Office de la Propriété Intellectuelle du Canada

Un organisme d'Industrie Canada

Canadian Intellectual Property Office

An agency of Industry Canada

CA 2136618 C 2004/11/16

(11)(21) 2 136 618

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1993/05/29

(87) Date publication PCT/PCT Publication Date: 1993/12/09

(45) Date de délivrance/Issue Date: 2004/11/16

(85) Entrée phase nationale/National Entry: 1994/11/24

(86) N° demande PCT/PCT Application No.: GB 1993/001098

(87) N° publication PCT/PCT Publication No.: 1993/024638

(30) Priorité/Priority: 1992/05/29 (9211416.4) GB

(51) Cl.Int.^{5}/Int.Cl.^{5} C12N 15/82, A01H 5/00, C12N 15/11, C12N 5/10

(72) Inventeur/Inventor:
WALTER, MICHAEL H., DE

(73) Propriétaire/Owner:
SYNGENTA LIMITED, GB

(74) Agent: AIRD & BERLIS LLP

(54) Titre : EXPRESSION DE GENES CHEZ DES PLANTES TRANSGENIQUES
(54) Title: EXPRESSION OF GENES IN TRANSGENIC PLANTS

(57) Abrégé/Abstract:
A-gene promoter sequence has been isolated and characterised from the plant gene encoding the enzyme cinnamyl alcohol dehydrogenase (CAD), an enzyme which participates in the biosynthesis of lignins. The promoter may be used to control expression of exogenous genes placed under its control. Genes controlled by the CAD promoter are inducible by wounding.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/82, 15/53, 9/04
C12N 5/10, A01H 5/00

A1

2136618

(11) International Publication Number: WO 93/24638

(43) International Publication Date: 9 December 1993 (09.12.93)

(21) International Application Number: PCT/GB93/01098

(22) International Filing Date: 27 May 1993 (27.05.93)

(30) Priority data:
9211416.4 29 May 1992 (29.05.92) GB

(71) Applicant *(for all designated States except US):* ZENECA LIMITED [GB/GB]; Imperial Chemical House, 9 Millbank, London SW1P 3JF (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : WALTER, Michael, Herbert [DE/DE]; Talstrasse 103, D-7024 Filderstadt 1 (DE).

(74) Agent: HUSKISSON, Frank, Mackie; Zeneca Limited, Group Patent Services Department, P.O. Box 6, Bessemer Road, Welwyn Garden City, Herts AL7 1HD (GB).

(81) Designated States: AU, BB, BG, BR, CA, CZ, FI, HU, JP, KP, KR, LK, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: EXPRESSION OF GENES IN TRANSGENIC PLANTS

(57) Abstract

A gene promoter sequence has been isolated and characterised from the plant gene encoding the enzyme cinnamyl alcohol dehydrogenase (CAD), an enzyme which participates in the biosynthesis of lignins. The promoter may be used to control expression of exogenous genes placed under its control. Genes controlled by the CAD promoter are inducible by wounding.

EXPRESSION OF GENES IN TRANSGENIC PLANTS

The present invention relates to the regulation of gene expression in transgenic plants. In particular it is concerned with the isolation and use of DNA sequences which control the expression of genes in lignifying tissues and in response to pathogen attack.

The ability to isolate and manipulate plant genes has opened the way to gain understanding about the mechanisms involved in the regulation of plant gene expression. This knowledge is important for the exploitation of genetic engineering techniques to practical problems such as the expression of genes in genetically manipulated crop plants exhibiting improved quality and production characteristics.

Many examples have been reported in the literature of plant DNA sequences which have been used to drive the expression of foreign genes in plants. In most instances the regions immediately 5′ to the coding regions of genes have been used in gene constructs. These regions are referred to as promoter sequences. They may be derived from plant DNA; or from other sources, eg, viruses. It has been demonstrated that sequences up to 500-1000 bases in most instances are sufficient to allow for the regulated expression of foreign genes. The

types of regulation exemplified include tissue-specificity, regulation by external factors such as light, heat treatment, chemicals, hormones, and developmental regulation. However, it has also been shown that sequences much longer than 1 kb may have useful features which permit high levels of gene expression in transgenic plants.

These experiments have largely been carried out using gene fusions between the promoter sequences and foreign structural genes such as bacterial genes, etc. This has led to the identification of useful promoter sequences.

In the work leading to the present invention we have identified a gene which expresses an enzyme involved in biosynthesis of lignin in vascular plants. We have shown that this gene encodes cimmamyl alcohol dehydrogenase (CAD) which is one of the enzymes specific to the lignin branch pathway of phenylpropanoid metabolism. The gene in question is encoded by the cDNA clone pTCAD19 among others which are the subject of our published copending International Patent Application Number WO 93/05159.

The enzyme encoded by the CAD gene catalyses the oxidation of cinnamyl aldehydes to cinnamyl alcohols which are the direct precursors of the lignin polymer. We have isolated the gene encoding this enzyme from tobacco.

We have shown that CAD mRNA is expressed throughout plant development with the highest levels found in lignifying tissues. In these tissues the CAD enzyme can be found in high concentration in the xylem of plants.

An object of the present invention is to

provide a promoter sequence suitable for the control of foreign gene expression in plants.

According to the present invention there is provided a DNA construct for use in transforming plant cells which comprises an exogenous coding sequence under the control of upstream promoter and downstream terminator sequences, characterised in that the upstream promoter is or has functional homology to a promoter of a gene of the lignin biosynthesis pathway.

Thus, the present invention comprises the use of the promoter(s) of the cinnamyl alcohol dehydrogenase or other genes involved in the lignin biosynthesis pathway to control the expression of novel and exogenous proteins and genes in these tissues.

Preferably, the promoter is the CAD promoter.

The clone gNtCAD9-6.3SH contains the promoter fragment of this gene. This clone has been deposited at the National Collections of Industrial and Marine Bacteria (NCIB), at 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland, on 2nd April 1992 under the reference NCIB Number 40499.

