



(12) PATENTSKRIFT

Patent- og
Varemærkestyrelsen

(51) Int.Cl.: C 12 N 15/56 C 12 N 9/24 C 12 P 21/02

(21) Patentansøgning nr: PA 1989 03746

(22) Indleveringsdag: 1989-07-28

(24) Løbedag: 1989-07-28

(41) Alm. tilgængelig: 1990-01-30

(45) Patentets meddelelse bkg. den: 2004-11-29

(30) Prioritet: 1988-07-29 US 226303 1989-05-17 US 353312

(73) Patenthaver: Syngenta Participations AG, Schwarzwaldallee 215, CH-4058 Basel, Schweiz

(72) Opfinder: Frederick Meins, Leimgrubenweg 52, 4125 Riehen, Schweiz
Hideaki Shinshi, Ottominami 2-9-24, Tsuchinra City, Ibaraki 300, Japan
Herman C. Wenzler, 2300 Longmire, Apt. C, College Station, Texas 77840-5201, USA
Jan Hofsteenge, Mischelistrasse 15, 4153 Reinach, Schweiz
John Ryals, 415 Dunhill Circle, Durham, NC 27713, USA
Christoph Sperisen, Landskronstr. 4, 4056 Basel, Schweiz

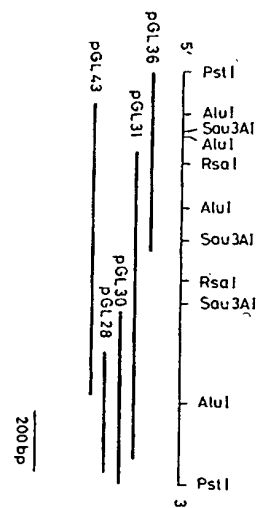
(74) Fuldmægtig: Budde, Schou & Ostenfeld A/S, Vester Søgade 10, 1601 København V, Danmark

(54) Benævnelse: Rekombinant DNA-sekvens, som koder for et polypeptid, som har beta-1,3-glucanase-aktivitet, og signalpeptid, som den rekombinante DNA-sekvens koder for

(56) Fremdragne publikationer:
Ingen

(57) Sammendrag:

Rekombinant-DNA-sekvenser indeholdende cDNA-sekvenser eller genomiske DNA-sekvenser, som koder for polypeptider med β -1,3-glucanase-aktivitet, beskrives. Endvidere beskrives fremgangsmåder til isolering og karakterisering af DNA-sekvenserne.



Den foreliggende opfindelse angår rekombinante DNA-sekvenser, som koder for et polypeptid, der har β -1,3-glucanaseaktivitet. Endvidere angår opfindelsen et signalpeptid for tobaks-
5 β -1,3-glucanase.

Glucan-endo-1,3- β -glucosidaser (β -1,3-glucanase) er enzymer, som indgår i en række biologiske processer, herunder planters forsvarsreaktion mod patogener (1). De induceres som reaktion på stresshormonet ethylen (2, 3),
10 infektion (4-10) og udløserstoffer ("elicitors") (9, 10), og har antifungal virkning in vitro (12). En basisk β -1,3-glucanase fra tobak med M_r ~33.000 er en væsentlig komponent (~5-10%) af det opløselige protein i dyrket tobaksvæv og de nedre blade og rødder på tobaksplanter
15 (13, 14). Denne β -1,3-glucanase hæmmer kompleks hormonal og udviklingsmæssig regulering. Den induceres i tobaksblade behandlet med ethylen (3) eller inficeret med tobaksmosaikvirus (TMV) (8, 10) og nedreguleres på mRNA-niveauet i dyrkede tobaksplanteceller ved kombinationer af
20 plantehormoner auxin og cytokinin (3, 14-16).

Glucanaser er enzymer, som hydrolyserer polysaccharider. De er derfor vigtige komponenter i nedbrydningen af polysaccharidmaterialer fra planter. De er som sådan nyttige til biomasseomdannelsesprocesser (kompostering).
25 Under visse omstændigheder kan denne aktivitet også udstrækkes til papirfremstillingsprocesser.

Identifikationen af DNA-sekvenserne, som koder for disse enzymer, vil gøre det muligt at foretage ændringer i planter indeholdende disse gener til kontrollering af
30 ekspressionen af de tilsvarende proteiner i disse planter. Endvidere ville fjernelse af disse DNA-sekvenser fra deres naturlige genomer gøre det muligt at transformere andre organismer med disse DNA-sekvenser og dermed overføre fordelene ved sådanne enzymer til andre
35 organismer end dem, hvori enzymerne forekommer naturligt.

Desuden kunne man kontrollere timingen og omfanget af
ekspressionen af disse enzymer og derved optimere disse
transformerede planter værdi. Det ville være særligt
5 interessant at transformere en planteart med gener, som
koder for glucanase, afledt af en anden planteart.

Den foreliggende opfindelse angår rekombinante DNA-sekven-
ser, som koder for polypeptider, der har β -1,3-glucanase-
aktivitet. Disse rekombinante DNA-sekvenser kan være cDNA-
10 sekvenser eller de kan være genomiske DNA-sekvenser udvundet
i praktisk taget ren form. DNA-sekvenserne omfatter såvel
de sekvenser, som koder for præ-pro-enzymet, som de, der
koder for det modne enzym, der dannes naturligt som følge
af en posttranslationel behandling.

15 Endvidere omfatter opfindelsen også DNA-sekvenser, som
udviser væsentlig sekvenshomologi med sådanne rekombi-
nant DNA-sekvenser eller genomiske DNA-sekvenser.

I en foretrukken udførelsesform koder DNA-sekvenserne for
 β -1,3-glucanaser, som forekommer naturligt i planter.
20 Særlig interesse har glucanaserne fra tobaksplanter.

Opfindelsen forklares nærmere i det følgende under henvis-
ning til tegningen, hvor

fig. 1 er et partielt restriktionskort, som viser længder-
ne og positionerne for cDNA-klonerne af tobak- β -1,3-glucanase,
25 nase, som er sekvenseret,

fig. 2 viser Southern blot-analyse af DNA fremstillet ud
fra tobaksblade, *Nicotiana tabacum* L.cv. Havana 425 (TA),
N. sylvestris (SY) og *N. tomentosiformis* (TO) spaltet med
den anførte restriktionsendonuclease og hybridiseret med
30 insertet tobak- β -1,3-glucanase-cDNA-plasmid pGL43, hvor
skalaen til venstre angiver størrelsen af markør DNA'er i
kilobaser,

fig. 3 viser et autoradiogram af polypeptider fra tobaks-protoplaster inkuberet i 45 minutter med ^{35}S -methionin (PULSE) og derefter ca. 12 timer med 1 mM methionin (CHASE), hvor immunadsorption gennemføres med anti-tobak- β -1,3-glucanase; R: produkter opnået ved in vitro-translation af total RNA fra celler induceret til at producere β -1,3-glucanase, T: protoplaster behandlet med tunicamycin, skala til venstre: M_r for proteinstandarder, GLU: position for rensed β -1,3-glucanase,

