60
GGGTGATTGACTAAAATTTTTAAAAATTTTGAAGGTTTTAATGAGAATTTTTAAACAATT
70 80 90 100 110 120
TTGTATGTTAAACTAAAACTTTCAAAAAAAATTTTGAAAGGTTTAATGAGAATTTTAAAA
130 140 150 160 170 180
ATTTTGAGCGGGCTAATTAAAATTTTTAAAAAATGTATAATAAAAAAATTCAAAAACTCT
>Apa1
190 200 210 220 230 240
TTGAGGCCATAAAGGTCATCGGGCCCTTAAATACATCAGCTTGTTGTTTCCTCATATTAC
>Hpa1
250 260 270 280 290 300
TCATGTTATTTCAGTTAACAGATATAATGGCTATCATTTGATTTAGGAGTGAAATCTAAA
>Pac1
310 320 330 340 350 360
AATTCGAAAAGTATAAAAACTAAAAAGGATTAAATTGAAGAACATTAATTAAATCAACAA
>Hpa1
370 380 390 400 410 420
TTTACTATTCCAATAACAGAATTTTGAGTTAACAAATTTAACTGCTACAATTTGGTTCGA
>Bcl1
430 440 450 460 470 480
GACCAAAATTACAAAACCCGAAAAGTATTGGGACTAAAATTGATCAAATTAGAGTACATG
490 500 510 520 530 540
GGTTAAATTCACAACTTACTTATGGTACAAGGATTAATAGCATAATTTCTCCTTAGGCAA
>Hind3
550 560 570 580 590 600
ATGCCAGTTAGTTAAAGATGTACCTTGCCCAACCGAAAGCTTCCTTAAACTTCCCGCAAT
>Hind3
610 620 630 640 650 660
TTTTTAAATTTCTTTTTCCCTTAGAAAAAAGAACAAAAATGTAAGCTTTGCTTGTCAGAG
670 680 690 700 710 720

FIGURE 8A

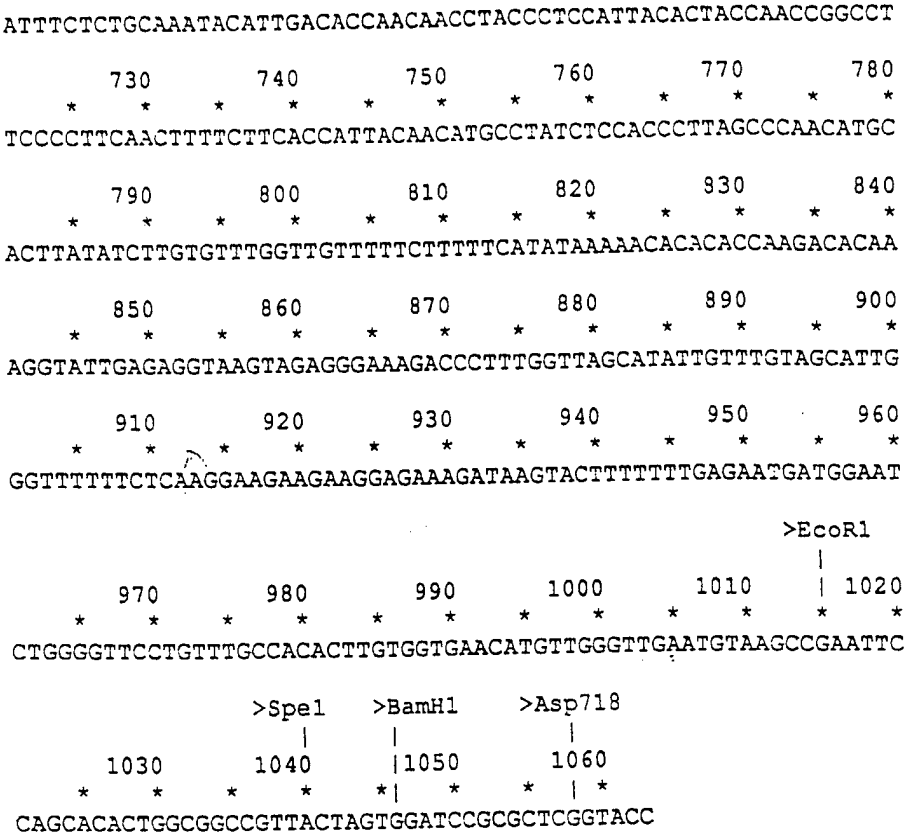

ATTTCTCTGCAAATACATTGACACCAACAACCTACCCTCCATTACACTACCAACCGGCCT
730 740 750 760 770 780
TCCCCTTCAACTTTTCTTCACCATTACAACATGCCTATCTCCACCCTTAGCCCAACATGC
790 800 810 820 830 840
ACTTATATCTTGTGTTTGGTTGTTTTTCTTTTTCATATAAAAACACACACCAAGACACAA
850 860 870 880 890 900
AGGTATTGAGAGGTAAGTAGAGGGAAAGACCCTTTGGTTAGCATATTGTTTGTAGCATTG
910 920 930 940 950 960
GGTTTTTTCTCAAGGAAGAAGAAGGAGAAAGATAAGTACTTTTTTTGAGAATGATGGAAT
>EcoR1
970 980 990 1000 1010 1020
CTGGGGTTCCTGTTTGCCACACTTGTGGTGAACATGTTGGGTTGAATGTAAGCCGAATTC
>Spe1 >BamH1 >Asp718
1030 1040 1050 1060
CAGCACACTGGCGGCCGTTACTAGTGGATCCGCGCTCGGTACC

FIGURE 8B