We further provide novel plant cells, and plants transformed with constructs according to the present invention. We illustrate the invention hereinafter using tobacco as a model plant species.

The constructs can be used in the expression of exogenous genes and proteins in vascular tissues, particularly but not exclusively, such as poplar, eucalyptus, pine, and other woody plants as well as forages such as festuca, alfalfa, maize, sorghum, penesitum, amongst others.

Not only will this promoter be expressed in

2136618

modified plants in a given tissue but it will also be induced in response to environmental and endogenous signals such as wounding, infection, ethylene production and others.

5 Promoters for use in the invention may be derived from genes such as methyl transferase, and cinnamyl CoA reductase. Such promoters may be isolated from genomic libraries by the use of cDNA probes, as has been done in the case of CAD. We 10 particularly prefer to use the promoter of the CAD gene.

The downstream (3') terminator sequences may also be derived from the CAD gene or they may be derived from other genes. Many possibilities are 15 available from the literature: the selection of the terminator being of rather lesser importance.

By the term 'exogenous coding sequence' we mean a sequence of DNA, other than that which follows the promoter region in the natural CAD 20 gene, that is adapted to be transcribed into functional RNA under the action of plant cell enzymes such as RNA polymerase. Functional RNA is RNA which affects the biochemistry of the cell: it may for example be mRNA which is translated into 25 protein by ribosomes; or antisense RNA which inhibits the translation of mRNA complementary (or otherwise related) to it into protein. In principle any exogenous coding sequence may be used in the present invention.

30 Where the exogenous coding sequence codes for mRNA for a protein, this protein may be of bacterial origin (such as enzymes involved in cell wall modification and cell wall metabolism), of eukaryotic origin (such as pharmaceutically active

polypeptides) or of plant origin (such as enzymes involved in the synthesis of phenolic compounds, ethylene synthesis, sugar metabolism, cell wall metabolism,), or genes or parts thereof in sense
5 and antisense orientation.

A wide variety of exogenous coding sequences is known from the literature, and the present invention is applicable to these as well as many others yet to be reported. As well as functional
10 mRNA, the exogenous gene may code for RNA that interferes with the function of any kind of mRNA produced by the plant cell: for example, antisense RNA complementary to mRNA for genes such as stilbene synthesis, phytoalexin synthesis and
15 flavour and pigment synthesis.

Of particular interest is the ability of the CAD gene promoter to respond to exogenous stimuli.

The construction of vectors and constructs of the present invention will be described in more
20 detail in the Examples below. For convenience it will be generally found suitable to use promoter sequences (upstream - i.e. 5' - of the coding sequence of the gene) of between 100 and 2000 bases in length.

25 A particularly preferred embodiment of the invention is a promoter for use in the expression of exogenous genes in plants, comprising the promoter of cinnamyl alcohol dehydrogenase.

More particularly the promoter of the
30 invention comprises the nucleotide sequence shown in Figure 2. The invention includes modifications or the use of only parts of the said sequence which, while retaining sufficient homology to the said sequence in order to maintain functionality,

enhance or alter its tissue-specificity or response to external stimuli.

The invention further provides a recombinant gene construct comprising, in sequence, a promoter according to this invention, a coding region and a gene terminator.

The invention further comprises a recombinant plant genome containing the said construct.

Plant cells may be transformed with constructs of the invention according to a variety of known methods (Agrobacterium Ti plasmids, electroporation, microinjection, microprojectile bombardment, etc). The transformed cells may then be regenerated into whole plants in which the new nuclear material is stably incorporated into the genome. Both transformed monocot and dicot plants may be obtained in this way. The transformation and regeneration methods employed are not particularly germane to this invention and may simply be selected from those available from the literature.

Examples of genetically modified plants according to the present invention include tomatoes, fruits such as mangoes, peaches, apples, pears, strawberries, bananas and melons; and field crops such as maize (corn), sunflowers, sugarbeet, canola, and small grained cereals such as wheat, barley and rice, ornamental plants such carnations, petunias, roses, chrysanthemums etc.

Plants produced by the process of the invention may contain more than one recombinant construct. As well as one or more constructs containing the cinnamyl alcohol dehydrogenase promoter, they may contain a wide variety of other

recombinant constructs, for example constructs having different effects on plant development, structure and defense. In particular, constructs which affect lignin structure, composition and quality and quantity, cellulose and hemicellulose structure and amount, and plant defense genes are included in this invention.

Of particular interest is the ability of the CAD gene promoter to respond to exogenous stimuli. Thus, a further aspect of the present invention is a process of activating exogenous coding sequences in plants under the control of the CAD promoter which comprises the application of exogenous stimuli such as viruses, fungi and bacteria, as well as ethylene and other chemicals. This may find particular use when the expression of novel characters in the plant may need to be controlled exogenously.

We now describe the isolation of genomic clones from a tobacco genomic library encoding the cinnamyl alcohol dehydrogenase gene and related sequences. Genomic clones representing one of the closely related CAD genes found in tobacco have been isolated and characterised by DNA sequence analysis. The clone gTCAD9 represents the whole of the gene with exon sequence identical to the clone pTCAD14 with only two mismatches. The genomic clones described in the Examples cover all of the coding region and the complete transcriptional initiation region of the CAD gene. The subclone gNTCAD9 -6.4SH contains approximately 2800 bp of the promoter fragment of this gene.

The invention will be further described with reference to the following drawings, in which:

2136618

Figure 1 shows a restriction map of gNtCAD9 and a diagram of the structure of the CAD gene;

Figure 2 shows the nucleotide sequence of the CAD promoter and the CAD structural gene, most of which is contained in the 6.4 kb SalI -HindIII fragment of gNTCAD9 (SEQ ID: 1).

Figure 3 outlines a scheme for construction of the plant transformation vector pNtCAD9Prom-GUS1.