fig. 4-11 viser sekvenser, hvor etbogstavssymboler anvendes til betegnelse af nucleotider og aminosyrer; (N) og (C): N- og C-terminale aminosyrer i moden β -1,3-glucanase, (N-glyc): formodet N-glycosyleringssted, og formodede polyadenyleringssignaler er understreget, idet

fig. 4 viser sekvens I, som er sekvens DNA-I svarende til den sammensatte cDNA-sekvens fra kloner pGL28, pGL30 og pGL31, hvis sekvenser er identiske; under sekvensen er vist den udledte aminosyresekvens AA-I,

fig. 5 viser sekvens II, som er sekvens DNA-II svarende til den sammensatte cDNA-sekvens afledt af klonen betegnet pGL43; under denne sekvens er vist den udledte aminosyresekvens AA-II,

fig. 6 viser sekvens III, som er sekvens DNA-III svarende til den sammensatte cDNA-sekvens afledt af klonen med betegnelsen pGL36; under denne sekvens er vist den udledte aminosyresekvens AA-III,

fig. 7 viser sekvens IV, som er sekvens DNA-IV svarende til den sammensatte cDNA-sekvens afledt af sekvens DNA-I og hidrørende fra peptid-sekvensering vist som aminosyresekvens AA-IV anbragt under sekvensen DNA-IV; XXX indicerer forskelle i indkodede aminosyrer af sekvens DNA-I og dem, som findes ved peptidsekvensering,

fig. 8 viser sekvens V, som svarer til det genomiske DNA isoleret i klon pBSGluc39.1,

fig. 9 viser sekvens VI svarende til det genomiske DNA som
5 isoleret i klon pBSGluc39.3,

fig. 10 viser sekvens VII, hvor sekvens DNA-VII svarer til cDNA-sekvensen udledt af det isolerede genomiske DNA i klon pBSGluc39.1, sekvens V; under denne sekvens er vist den afledte aminosyresekvens AA-VII, og

10 fig. 11 viser sekvens VIII med sekvens DNA-VIII svarende til cDNA-sekvensen udledt af det isolerede genomiske DNA i klon pBSGluc39.3, sekvens VI, og anbragt under denne sekvens er vist den afledte aminosyresekvens AA-VIII.

Deponeringer

15 Plasmid pBSGluc39.1 er deponeret ved American Type Culture Collection, 29. december 1988, under deponeringsnummeret ATCC 40526, og

plasmid pBSGluc39.3 er deponeret ved American Type Culture Collection, 29. december 1988, under deponeringsnummeret
20 ATCC 40527.

Den foreliggende opfindelse angår en rekombinant DNA-sekvens indeholdende en DNA-sekvens, som koder for et polypeptid med β -1,3-glucanase-aktivitet, eller en DNA-sekvens med væsentlig sekvenshomologi med sådan en β -1,3-glucanase-kodende sekvens. Dette rekombinant-DNA kan indeholde
25 et molekylært DNA-fragment fra genomisk DNA, som koder for et polypeptid med β -1,3-glucanase-aktivitet, eller en DNA-sekvens med væsentlig sekvenshomologi til sådant et fragment. Endvidere kan dette rekombinant-DNA være et cDNA
30 fremstillet ud fra mRNA kodende for et polypeptid med

β -1,3-glucanase-aktivitet eller ud fra et fragment af sådant mRNA, et DNA med væsentlig sekvenshomologi med sådant et cDNA eller et rekombinant DNA indeholdende
5 sådant et cDNA eller et DNA med væsentlig sekvenshomologi med sådant et cDNA. Den foreliggende opfindelse angår også en genomisk DNA-sekvens, som koder for et polypeptid med β -1,3-glucanase-aktivitet, i stort set ren form.

Sædvanligvis vil kildeorganismen for det genomiske DNA
10 eller for mRNA'et, som koder for et polypeptid med β -1,3-glucanase-aktivitet, være en plante. Anvendelige planter er enkimbladede og tokimbladede planter, f.eks. majs, hvede, ris, byg, tobak, tomat, agurk, kartoffel, almindelig bønne, sojabønne, ærter eller gulerødder. Den fore-
15 trukne plante er tobaksplanten.

Det fremgår af det efterfølgende, at i tobak koder den genomiske DNA-sekvens for et såkaldt "præ-pro-enzym", som undergår posttranslationel behandling eller processing til dannelse af den modne β -1,3-glucanase. Det vil sige, at
20 den C-terminale ende og den N-terminale ende af præ-pro-enzymet afspaltes efter translation. Disse spaltningssteder er vist over nucleotidsekvenserne i sekvenser I-IV som (C) for det C-terminale spaltningspunkt og som (N) for det N-terminale spaltningspunkt. Den foreliggende
25 opfindelse omfatter følgelig de DNA-sekvenser, som koder for præ-pro-enzymet, samt de DNA-sekvenser, som koder for det modne enzym. De sidstnævnte DNA-sekvenser kan opnås ved anvendelse af konventionelle restriktions- og spaltningsprocesser, enten ud fra det genomiske DNA eller ud
30 fra cDNA'et, som koder for præ-pro-enzymet. Opfindelsen omfatter også DNA-sekvenser med væsentlig sekvenshomologi til de naturligt forekommende genomiske DNA-sekvenser og til cDNA-sekvenser afledt af naturligt forekommende mRNA-sekvenser.

35 Der foretrækkes en rekombinant DNA-sekvens indeholdende en

cDNA-sekvens, som koder for en moden β -1,3-glucanase, eller et fragment deraf indeholdende 5'-enden af kodningssekvenserne, især et fragment, som har mindst 700 nucleotider eller mindst 900 nucleotider. Specielt foretrækkes en rekombinant DNA-sekvens indeholdende en cDNA-sekvens, som koder for en præ-pro- β -1,3-glucanase, eller et fragment deraf, indeholdende 5'-enden af kodningssekvenserne, især et fragment med mindst 800 nucleotider eller mindst 1000 nucleotider. Helt specielt foretrækkes DNA-sekvenser, som indeholder cDNA, der koder for moden tobaks- β -1,3-glucanase eller tobaks-præ-pro- β -1,3-glucanase, f.eks. et rekombinant DNA, som indeholder en DNA-sekvens som vist i sekvens DNA-I, sekvens DNA-II, sekvens DNA-III, sekvens DNA-IV, sekvens DNA-VII eller sekvens DNA-VIII. Desuden foretrækkes især stort set rent genomisk DNA, som koder for tobaks- β -1,3-glucanase, f.eks. stort set ren genomisk DNA gluc39.1 (sekvens V) eller genomisk DNA gluc39.3 (sekvens VI) samt rekombinant DNA, som indeholder det genomiske DNA.