EXAMPLE 1

1.1 Isolation of CAD genes

A tobacco genomic library (*N.tabacum* cv NK 326) was purchased from Clontech. This library contains MboI partially digested genomic DNA cloned into lambda EMBL3. The library was screened with the pTCAD14 cDNA insert and positive phages were purified by four successive cycles of plaque purification. Four positive clones were isolated.

Restriction fragment mapping and DNA sequence analysis of these clones indicated that all 4 clones are overlapping clones of the CAD gene and are related. Only clones 2 and 9 contain promoter sequences and only clone 9 was characterised in detail. A restriction map of this clone is shown in Figure 1.

1.2 Characterisation of the CAD gene promoter sequence

A 6.4 kb SalI - HindIII fragment was isolated from gNTCAD9 and the nucleotide sequence covering 2765 bp of the 5′-flanking (promoter) sequences and most of the coding sequences has been determined (Figure 2). Primer extension experiments place the transcription start point at nucleotide position 2766.

Full sequence data of about 6.8kb have been

obtained of a tobacco CAD gene contained on clone gNtCAD9. The following are thereby confirmed:
(A) Transcription start site mapped to position 2766 from the upstream Sal 1 border, now designated
5 position + 1, placing the ATG start codon at + 110
(B) Position and exact size of the four introns:
(1) at + 198, size 149 bp; (11) at + 312, size 1290 bp;(111) at + 540, size 563 bp; (IV) at 980, size 632bp. The intron position and sizes differ from
10 the ones in ADH genes.
(C) Three nucleotide mismatches to cDNA cNtCAD14, which are all silent with respect to the deduced protein. They can be ascribed to cultivar differences between cDNA and genomic library.

15 1.3 Fusion of CAD promoter to the GUS reporter gene

A XbaI site at position +84 was used to isolate the CAD promoter from gNtCAD9 and create a transcriptional fusion of 2849 bp of promoter and
20 leader sequences to the GUS reporter gene in pBI 101.4. The promoter fragment was isolated after partial XbaI digest due to a second XbaI site in the promoter. Correct fusion and fusion borders were confirmed by restriction analysis and DNA
25 sequencing.

1.4 Construction of plant transformation vector - pCAD-gus

The plasmid subclone gNtCAD9-6.4SH containing the 6.4 kb SalI-HindIII fragment was digested with
30 SalI and partially digested with XbaI to recover a 2853 bp SalI-XbaI fragment, which contains all of the available CAD promoter sequence and 87 bp of leader sequence. This promoter fragment was inserted into the appropriately cut polylinker of

2136618
the promoterless GAS cassette of the binary plant transformation vector pBI101.4 in the 5'-3' orientation (Figure 3). The correct insertion has been confirmed by DNA sequencing of the fusion
5 borders.

EXAMPLE 2

Generation of transformed plants

The vector pNtCAD9Prom-GUS1 (from Example 1.4) is transferred to Agrobacterium tumefaciens LBA4404
10 (a micro-organism widely available to plant biotechnologists) and was used to transform tobacco plants. The fusion construct was introduced into Agrobacterium tumefaciens LBA 4404. Tobacco cv. SR1 leaf discs were transformed via cocultivation
15 with Agrobacteria harbouring the construct. Transgenic shoots were rooted and further propagated by standard procedures. Transformation of tobacco leaf disks follows established methods. Transformed plants were identified by their ability
20 to grow on media containing the antibiotic kanamycin. Plants were regenerated and grown to maturity.

The tissue-specific and developmental expression and expression in response to external
25 stimuli of the β-glucuronidase (GUS) gene as determined by the CAD gene promoter is demonstrated by analysis of stems, roots, leaves, seeds, flowers, pollen and in response to wounding and ethylene treatment for GUS enzyme activity.

30 EXAMPLE 3

Expression of CAD promoter - GUS fusion in transgenic tobacco

From 20 transformants regenerated 11 have been analysed in some detail. Quantitatively

(fluorometric analysis), GUS activity is highest in roots followed by stems and leaves. Histochemical staining by X-gluc indicates strong expression in the vascular system of stems and leaves
5 preferentially in the developing xylem, where lignification is known to occur. No activity was detectable in pith or epidermal tissue. However, leaf and stem hairs ((trichomes)) exhibit considerable GUS expression in certain stages of
10 their development. One example is the region around axillary buds, where strong staining of trichomes was observed. Upon wounding of stems a local induction of promoter activity was seen during the development of a wound plug but only
15 10-20 days after wounding.

Claims:

1. A DNA construct for use in transforming plant cells which comprises an
5 exogenous coding sequence under the control of upstream promoter and downstream terminator sequences, characterised in that the upstream promoter is a promoter of a cinnamyl alcohol dehydrogenase gene, which promoter has a nucleotide sequence comprising the first 2765 nucleotides of the sequence shown in Figure No. 2.

10 2. An isolated cinnamyl dehydrogenase gene promoter, having the nucleotide sequence comprising the first 2765 nucleotides of the sequence shown in Figure No.

3. A plant cell transformed with a DNA construct as claimed in claim 1.

15

20

25

30

Claims:

1. A DNA construct for use in transforming plant cells which comprises an
5 exogenous coding sequence under the control of upstream promoter and downstream terminator sequences, characterised in that the upstream promoter is a promoter of a cinnamyl alcohol dehydrogenase gene, which promoter has a nucleotide sequence comprising the first 2765 nucleotides of the sequence shown in Figure No. 2.

10 2. An isolated cinnamyl dehydrogenase gene promoter, having the nucleotide sequence comprising the first 2765 nucleotides of the sequence shown in Figure No.