Endvidere foretrækkes en rekombinant DNA-sekvens, som koder for et polypeptid med formlen

```

      Q S I G V C Y G M L G N N L P N H W
E V I Q L Y K S R N I G R L R L Y D P N
H G A L Q A L K G S N I E V M L G L P N
S D V K H I A S G M E H A R W W V Q K N
V K D F W P D V K I K Y I A V G N E I S
P V T G T S Y L T S F L T P A M V N I Y
K A I G E A G L G N N I K V S T S V D M
T L I G N S Y P P S Q G S F R N D A R W
F V/T D P I V G F L R D T R A P L L V N I
Y P Y F S Y S G N P G Q I S L P Y S L F
T A P N V V V Q D G S R Q Y R N L F D A
M L D S V Y A A L E R S G G A S V G I V
V S E S G W P S A G A F G A T Y D N A A
T Y L R N/Y L I Q/P H A K E G S P R K P G P
I E T Y I F A M F D E N N K N P E L E K

```

H F G L F S P N K Q P K Y N I/L N F G

især med formlen

										M	S	T	S	H	K	H	N	T	P
Q	M	A	A	I	T	L	L	G	L	L	L	V	A	S	S/T	I	D/E	I	A
G	A	Q	S	I	G	V	C	Y	G	M	L	G	N	N	L	P	N	H	W
E	V	I	Q	L	Y	K	S	R	N	I	G	R	L	R	L	Y	D	P	N
H	G	A	L	Q	A	L	K	G	S	N	I	E	V	M	L	G	L	P	N
S	D	V	K	H	I	A	S	G	M	E	H	A	R	W	W	V	Q	K	N
V	K	D	F	W	P	D	V	K	I	K	Y	I	A	V	G	N	E	I	S
P	V	T	G	T	S	Y	L	T	S	F	L	T	P	A	M	V	N	I	Y
K	A	I	G	E	A	G	L	G	N	N	I	K	V	S	T	S	V	D	M
T	L	I	G	N	S	Y	P	P	S	Q	G	S	F	R	N	D	A	R	W
F	V/T	D	P	I	V	G	F	L	R	D	T	R	A	P	L	L	V	N	I
Y	P	Y	F	S	Y	S	G	N	P	G	Q	I	S	L	P	Y	S	L	F
T	A	P	N	V	V	V	Q	D	G	S	R	Q	Y	R	N	L	F	D	A
M	L	D	S	V	Y	A	A	L	E	R	S	G	G	A	S	V	G	I	V
V	S	E	S	G	W	P	S	A	G	A	F	G	A	T	Y	D	N	A	A
T	Y	L	R	N/Y	L	I	Q/P	H	A	K	E	G	S	P	R	K	P	G	P
I	E	T	Y	I	F	A	M	F	D	E	N	N	K	N	P	E	L	E	K
H	F	G	L	F	S	P	N	K	Q	P	K	Y	N	I/L	N	F	G	V	S
G	G	V	W	D	S	S	V	E	T	N	A	T	A	S	L	V/I	S	E	M

- 5 hvori bogstaverne afskilt med en skråstreg repræsenterer den ene eller den anden aminosyre. Disse aminosyresekvenser svarer til henholdsvis præ-pro-enzymet og det modne enzym og er udledt af aminosyresekvenserne AA-I til IV, VII og VIII.
- 10 Endvidere omfatter opfindelsen signalpeptidsekvenserne, som er karakteristiske for β -1,3-glucanaserne. Signalpeptider er aminosyrekomponenter i translaterede proteiner og muliggør transport af sådanne proteiner fra syntese-
- 15 ført, afspaltes signalpeptidet, og det modne protein aktiveres eller gøres tilgængeligt for yderligere

processing. Opfindelsen angår især et signalpeptid for tobaks- β -1,3-glucanase og rekombinant-DNA-sekvenser, som koder for et sådant signalpeptid, f.eks. peptidet med
5 formelen

M S T S H K H N T P Q M A A I T L L G L
L L V A S S/T I D/E I A G A

hvor i bogstaverne adskilt med en skråstreg repræsenterer den ene eller den anden aminosyre, og en rekombinant DNA-sekvens, som koder for dette peptid.

- 10 Bestemte DNA-sekvenser er fremstillet og karakteriseret som cDNA eller genomiske sekvenser i overensstemmelse med fremgangsmåderne beskrevet i de efterfølgende eksempler. Resultaterne af karakteriseringerne angives i det følgende.
- 15 Der isoleres tre forskellige klasser cDNA-kloner, hvis kodningssekvenser ligner hinanden. Desuden er der heterogenitet i aminosyresekvensen, som ikke er repræsenteret i mRNA'erne svarende til de isolerede cDNA-kloner. Dyrkede tobaksplanteceller induceret til at producere β -1,3-glucanase har således mindst fire transkriptionelt aktive β -1,3-glucanase-gener. Dette og det lille antal restriktionsfragmenter, som påvises i Southern blot-forsøgene, indikerer, at tobak- β -1,3-glucanase indkodes af en lille genfamilie. Moderne tobaksplanter er opstået ved hybridisering af to ancestrale arter, nemlig *N. sylvestris* og *N. tomentosiformis* (32, 33). Sammenligning af restriktionsfragmenterne dannet af DNA fra tobak og de ancestrale arter tyder på, at ca. 2-3 medlemmer af tobaks- β -1,3-glucanase-familien har deres udviklingsmæssige baggrund i begge disse stamarter.
25
30

Computerundersøgelser i EMBL-version 14 (38) og NBRF-version 15 (39) databankerne giver kun et protein med signi-

fikant sekvenslighed med tobak- β -1,3-glucanase. Efter indføring af 3 huller eller spring i det modne tobakenzyms sekvens er aminosyresekvensen for bygendosperm (1 $\rightarrow$ 3, 1 $\rightarrow$ 4)-
5 β -glucanase-isozym II (40) identisk med hensyn til 47% af positionerne. Desuden er der stærkt bevarede områder (12 af 15 positioner) ved N- og C-terminalerne i de modne proteiner. Sekvensen for de første 22 N-terminale aminosyrer hos en formodet β -1,3-glucanase oprenset fra N.
10 plumbaginifolia (41) er identisk med den N-terminale sekvens hos tobaksenzymet, bortset fra den N-terminale aminosyre, som er Glu.

Forsøg med impulsundersøgelser ("pulse chase") viser, at præ-pro- β -1,3-glucanase undergår sekvensmæssig processing.
15 Først fjernes det N-terminale signalpeptid, og en oligosaccharidsidekæde adderes som rapporteret for frøoplageringsproteinerne phaseolin og concanavalin A (42, 43). Senere fjernes oligosaccharidet og den C-terminale forlængelse til dannelsen af det modne enzym. Tab af C-terminale
20 forlængelser er rapporteret for planteproteinerne pro-thaumatocin (44), pro-concanavalin A (43,45,46) og isolectin 3 i hvedekimagglutinin (47). Der konstateres ingen væsentlige sekvensligheder mellem disse peptider og den C-terminale forlængelse af tobaksenzymet.

25 cDNA-Sekvensen vist som sekvenser DNA-VII er udledt af de genomiske DNA-sekvenser for klon pBSGluc39.1, sekvens V. Nucleotider 1 til 121 svarer til nucleotider 1434 til 1554, og nucleotider 122 til 1407 svarer til nucleotider 2345 til 3630 i den genomiske sekvens. Nucleotider 1555
30 til 2344 i genomisk sekvens V repræsenterer en intron. cDNA-Sekvensen vist som sekvenser DNA-VIII er udledt af de genomiske DNA-sekvenser for klon pBSGluc39.3, sekvens VI. Nucleotider 1-121 svarer til nucleotider 1618 til 1738, og nucleotider 122-1431 svarer til nucleotider 2486 til 3795.
35 Nucleotider 1739 til 2487 i den genomiske sekvens VI repræsenterer et intron.