3. A plant cell transformed with a DNA construct as claimed in claim 1.

15

20

25

30

1/8

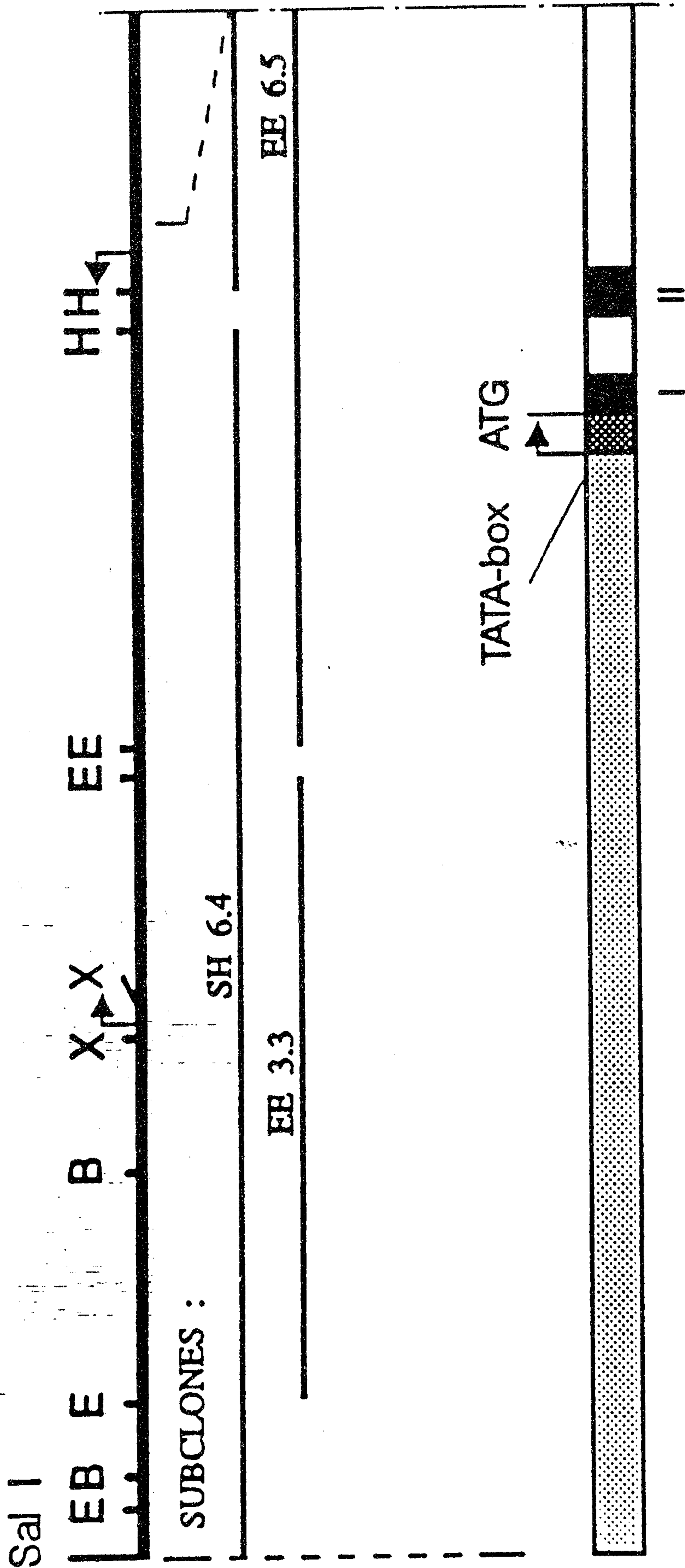


FIG. 1(1/2)
Tobacco Genomic CAD Clone gNtCAD9 (14 kb Insert)

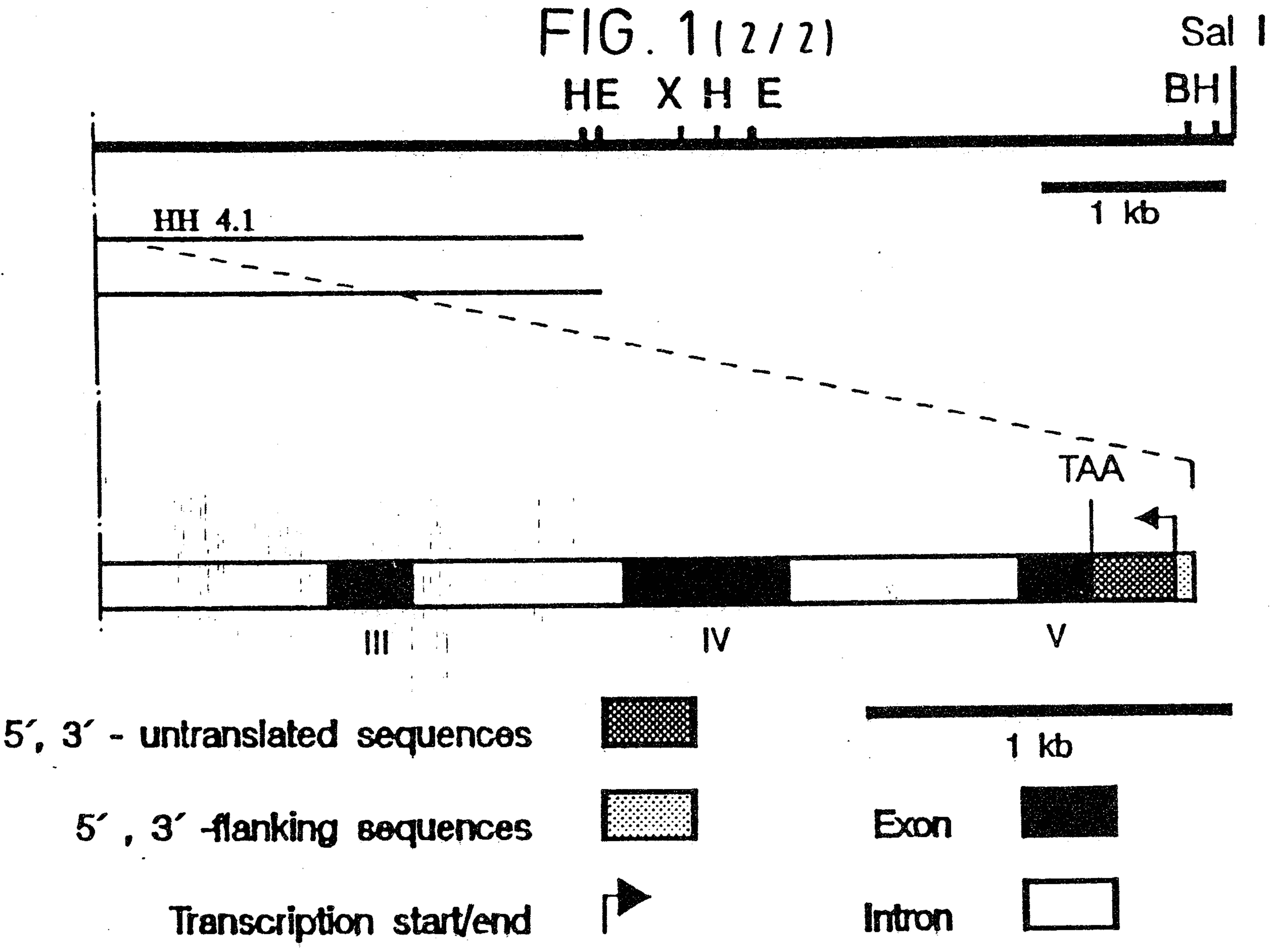


FIG. 1 (2/2)
HE X H E
Sal I
BH
1 kb
HH 4.1
TAA
III
IV
V
5', 3' - untranslated sequences
5', 3' -flanking sequences
Transcription start/end
1 kb
Exon
Intron

FIG. 2 (1/5)

SEQ ID NO: 1
SEQUENCE TYPE: Nucleotide
SEQUENCE LENGTH: 6877
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: genomic

ORIGINAL SOURCE ORGANISM: tobacco cv NK326
IMMEDIATE EXPERIMENTAL: tobacco genomic DNA library

FEATURES: from 1 - 2765 promoter
from 2875 - 6579 coding seqence and introns

PROPERTIES: genomic DNA of cinnamyl alcohol dehydrogenase - gNtCAD9

```
  1  GTCGACCTGC AGGTCAACGG ATCATATATT ATGTGTCTTT ACCTATACTA TTATCACTAA
 61  ATCAATAAAT GTCTAACTGC AATGATCGTG TTAAATTGTG TATCAGAGAA TTGAGTGACT
121  GATATATAGT AGGCATTTGA AAGTTGAAAG ATGAATTTCA CACTAAACTC AGATAAGATG
181  ATTTTGTGGG GTAGGCATTT GATTAAATGC AAAACCTACT CTCATGAATG TCAGTTTAAC
241  AACACAATCA ACCAAATGAA TTCTAAAAAC AAGAATTTTA CTGGGAAAAA AAACGTTTCG
301  AATAAATAAG AATATCGTAA AATTAATATT AACCAACAAA TTTGCATTGG TTATAACTTA
361  TAATCACGTG AAATGATTTG TACCAATACC TTATTTGGAT CCAAAGCCTG GCTAATTTGT
421  TTCTTACCTT CTTTTGCACA AGCAACCTCT TTTTAGCTTC TTCTACCTAT ATTACCTACT
```

FIG. 2 (2/5)

```
 481  AACTTTTAAG AGTCCAAAAC CTAAAATAAA TCAAAAGTTT TTTTTAAAAA AAAAAAAAAA
 541  AAAAAAAAAA CCTTACCTAA AACATGACGA TCTAGTGACG CATACTTCCA CCTAACCCCG
 601  GGGGACCCAA ACTGCTTGTC CAGAACAATT AATGCTTTCT TTTAGGAAGT TATCATTCAC
 661  ATAATTCATC TAGGCCTGTG ATAAAGATAA TTGGTCATTA TTGGAAAAGT CGTCTCACGT
 721  AGGAAGAGAT TGCTAAACTA GAATTTGAAA TTTGAGTCCT CTGAATTTGA CGTCGAATTC
 781  ACTGCATATT TTAATTACTG AGTTCATAAT TATATATTAG TATAAATTTA ATTGATTCTC
 841  TCGTATAAAT ATAGACTTTA TGCAAAAGCT ACGAGTTCGA TCAAACCCGT AAATAATATG
 901  TTACCTCCGT TTCCTAGTAG GAAGACAGAA ACAAATATCA AAAAGAAAGG AAAAAAACAT
 961  TTCACACTCT ACGTACCTTT CTATAATTTC AAAATAAAAG TTGGTGAAAG GGAAAAATGG
1021  TTAGAGGAAA AATAAGATAG GAGAGAACAG AAACTATGTA TTCGTAAGGA ATTTAAAATT
1081  AAAATATATT AATTTTGTTT TACAAGTTAT TTAAGAACTT TAGAAGACGA ATAACAACAA
1141  TGTTTTTTTT CTGTTTTAGT AATTTTCATA ACATATGAAC TAAATGCTCT TTTAAAAAGT
1201  TTTTAGGGAA AAGAAAAACG AAAAATTAAA GTGGTTGTAT CAAGAAAGCA AAACATATGA
1261  TCAGATTTTC CGGCATATTT ATAGGATTAT TAACGATGGT TTACTCAGTT ACAAGGTTCC
1321  CGAGAAATTA ATTTTGAAAA TATTAACAGA AAAAAGCCCA AATAAACACT TTTTTCTTTA
1381  AGATTTATTT ATTTAGGCCT CGTTAGTTTG CATGGAGGTT TGAATCTTAA TCATTCAGAT
1441  CTCAGATATT GACTGCGTTT CTTTTTAAGT ATAAATCTTA ATTATTCAGA TCTTAATCAT
1501  TAAATTTTTT TAAACCACTG AATGAGTCTG AAATCTGAAT AGGTCTTATT GATTACGATT
1561  TATAACAGAG ACTTAATTTT ATTAAGATGT TATCATCCAT ATTCACTACT AACCACCGCC
1621  ATTATCACCT ACCACCATGC ACCATCCAAC CAATACCAAA CTCAACCACC ACCAAATTCA
1681  AACACCATCA CTAACCACCA CTATCTCACT GCTATCATTA TCGACTACCA TCACCTCCAA
1741  CCCCACCATT ATCGATTACC ATCACCCACC ACCCCATAAT TATCAACTAA TACCACTACC
1801  ATCAATCACA AACATCTATC CACTCACCTC AACTACTACC ATCAACCCCC CCACCACCAT
1861  AATCAACCAT AGCTACCATT GTCCTCCACG CCACCATCAC CTTCCTCGAC CATAACTACT
1921  ACTACCCAAT CACCACTAGT TATTACTTGG AACCATCACC AACCGCCGCC TCCAGTCGGA
1981  TCCATAAGCC TCACCGTCAA CCATCACTAC ACCAACTATC ATATTATTTT TCAAAAATAT
2041  ATTTATTATT AATAGAATAT TAAATTAATT AATATTTTAT TTGAATTTTT TATTTATTAT
2101  TTTTAGATAA TGATAAATTT TATACATTCA GATGTTGAAA ATCAAACAGT CTTAATCATT
2161  TAGTATTCAT ATATTAATAT ATATCTTAAA TAATCAGATG TCTATTCTGA TTCAAACGTT
2221  TTAATCTTAA TACACATCTT AATATTTACA TGTCTCTTTA TCTTTAAATA AACAAATGAG
2281  GCCTTAGCTG TCTTCACATA ATTCTTTATT GCTGTGCTAT TAATGAGTAC GAATTAGTAA
2341  TTCTCTTAAC GAACAGTGAG TAGTTTGCTA ACTTTATATT CGTTTATTGT ACGTATATTG
```