Opfindelsen forklares nærmere ved hjælp af de efterfølgende eksempler.

Eksempel 1

5 Plantemateriale og dyrkningsmetoder

Nicotiana tabacum L. cv. Havana 425, N. sylvestris og N. tomentosiformis dyrkes fra frøstadiet i et væksthuse. Den klonede linie af Havana 425 marvvæv, 275N (17) dyrkes under betingelser, som inducerer β -1,3-glucanase-akkumulering (15). Suspensioner af bladprotoplaster fremstilles ud fra axeniske Havana 425-skudkulturer og dyrkes i flydende K3-medium som beskrevet af Paszkowski et al. (18).

Eksempel 2

15 Konstruktion og screening af et cDNA-bibliotek

Et cDNA-bibliotek fremstilles af poly(A⁺) RNA isoleret fra 275N-væv (15) som beskrevet (19) og indsat i Pst I-stedet i plasmid pBR322 ved anvendelse af den homopolymere dC-dG-tailingmetode (20). Rekombinantkolonier screenes ved hjælp af kolonihybridisering med ³²P-mærket Pst I-insert af tobaks- β -1,3-glucanase cDNA-klon pGL43, en partiel cDNA-klon for tobak- β -1,3-glucanase (15) som en hybridiseringsprobe. Selv om β -1,3-glucanase-mRNA er en stærkt repræsenteret komponent af poly(A⁺) RNA'et anvendt til fremstilling af cDNA-bibliotekerne (15), kan kun et lille antal partielle cDNA-kloner identificeres i screeninger af flere biblioteker.

Eksempel 3Deoxynucleotidsekvensanalyse

Restriktionsfragmenter kloneres i M13-vektorerne mp18 og
5 mp19 (21) og sekvenseres i begge orienteringer ved hjælp
af dideoxynucleotid-kædetermineringsmetoden (22). Nucleo-
tidsekvenser analyseres under anvendelse af computerpro-
grammet ifølge Queen og Korn (23).

Fem af de syv positive rekombinantkloner, som viser sig at
10 have de længste cDNA-inserter, sekvenseres (fig. 1). Tre
af klonerne indeholder ikke 3'-områderne af mRNA'et. Den
sammensatte sekvens viser en enkelt, uafbrudt åben aflæs-
ningsstruktur på 1083 nucleotider begyndende ved position
1. Baseret på tilgængelig information for disse syv kloner
15 foretages følgende bestemmelse af translationsstarten:
Consensussekvensen for translationsstarten i planter er
AACAAUGGC, hvori purinet ved -3 er vigtigt af hensyn til
effektiv translation i et reticulocytlysatsystem, men
ikke i et hvedekimsystem (28). Det første tilgængelige
20 initi-eringskodon i strukturen er ved position 7, som
giver sekvensen UCAAAUGCG. Translation startende med dette
kodon danner et polypeptid med en længde på 359 rester
med M_r 39.173 svarende i størrelse til det primære
translations- produkt af tobaks- β -1,3-glucanase (15, 29).

25 De overlappende sekvenser af inserterne af kloner pGL28,
pGL30 og pGL31 er identiske, hvilket tyder på, at de re-
præsenterer det samme mRNA (sekvens DNA-I). Sammenligning
af inserterne i pGL31 og pGL36, som overlapper med en
længde på 328 nucleotider, viser to stille substitutioner
30 ved positioner 555 og 564 og to substitutioner ved posi-
tioner 547 og 548, som resulterer i en enkelt aminosyre-
substitution af Val med Thr. Sekvensen af pGL43 insertet
er forskellig fra sekvenserne for pGL31 og pGL36 inser-
terne ved 13 positioner; 12 er stille forskelle i den

tredje såkaldte "wobble" position i kodon'et, og én i position 1006 resulterer i den bevarende ændring fra Leu til Ile, som også påvises ved aminosyresekvensering af det
5 modne enzym.

Den sammensatte cDNA-sekvens har to formodede polyadenyleringssignaler for planter, AAG AAA i position 1285 og ATT AAT i position 1321 (30). De overlappende sekvenser af cDNA inserterne fra pGL28 og pGL30 er identiske op til
10 position 1310, hvor poly(A) halen af pGL28 starter 18 nucleotider efter det første polyadenyleringssignal. Poly(A) halen af pGL30 starter 24 nucleotider efter det andet polyadenyleringssignal. Dette tyder på, at der er en skiftende polyadenylering af β -1,3-glucanase-transkriptionsterne som rapporteret for andre plantegener (31).
15

Eksempel 4

Southern Blot-analyse

5 μ g DNA fremstillet af tobaksblade (*N. tabacum*), *N. sylvestris* og *N. tomentosiformis* (32, 33) ved anvendelse af fremgangsmåden ifølge Murray og Thompson (24) spaltes
20 fuldstændigt med restriktionsendonucleaseerne EcoRI, HindIII eller XbaI. De dannede fragmenter adskilles ved elektroforese på en 0,8% (v/r) agarosegel, overføres til en Zeta-probe "Nylon"-membran (BioRad) og
25 hybridiseres som anbefalet af producenten. Den afsluttende vask gennemføres ved 50°C i 0,2 x SSC (1 x SSC er 0,15 M NaCl/15 mM natriumcitrat, pH 7), 0,1% (v/r) natriumdodecylsulfat (SDS). Pst I insertet af tobaks- β -1,3-glucanase cDNA-klonen pGL43 (15) anvendes som probe og mærkes
30 under anvendelse af et random primer-mærkningssæt (Boehringer).

Restriktionsendonucleasefragmenterne, som hybridiserer til insertet fra tobaks- β -1,3-glucanase-cDNA-klon pGL43, er

vist i fig. 2. Hver af restriktionsendonucleaseerne danner fire fragmenter med tobak DNA. EcoRI- og XbaI-fragmenterne svarer med hensyn til størrelse og hybridiseringsintensitet til to fragmenter dannet med DNA fra hvert af forældrearterne, *N. sylvestris* og *N. tomentosiformis*. HindIII danner fragmenter med *N. sylvestris* DNA, som ikke med hensyn til størrelse svarer til fragmenterne dannet med tobak DNA. Analyser under anvendelse af kombinationer af restriktionsendonucleaseer og andre hybridiseringsprober tyder på, at forskelle i hybridiseringsintensiteten viser forskelle med hensyn til homologi og ikke med hensyn til kopiantal.