FIG. 2 (3/5)

2401	TGTATTATAA	GAATAAAAAA	ATATTTAGCA	CTTATTTGGT	TAGTATTAAT	TTCCACGGAC
2461	ATTTTTAAAG	CATATTTTCC	AAGCTGTATC	TATTTTTTTT	TCCTTCCTCT	CATCTCAAGA
2521	TGCAAGTGTC	ATTTAATTTG	TACAACATGT	CGATTAATAT	AGTGATTACA	ATTATTACCT
2581	TACTGTTCAC	GTATAGGTTG	GAAATATAAT	ATAAGCATCT	ACTCAATCCA	ATTATAAAAC
2641	TTGATGTCTA	TATCTTAGCT	ACTGAGCAAA	AAATCAACTT	TTGTAAATTC	TAGAAGAAAC
2701	GCCACCCAAC	CTCTGTCTTC	CCCTTCTTTA	AACTATATAT	ATGTGGTGTT	ATTAACCAAA
2761	CTGTACTAAA	ATAGCTCACC	ACTATTCCTT	TCTCTTTCCC	TTGAACTGTG	TTTTCGTTTT
2821	TTCTGCTCTA	AAACAATAGT	GTGTTCCTTC	TAGATTTTAA	GTTTAAAGAA	CATCATGGGT
2881	GGCTTGGAAG	TTGAGAAAAC	AACTATTGGT	TGGGCTGCTA	GAGACCCTTC	TGGTGTACTT
2941	TCACCTTATA	CCTATACTCT	CAGGTACATT	ACATTACTAC	TTATGTTCTT	TCTATAATCG
3001	TAATATCCGA	ATCAGCTTAT	GCTTACCTCG	ACTACTTATT	TTTATCTTAT	AATAACTATT
3061	CTACCAATTT	TAAGATGAGT	TATGAACTTT	GTTTTTTGTT	TTTCATGTCC	AGAAACACAG
3121	GACCTGAAGA	TGTGGAAGTC	AAAGTTTTGT	ATTGTGGGCT	CTGCCACACT	GATCTTCACC
3181	AAGTTAAAAA	TGATCTTGGC	ATGTCCAACT	ACCCTCTGGT	TCCTGGGTAT	TTGCTCTTCC
3241	CTCTCTATTT	TCCCTGTTTT	TTCTCTCTCT	CTATATATAT	ATATAAAACA	TAATTTGAGA
3301	GTTCAAGTTT	TGTGGACAAA	ATGAGGATCA	CTATTGATCC	CCTAAGGATG	GAGGAAAATT
3361	GATTAATTAG	AAGAGAAACA	AGAGTTTGAC	AGTAATTTTT	TCAAGGATAA	TTTAACTTGT
3421	TTTAGCATAG	GTTGTTGCAG	TATTCTTGAG	ATTACGAATT	TCATTTATGA	TAAATACTTG
3481	TAAGTATATT	TTAAGTGATC	TGATAGCTAG	AAATTCTTTA	CACGCTTTTA	AATATATGAA
3541	TGTTAAAATT	CTCCTACAAT	GGTACCCTTT	GAATCTTTGT	TTGATGTCAA	TTTCAGTTTT
3601	TTCCTTAGCC	ACTGAATAAA	TCAAGAGAAA	TGTAGTGGTC	CTTGTGCCTT	TTAGTACCAA
3661	TCCACCATCC	TCGGTGACTT	AATAAATTTG	CAACTGTTAG	TTGTTAGGAT	TTGATTTATT
3721	ATTCCAAATG	TCTTTTTTTC	CTTCCTTTAT	TAAACGAATG	ATGAGTACTA	TTGAATTGTT
3781	TAATTAGGAG	TTGTGGCAGA	CCACGTATTT	TCGTGGCCTA	AATGCCAAAG	CAAACATATA
3841	ATAACGTACG	ACTAAGTTCA	CTTCTTTCAT	TGGATTTGTA	AAAGTTGTTG	GTGACACTTA

FIG. 2 (4/5)

```
3901  CGTTTTGTTC GAATCCAATG CTATTATTGT TAGATAAATT TTTACTTTTA GTCCCTAAAA
3961  TTATTGATAA GTTATAATTT TAATCCTTGT ACTATTTGGC GAAGTACATT TAGTCTTCAA
4021  TTTAATTAAA ATTAGACATT TGATCCTGCT TCTGCTTGTG AATCTTCACA AAGTTAATCA
4081  ATTATGAATT CATGAATTAT GTTAAATTTG TGTTGTTACT CCCTATCTAA GGATACTAGA
4141  GTTCTAATTC ATAAGTAGCG GTGGTAAAAT AGATAAAAAT AATAATTATC CATTCATTTT
4201  ATCCACTAAA AAATAAATAA CCAATGAATT TAACTTTTAT ATTTACAAAC CCTTAATTGA
4261  AGGGTTTCTC AAGTTTGAGA TACTAAGAAT TCTCTCAAAA GTAATCATAT TGAAGAAGTT
4321  ATGAATCATA TGTATATCCA TATTATCTGT CAATTAACTA TTTTTTATCC ATATTAAATA
4381  TAGATAAGGT TGAATATTAT ATTTATTAGA CGCCCATTTG TCACCCCTAT AATCACAATT
4441  GCAATATCCC TTTTGTTTTT TACTATGTGT TGATCCTTAT CCCTTTGTAA CTTCATATTT
4501  GTGGTGAATG ATGTAGACAT GAAGTGGTGG GAGAAGTGGT GGAGGTAGGA CCAGATGTGT
4561  CAAAATTCAA AGTGGGGGAC ACAGTTGGAG TTGGATTACT CGTTGGAAGT TGCAGGAACT
4621  GTGGCCCTTG CAAGAGAGAT ATAGAGCAAT ATTGCAACAA GAAGATTTGG AACTGCAATG
4681  ATGTCTACAC TGATGGCAAA CCCACCCAAG GTGGTTTTGC TAAATCCATG GTTGTTGATC
4741  AAAAGTAAGT CTTTTACCCT TCCATATATA TAGAGAAGTA CCTTTCTTTA ATATTAACTT
4801  ATAGCAAGTG ACATTGTATA TAAGTTACAC CCTGTAAATT AAAGAAACTA TCAGTATAAC
4861  TTACCTGTCG TAATAGGTTG GTTGCCTTAT TTTGAGGTAA TTAGTTTTAC TTATTGTGGA
4921  CAGAGCTACA ACTATTTAGA TAATAGGATG GTATAAAACA ATTTCTACAC TATCATTGTG
4981  TTTGTTCAAC TCTTTTTTAA ATCCATTGGG TATCGAATCA TGACAAATGT TACTAGGACG
5041  ATCAAGAAAT CATGTAGGGC AGCCCTGTAC ACTAAGCTCG CGCTATGCGG TCCGGGGAAA
5101  GTCGGACCAC AAAGGTATAT TGTATGTAGT CTAACCCTGC ATTTCTGCAG AGGTGCTCGA
5161  ACCCGTATCC TCTTTGGTCA CATGGTAATA ACTTAACCAG TTACGCCAAG GTTCCCCTTC
5221  AAGAAACCAT ATAATTAAAG TGGTTAATTC ACATTTCTTG AATTAAAAGA TCACAATGGA
5281  CTGAATGTTA TTTTGCCTTG TATATAGGTT TGTGGTGAAA ATTCCAGAGG GTATGGCACC
5341  AGAACAAGCA GCACCTCTAT TATGTGCTGG TATAACAGTA TACAGTCCAT TGAACCATTT
5401  TGGTTTCAAA CAGAGTGGAT TAAGAGGAGG AATTTTGGGA TTAGGAGGAG TGGGACACAT
5461  GGGAGTGAAA ATAGCAAAGG CAATGGGACA TCATGTTACT GTCATTAGTT CTTCAAATAA
```

```
5521  GAAGAGACAA GAGGCATTGG AACATCTTGG TGCAGATGAT TATCTTGTCA GTTCAGACAC
5581  TGATAAAATG CAAGAGGCTT CTGATTCACT TGACTATATT ATTGATACTG TCCCTGTTGG
5641  CCATCCTCTT GAACCTTATC TTTCTTTGCT TAAAATTGAT GGCAAACTTA TCTTGATGGG
5701  AGTTATCAAC ACCCCCTTGC AATTTATCTC CCCCATGGTT ATGCTCGGTA AGTCATTTAA
5761  CTTATTTACG CTGACAGTGT AACACGTTCG GACGAAATGC ATTATTACAA CAAGTTAAAA
5821  TTGGTTTCTA ACAATATATC GTGTTAGTAC TTGCTCTAAA AAGTCGATTG AAAATTTAAA
5881  CGTTGAATTT GTCTATATAT GAAAAAAGAT AGAACAAAAG GATTGGAAGA GTGGTCATTT
5941  TAGTCAATGA AAGCCATGAT TTGTTGAAAT ATGCAGGTGC CATTGGTCTA TTTTTGCGGC
6001  ATAATAATGT TATGAGCAGA ATTAATATAT AGACAAATTC TAGGCTTGTA GGGTTGTGTT
6061  ATTTAAGGAA CCATGTCTGC AATAATGGAC TGCAATAACA ACAGAGACCA TACACGTATT
6121  AAAGCCAATG TATTTGATTA ACATTTGTTT AATGCAACCT CTTATTACTA TTATGTTTTA
6181  AGATTTTTTA AGCACAGACA GTGTGTAAGT TTTTTTATAC CATCAAGCAA AGTTGATATA
6241  TTCCGAAAAG AAAAGTTGAT GTACAATAAT AGATAACATT TCATTAGTGT GGTATGGTAA
6301  TTTGAAAGTA GTACTCGGTA TAATTTTTCA TTTTACATTG TCGGCGTATA TAACTTTTAG
6361  ACTTTATTAT TATGATCAGG GAGAAAGAGC ATCACAGGAA GCTTTATTGG TAGCATGAAG
6421  GAAACAGAGG AAATGCTAGA TTTCTGCAAA GAGAAAGGTG TGACTTCACA GATTGAGATA
6481  GTGAAAATGG ATTATATCAA CACTGCAATG GAGAGGTTGG AGAAAAATGA TGTGAGGTAC
6541  AGATTTGTGG TTGATGTTAT TGGAAGCAAG CTTGACCAGT AATTATATTA CACAAGAAAA
6601  ACAACATGGA ATGGTTCACT ATTATACAAG GCTGTGAGAA TACTAAACTT TGATGTCGTC
6661  TTTTGTATCC TTTTGTTTTA TTTGCCACCT GTATTTTCTT ATTTGGTGAT CGAGAGTGAC
6721  GTTTATGTAT TATTTTCTTT CTTCAAAACA ATTTAATGTA TGAATTTGGA TGTTGGTGAC
6781  GATTTTGAAA TATACCAACC AAAACTTTGT TTGGTACGTG AAGCTATCTC CACTTCTCGT
6841  GACTAGGGTG TCGATCCATC TTTAAATCGA TTAACCG  6877
```