Eksempel 5

15 Analyse af proteiner

β -1,3-Glucanase renses fra 275N væv ved hjælp af "CM-Sephadex"® G-50-kromatograferingstrinet for chitinase (16). Fraktioner indeholdende β -1,3-glucanase hældes sammen, koncentrerer ved ultrafiltrering (nedsænkkelig CX-10, Millipore), indstilles til 100 mM NaCl/10 mM Tris-HCl, pH 8,4 og ledes gennem en søjle indeholdende regenereret chitin (25) til fjernelse af spor af chitinase. Slutproduktet har en renhed på >95% bestemt ved SDS-PAGE og farvning af proteinet med coomassie-blåt R-250. Til fremstilling af tryptiske peptider indstilles β -1,3-glucanase (320 μ g i 1 ml) på 8 M urinstof og dialyseres mod 2 M urinstof/0,1 M NH_4HCO_3 . Der tilsættes trypsin (7 μ g, Worthington), og blandingen inkuberes i 6,5 timer ved 37°C. Derefter tilsættes yderligere 7 μ g trypsin, og reaktionen afbrydes efter 24 timer ved indstilling af pH-værdien på 3 med HCl. CNBr-spaltning, fraktionering af peptider med hjælp af HPLC, aminosyreanalyse og aminosyresekvensering gennemføres som beskrevet (26).

Sekvensen for 4 tryptiske peptider og et CNBr-peptid af det rene enzym dækkende i alt 70 aminosyrerester er den samme som de udledte sekvenser, hvilket tyder på, at de
5 klonede cDNA'er er afledt af mRNA'et for β -1,3-glucanase. Yderligere aminosyrer findes i to positioner i proteinet, 284 og 287, og disse er ikke repræsenteret i de isolerede cDNA-kloner.

Aminosyre- og DNA-sekvensinformationen anvendes til at
10 udlede primærstrukturen for præ-pro- β -1,3-glucanase. Under anvendelse af reglerne for proteinprocessing af Von Heijne (34) identificeres et yderst sandsynligt spaltningssted mellem Ala 21 og Gln 22. Processing ved dette sted stemmer overens med iagttagelsen af, at den N-terminale ende af
15 modent β -1,3-glucanase er blokeret. N-terminale glutaminer blokeres ofte ved spontan eller enzymatisk omdannelse til en pyroglutamylrest (35). Et rensat N-terminalt blokeret peptid dannet ved CNBr-spaltning af det modne enzym med sammensætningen Glu_{0,6}Ser_{1,0}Ile_{1,0}Gly_{2,0}Val_{1,1}Tyr_{1,0}Hse_{1,0}
20 (Hse er homoserin) svarer til den N-terminale sekvens Gln Ser Ile Gly Val Cys Tyr Gly Met. Disse resultater viser, at det primære translationsprodukt for β -1,3-glucanase har et hydrofobt signalpeptid med en længde på 21 rester, og at den N-terminale ende af det modne enzym er Gln 22.

25 Det primære translationsprodukt af β -1,3-glucanase har en M_r -værdi, som er ca. 4000 større end værdien for det modne enzym (15, 29). Tilstedeværelsen af et signalpeptid kan kun dække for en forskel i M_r -værdi på ca. 2000, hvilket tyder på, at proteinet undergår yderligere processing.
30 Spaltning af enzymet med CNBr giver et peptid, som mangler homoserin, hvilket indicerer, at det er den C-terminale ende af det modne protein. Dets aminosyresammensætning og sekvens bestemmes, og resultaterne viser sig at svare til rester 308 til 337 af de udledte sekvenser. Da stop-
35 kodon'et forekommer 22 aminosyrerester efter Gly 337, tyder det på, at en C-terminal forlængelse er fjernet

under processing.

Eksempel 6

Impuls-opsporingsforsøg

- 5 5×10^4 bladprotoplaster inkuberes i 45 minutter ved 26°C i mørke under omrystning ved 50 o/m i hullerne på en 24-hullers Falcon vævskulturplade (Becton Dickinson), som indeholder 500 μl K3-medium uden fytohormoner og ca. 20 μCi ^{35}S -L-methionin (specifik aktivitet ca. 1275 Ci/nmol, 10 Amersham). Hvor det indiceres efterfølges impulsmærkningen ved inkubation i 11,75 timer med 1 mM L-methionin sat til hullerne. Til et kopisæt af huller sættes 20 $\mu\text{g}/\text{ml}$ tunicamycin (Calbiochem) 20 minutter inden tilsætningen af ^{35}S -L-methioninet. Proteinekstrakter fremstilles ved 15 tilsætning af 100 μl 6 mM dithiothreitol, 6% (v/v) "Triton" X-100, 200 mM Tris-HCl, pH 8,0 til hvert hul, hvorpå der inkuberes i 10 minutter ved 26°C , hvorefter prøverne fryses i flydende nitrogen. Ekstrakterne opdeles i to alikvoter. Proteinet i en alikvot fældes med acetone 20 og analyseres ved hjælp af SDS-PAGE (17). Den anden alikvot immunadsorberes med kanin-anti-tobak- β -1,3-glucanase IgG (27) bundet til protein A-"Sephadex"® CL-4B (Pharmacia, og de immunadsorberede polypeptider analyseres ved hjælp af SDS-PAGE (15). ^{35}S -Methionin-mærkede in vitro 25 translationsprodukter fremstilles med et kanin-reticulocytsystem. Total RNA fra 275N væv induceret til at producere β -1,3-glucanase er kilden for mRNA (15). Geler kalibreres med SDS-PAGE-standarder med lav molekylvægt (BioRad).
- 30 Processingen af β -1,3-glucanase udledt af strukturundersøgelser (eksempel 5) bekræftes ved impuls-opsporingsforsøg. Hovedpolypeptidet mærket ved impulsmærkning af tobaksprotoplaster med ^{35}S -methionin i 45 minutter immunadsorberes ved hjælp af anti- β -1,3-glucanase-anti-

- stof (fig. 3). Det har en M_r -værdi, som er lidt større end M_r -værdien for in vitro translationsproduktet anvendt ved den samme SDS-PAGE som en standard for præ-pro-enzymet.
- 5 Efter opsporingen comigrerer størsteparten af det immunreaktive materiale med den modne form af β -1,3-glucanase, hvilket viser, at størsteparten af precursorformen af β -1,3-glucanase undergår processing til den modne form under den ca. 12 timer lange opsporing (chase).
- 10 Den udledte sekvens for precursoren har et enkelt formodet sted for N-glycosylering, Asn X Ser/thr (36), som starter ved Asn 350 i den C-terminale forlængelse. Dette tyder på, at processing af β -1,3-glucanase omfatter glycosylering. Til afprøvning af denne formodning inkuberes kopiproto-
- 15 plastsuspensioner anvendt i impuls-opsporings-forsøgene med tunicamycin, som hæmmer syntesen af lipidbundne oligosaccharider, og som er kendt for at blokere N-glycosylering af plantepolypeptider (37). Tunicamycinbehandling hæmmer ^{35}S -methionininkorporering i polypeptider med ca.
- 20 50%. Det immunreaktive polypeptid opnået efter impulsmærkning i nærværelse af tunicamycin har en M_r -værdi, som er ca. 2000 mindre end værdien for polypeptidet opnået med ubehandlede protoplaster. På den anden side har det immunreaktive polypeptid opnået efter opsporingsbehand-
- 25 lingen i nærværelse af tunicamycin samme M_r -værdi som modent β -1,3-glucanase. Plante-N-glycaner (37) og den N-terminale signalsekvens af β -1,3-glucanase har samme M_r -værdi på ca. 2000. Resultaterne indicerer derfor, at precursoren mister signalpeptidet og får tilført en
- 30 oligosaccharidsidekæde. Dette sker på stedet mærket (N-glyc) i sekvenserne. Senere deglycosyleres det dannede proenzym, og den C-terminale forlængelse fraspaltes til dannelse af det modne enzym. Processing til den modne form kræver imidlertid ikke glycosylering.