FIG. 2 (5/5)

2136618

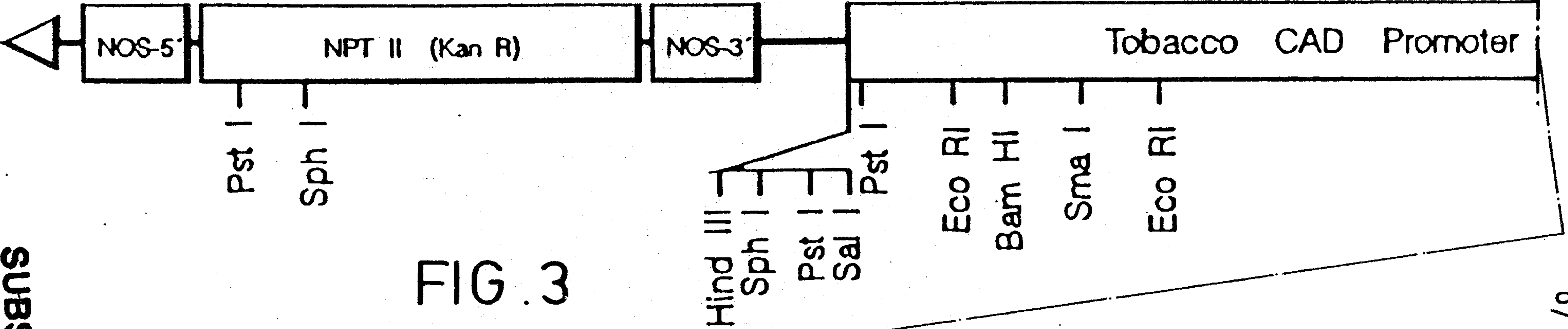
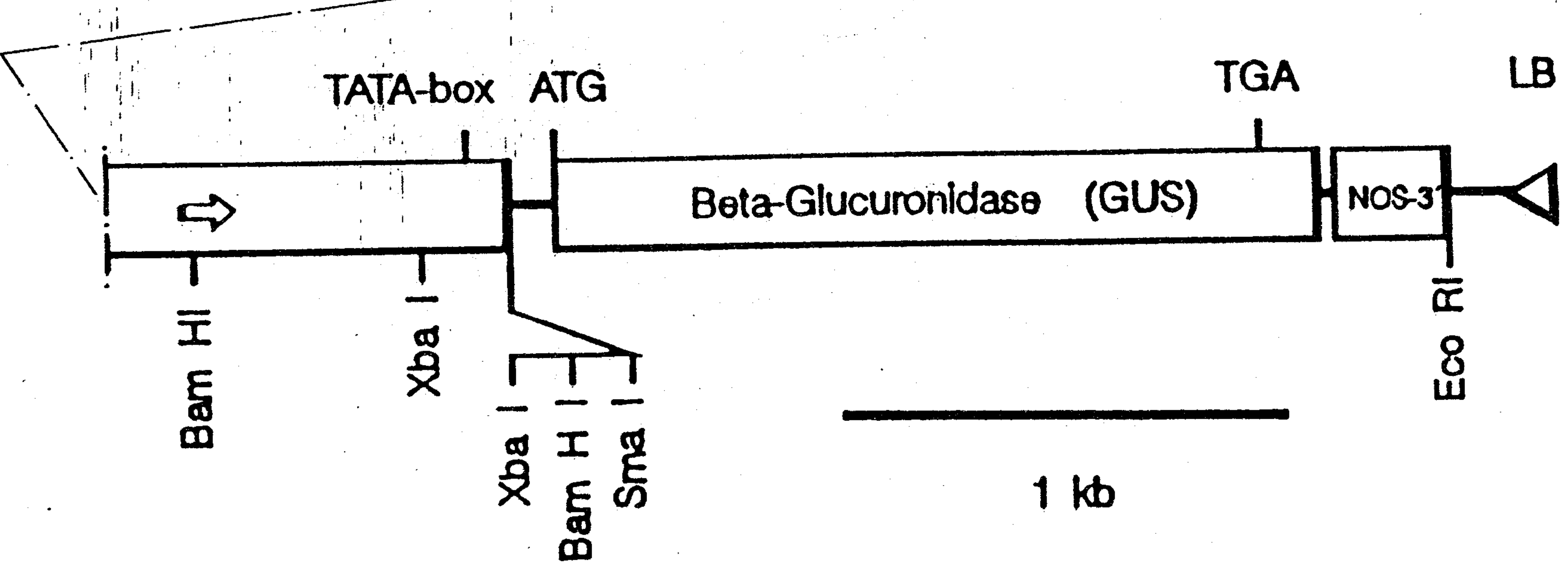


FIG. 3



SUBSTITUTE SHEET