Eksempel 7Isolering af en genomisk klon

DNA med høj molekylvægt fremstilles ud fra blade af
5 *Nicotiana tabacum* cv. Havana 425 ved anvendelse af CETAB-
metoden (24). 100 µg DNA spaltes fuldstændigt med *SacI*.
Dette DNA fraskilles ved elektroforese på en 0,8%
agarosegel. Det område af gelen, som svarer til et
molekyl- vægtsområde på 3-7 kb, udskæres i 6 lige store
10 strimler, og DNA'et elektroeluieres fra disse strimler og
fældes som separate fraktioner. DNA'et resuspenderes, og
en alikvot fra hver fraktion behandles på en agarosegel og
analyseres ved hjælp af Southern blotting. Fraktionen, som
indeholder DNA, der hybridiserer til β -1,3-glucanase
15 cDNA-proben (eksempel 4), hældes sammen og anvendes til
konstruktion af biblioteker.

Kloningsvektoren λ OngC fra firmaet Stratagene Corp.
spaltes med *SacI*. 1 µg af dette DNA blandes med 0,1 µg af
det *SacI*-spaltede tobak DNA, og DNA'et liggeres med T4-DNA-
20 ligase under anvendelse af producentens forskrifter. Det
ligerede DNA pakkes derpå i fagpartikler under anvendelse
af en in vitro-pakningsmetode ifølge Stratagene. Fagerne
udplades med bakterier som foreslået i λ -håndbogen. Ca.
75.000 plaque screenes under anvendelse af en 32-P-mærket
25 β -1,3-glucanase cDNA-probe. 11 positive plaque identifi-
ceres. Disse plaque renses ved flere ganges udpladning og
screening.

DNA isoleres fra de rensede kloner, og klonerne analyseres
ved restriktionsspaltning under anvendelse af *HindIII*,
30 *EcoRI* og *SacI*. Der er to typer kloner, hvor den ene er
repræsenteret ved klonen 39.1, og den anden er repræsente-
ret ved klonen 39.3. 4,5 og 4,7 kb inserterne i henholds-
vis klon 39.1 og 39.3 underklones i bluescript plasmidet
spaltet med *SacI*, og underklonerne pBSGluc39.1 (*SacI*-

fragment afledt af λ -klon 39.1) og pBSGluc39.3 (SacI-fragment afledt af λ -klon 39.3) isoleres. DNA-sekvensen i SacI-fragmenterne indeholdt i underklonerne pBSGluc39.1 og pBSGluc39.3 bestemmes ved DNA-sekvensering og er vist i henholdsvis sekvens V og VI. Kodningssekvensen bestemmes, og den tilsvarende aminosyresekvens udledes. Disse sekvenser er vist i henholdsvis sekvens VII og sekvens VIII. En stor mellemliggende sekvens findes nær 5'-enden af kodningssekvensen.

Plasmid pBSGluc39.1 er deponeret ved American Type Culture Collection, 29. december 1988, under deponeringsnummeret ATCC 40526, og

plasmid pBSGluc39.3 er deponeret ved American Type Culture Collection, 29. december 1988, under deponeringsnummeret ATCC 40527.

Eksempel 8

Identificering af transkriptionsstartstedet

A. Primerforlængelseskortlægning

Der fremstilles en syntetisk DNA-primer, som er komplementær til baserne 2399 til 2416 (sekvens V), under anvendelse et synteseapparat fra Applied Biosystems og β -cyanoethylphosphoramidatmetoden. Oligonucleotidet renses ved højtryksvæskechromatografi med omvendt fase (HPLC). 5 pmol af oligoforbindelsen kinasebehandles (20) under anvendelse af 200 μ Ci ^{32} P-ATP (6000 Ci/mmol, 10 μ Ci/ μ l). Efter inkubering ved 37°C i 30 minutter fortyndes reaktionsblandingen til 100 μ l, hvorpå den ekstraheres med phenol/chloro-form, hvorefter der fældes tre gange med 50 μ g bærer-RNA. Bundfaldet fra sidste fældning resuspenderes i 1 X reverse-transkriptase-puffer (50 mM Tris-HCl, pH 7,5, 40 mM KCl, 3 mM MgCl₂) ved en koncentration på 2 nM.

Den specifikke aktivitet af det mærkede oligonucleotid bestemmes til ca. 3×10^6 Cvpmp/pmol.

Total-RNA isoleres fra *Phytophthora parasitica* var. nicotiana-inficeret tobak i alt væsentligt som beskrevet (48). Væv findeles under flydende nitrogen i en morter med en pistil. Det formalede væv sættes til formalingspuffer (48), idet der anvendes 2,5 ml pr. g væv. Et lige så stort rumfang phenol tilsættes derpå, og emulsionen homogeniseres i en Brinkman-polytron. Der tilsættes et halvt rumfang chloroform, hvorpå emulsionen blandes forsigtigt i 15 minutter. Faserne adskilles ved centrifugering, og den vandige fase fjernes. RNA fældes ved tilsætning af natriumacetat til 0,3 M og 2,5 rumfang ethanol. Bundfaldet samles ved centrifugering og resuspenderes i 2 ml sterilt vand. Der tilsættes lithiumchlorid med en slutkoncentration på 3 M, hvorpå blandingen henstår natten ved 4°C. Bundfaldet samles ved centrifugering, og pelletten vaskes med iskold 80%'s ethanol. Pelletten tørres og resuspenderes i 500 µl sterilt vand. Koncentrationen af denne total-RNA-præparation bestemmes spektrofotometrisk.

50 µg total-RNA lyofiliseres i et 500 µl Eppendorf-rør. RNA'et resuspenderes i 30 µl radioaktivt mærket probeopløsning, og der opvarmes til 70°C i 10 minutter. Røret afkøles langsomt til 37°C, og der inkuberes natten over, uden at fjerne røret fra 37°C-vandbadet tilsættes 2 µl 10 X revers-transkriptase-puffer (500 mM Tris-HCl, pH 7,5, 400 mM KCl, 30 mM MgCl₂), 1 µl 5 mg/ml bovint serumalbumin, 5 µl 100 mM dithiothreitol, 5 µl 10 X dNTP'er (10 mM af hver dNTP i H₂O), 3 µl H₂O, 2 µl RNasin (80 enh.) og 2 µl revers-transkriptase (400 enh.), og reaktionsblandingen inkuberes ved 37°C i 30 minutter. Reaktionen afbrydes ved tilsætning af 5 µl 3 M natriumacetat, pH 5, og 150 µl absolut ethanol. Røret henstår ved -20°C i 30 minutter, bundfaldet isoleres ved centrifugering, vaskes med 80%'s ethanol og henstilles til lufttørring. Bundfaldet resus-

penderes i 10-20 μ l påsætningsfarve (90% formamid, 0,05% bromphenolblåt, 0,05% xylencyanol, 1 mM EDTA), og forlængelsesprodukterne adskilles på en 6%^s sekvenseringsgel (20). Forlængelsesprodukterne gøres synlige ved autoradiografi.

Primerforlængelsesprodukterne køres sammen med en molekyl-
vægtsstandard, hvori den mærkede primer anvendes i dide-
oxy-DNA-sekvenseringsreaktioner med subklonen pBSGluc39.1
10 som skabelon. Der identificeres forlængelsesprodukter, som
svarer til positioner 1432, 1446 og 1447 i β -1,3-gluca-
nase-39.1-sekvensen, sekvens V. Da der findes et stort
intron mellem positioner 1554 og 2345 i pBSGluc39.1-
sekvensen, afspejler molekylvægtsskalaen ikke den korrekte
15 molekylvægt for forlængelsesprodukterne. Der gennemføres
derfor et andet primerforlængelseskortlægningsforsøg under
anvendelse af en primer, som er komplementær til positio-
ner 1530-1547. Under anvendelse af denne primer kortlægges
tre 5'-ender af glucanase-mRNA'et til adeninrester i
20 positioner 1434, 1448 og 1449.

B. S1-Nucleasekortlægning

Der fremstilles et syntetisk oligonucleotid, som er komplementært til positioner 1415-1504, til anvendelse som en probe i S1-nucleasekortlægning. Oligonucleotidet kinase-
25 behandles ved 5'-enden under anvendelse af ^{32}P -ATP og T4-polynucleotidkinase i overensstemmelse med leverandørens forskrifter. Efter phenolekstraktion og fældning med ethanol resuspenderes det mærkede oligonucleotid i formamidhybridiseringspuffer (49) ved en koncentration på ca.
30 2 nM. 50 μg lyofiliseret, total-RNA fra *Phytophthora parasitica*-inficeret tobak som ovenfor opløses i 30 μl af probeopløsningen, og rørene opvarmes til 65°C i 10 minutter, hvorefter de henstår til hybridisering natten over ved 48°C . S1-Nucleasebehandling og gelelektroforese
35 gennemføres i alt væsentligt som beskrevet under

anvendelse af en S1-koncentration på 400 enh./ml og en inkubationstemperatur på 30°C. Den passende S1-nuclease-koncentration bestemmes ved forforsøg. Efter gelelektroforese påvises tre bånd, som svarer til positioner 1434, 1448 og 1449 i pBSGluc39.1-DNA-sekvensen.

C. Bestemmelse af transkriptionsstartstedet

Primerforlængelse og S1-nucleasekortlægning bestemmer i begge tilfælde 5'-enderne i mRNA'et til positioner 1434, 1448 og 1449. Disse steder, som alle er adeninrester, svarer derfor til transkriptionsstartstedet for β -1,3-glucanase-genet.

D. Bestemmelse af translationsstartstedet

Ved at sammenligne de sammensatte cDNA-sekvenser med sekvensen af den ene eller den anden genomiske klon er det klart, at der findes et methionininitiator-kodon i strukturen (in-frame) anbragt 32 aminosyrer foran aminoterminalen glutamin i det modne protein. Det skønnes, at dette kodon repræsenterer translationsstartstedet, da det er det første methioninkodon efter transkriptionsstartstedet, og da det dannede signalpeptid har den egnede længde (25-32 aminosyrer) og indeholder en anden methionin 11 rester efter den første, hvilket er almindeligt i signalpeptider.

Litteraturhenvisninger

1. Boller, T. (1987) in Plant-Microbe Interactions, Molecular and Genetic Perspectives, eds. Kosuga, T. & Nester, E.W. (Macmillan, New York), Vol 2, pp. 385-413.
2. Abeles, F.B., Bosshart, R.P., Forrence, L.E. & Habig, W.H. (1971) Plant Physiol. 47, 129-134.
3. Felix, G. & Meins, F., Jr. (1987) Planta 172, 386-392.
4. Moore, A.E. & Stone, B.A. (1972) Virology 50, 791-798.
5. Pegg, G.F. & Young, D.H. (1981) Physiol. Plant Pathol. 19, 371-382.
6. Kearney, C.M. & Wu, J.H. (1984) Can. J. Bot. 62, 1984-1988.
7. Sanada, M., Matsushita, K., Shimokawa, H. & Itoh, R. (1986) Ann. Phytopath. Soc. Japan 52, 320-329.
8. Kauffmann, S., Legrand, M., Geoffroy, P. & Fritig, B. (1987) EMBO J. 6, 3209-3212.
9. Kombrink, E., Schroeder, M. & Hahlbrock, K. (1988) Proc. Natl. Acad. Sci. USA 85, 782-786.
10. Voegeli-Lange, R., Hansen-Gehri, A., Boller, T. & Meins, F., Jr. (1988) Plant Science, 54, 171-176.
11. Mauch, F., Hadwiger, L.A. & Boller, T. (1984) Plant

Physiol. 76, 607-611.

12. Mauch, F. (1985) Ph.D. Dissertation, Univ. of Basel, Basel.
13. Shinshi, H. & Kato, K. (1983) Agric. Biol. Chem. 47, 1455-1460.
14. Felix, G. & Meins, F., Jr. (1986) Planta, 167, 206-211.
15. Mohnen, D., Shinshi, H., Felix, G. & Meins, F., Jr. (1985) EMBO J. 4, 1631-1635.
16. Shinshi, H., Mohnen, D. & Meins, F., Jr. (1987) Proc. Natl. Acad. Sci. USA 84, 89-93.
17. Eichholz, R., Harper, J., Felix, G. & Meins, F., Jr. (1983) Planta 158, 410-415.
18. Paszkowski, J., Shillito, R.D., Saul, M., Mandák, V., Hohn, T., Hohn, B. & Potrykus, I. (1984) EMBO J. 3, 2717-1722.
19. Gubler, U. & Hoffman, B.J. (1980) Gene 25, 263-269.
20. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
21. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Gene, 33, 103-109.
22. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
23. Queen, C. & Korn, L.J. (1984) Nucleic Acid Res. 12,

581-599.

24. Murray, M.G. & Thompson, W.F. (1980) Nucleic Acid Res. 8, 4321-4325.
25. Boller, T., Gehri, A., Mauch, F. & Voegeli, U. (1983) Planta, 157, 22-31.
26. Hofsteenge, J. & Stone, S.R. (1987) Eur. J. Biochem. 160, 49-56.
27. Felix, G. & Meins, F., Jr. (1985) Planta, 164, 423-428.
28. Luetcke, H.A., Chow, K.C., Moss, K.A., Kern, H.F. & Scheele, G.A. (1987) EMBO J. 6, 43-48.
29. Shinshi, H. & Kato, K. (1983) Agric. Biol. Chem. 47, 1275-1280.
30. Dean, D., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. & Bedbrook, J. (1986) Nucleic Acids Res. 14, 2229-2240.
31. Heidecker, G. & Messing, J. (1986) Ann. Rev. Plant Physiol. 37, 439-466.
32. Gray, J.C., Kung, S.D., Wildman, S.G. & Sheen, S.J. (1974) Nature, 252, 226-227.
33. Jamet, E., Durr, A. & Fleck, J. (1987) Gene, 59, 213-221.
34. Von Heijne, G. (1983) Eur. J. Biochem. 133, 17-21.
35. Abraham, G.N. & Podel, D.N. (1981) Mol. Cell. Biochem. 38, 181-190.

36. Kornfeld, R. & Kornfeld, S. (1985) Ann. Rev. Biochem. 54, 631-664.
37. Chrispeels, M.J. (1984) Philos. Trans. R. Soc. Lond. Ser. B, 304, 309-322.
38. Hamm, G.H. & Cameron, G.N. (1986) Nucl. Acid Res. 14, 5-10.
39. George, D.G., Barker, W.C. & Hunt, L.T. (1986) Nucl. Acid Res. 14, 11-16.
40. Fincher, G.B., Lock, P.A., Morgan, M.M., Lingelbach, K., Wettenhall, R.E.H., Mercer, J.F.B., Brandt, A. & Thomsen, K.K. (1986) Proc. Natl. Acad. Sci. USA, 83, 2081-2085.
41. Bauw, G., De Loose, M., Inzé, D., Van Montague, M. & Vandekerckhove, J. (1987) Proc. Natl. Acad. Sci USA, 84, 4806-4810.
42. Bollini, R., Vitale, A. & Chrispeels, M.J. (1983) J. Cell. Biol. 96, 999-1007.
43. Herman, E.M., Shannon, L.M. & Chrispeels, M.J. (1985) Planta, 165, 23-29.
44. Edens, L., Heslinga, L., Klok, R., Ledeboer, A.M., Maat, J., Toonen, M.Y., Visser, C. & Verrips, C.T. (1982) Gene, 18, 1-12.
45. Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. & Burgess, J. (1986) J. Cell Biol. 102, 1284-1297.
46. Chrispeels, M.J., Hartl, P.M., Sturm, A. & Faye, L. (1986) J. Biol. Chem. 261, 10021-10024.

47. Raikhel, N.V. & Wilkins, T.A. (1987) Proc. Natl. Acad. Sci. USA, 84, 6745-6749.
48. Lagramini, L.M. et al. (1987) Proc. Natl. Acad. Sci. USA, 84, 7542.
49. Berk, A.J. et al. (1977) Cell, 12, 721.

P A T E N T K R A V

1. Rekombinant DNA-sekvens, kendetegnet ved, at den indeholder en DNA-sekvens, som koder for et polypeptid med
5 β -1,3-glucanase-aktivitet, og som har væsentlig sekvens-homologi med en DNA-sekvens valgt fra gruppen bestående af sekvens DNA-I, sekvens DNA-III, sekvens DNA-IV, sekvens DNA-V, sekvens VI, sekvens DNA-VII og sekvens DNA-VIII.
- 10 2. Rekombinant DNA-sekvens i følge krav 1, kendetegnet ved, at DNA-sekvensen, som koder for et polypeptid med β -1,3-glucanase-aktivitet, er et molekylfragment fra genomisk DNA.
- 15 3. Rekombinant DNA-sekvens ifølge krav 1, kendetegnet ved, at DNA-sekvensen, som koder for et polypeptid med β -1,3-glucanase-aktivitet, er et cDNA fremstillet ud fra mRNA eller ud fra et mRNA-fragment, eller en
20 DNA-sekvens med væsentlig sekvenshomologi med et sådant cDNA.
4. Rekombinant DNA-sekvens ifølge krav 1, kendetegnet ved, at den er et cDNA fremstillet ud fra mRNA eller ud fra et mRNA-fragment eller en DNA-sekvens med væsentlig
25 sekvenshomologi med et sådant cDNA.
5. Rekombinant DNA-sekvens ifølge krav 1, kendetegnet ved, at den koder for tobak- β -1,3-glucanase.
- 30 6. Renset genomisk DNA-sekvens, kendetegnet ved, at den koder for et polypeptid med β -1,3-glucanase-aktivitet, med væsentlig sekvenshomologi med en DNA-sekvens valgt fra gruppen bestående af sekvens V og sekvens VI.
- 35 7. Renset genomisk DNA-sekvens ifølge krav 6, kendetegnet ved, at β -1,3-glucanasen er tobak- β -1,3-glucanase.

8. Rekombinant DNA-sekvens ifølge krav 1, kendeteg-
net ved, at den indeholder en cDNA-sekvens, som koder for
moden β -1,3-glucanase.
- 5 9. Rekombinant DNA-sekvens ifølge krav 1, kendeteg-
net ved, at den indeholder et 5'-endekodningssekvensfrag-
ment af en cDNA-sekvens, som koder for moden β -1,3-glucanase.
- 10 10. Rekombinant DNA-sekvens ifølge krav 9, kendeteg-
net ved, at kodningssekvensfragmentet indeholder mindst
700 nucleotider.
- 15 11. Rekombinant DNA-sekvens ifølge krav 1, kendeteg-
net ved, at den indeholder en cDNA-sekvens, som koder for
præ-pro- β -1,3-glucanase.
- 20 12. Rekombinant DNA-sekvens ifølge krav 1, kendeteg-
net ved, at den indeholder et 5'-endekodningssekvensfrag-
ment af en cDNA-sekvens, som koder for præ-pro- β -1,3-
glucanase.
- 25 13. Rekombinant DNA-sekvens ifølge krav 12, kendeteg-
net ved, at kodningssekvensfragmentet indeholder mindst
800 nucleotider.
- 30 14. Rekombinant DNA-sekvens ifølge krav 5, kendeteg-
net ved, at den indeholder en cDNA-sekvens, som koder for
moden tobak- β -1,3-glucanase.
- 35 15. Rekombinant DNA-sekvens ifølge krav 5, kendeteg-
net ved, at den indeholder en cDNA-sekvens, som koder for
tobak-præ-pro- β -1,3-glucanase.

16. Rekombinant DNA-sekvens ifølge krav 1, kendetegnet ved, at den indeholder en DNA-sekvens valgt blandt sekvens DNA-I, sekvens DNA-III og sekvens DNA-IV.

5 17. Genomisk DNA-sekvens ifølge krav 7, kendetegnet ved, at den er DNA'et fra sekvens V.

18. Genomisk DNA-sekvens ifølge krav 7, kendetegnet ved, at den er DNA'et fra sekvens VI.

10

19. Rekombinant DNA-sekvens ifølge krav 5, kendetegnet ved, at den koder for et polypeptid med formlen

		Q	S	I	G	V	C	Y	G	M	L	G	N	N	L	P	N	H	W	
15	E	V	I	Q	L	Y	K	S	R	N	I	G	R	L	R	L	Y	D	P	N
	H	G	A	L	Q	A	L	K	G	S	N	I	E	V	M	L	G	L	P	N
	S	D	V	K	H	I	A	S	G	M	E	H	A	R	W	W	V	Q	K	N
	V	K	D	F	W	P	D	V	K	I	K	Y	I	A	V	G	N	E	I	S
	P	V	T	G	T	S	Y	L	T	S	F	L	T	P	A	M	V	N	I	Y
20	K	A	I	G	E	A	G	L	G	N	N	I	K	V	S	T	S	V	D	M
	T	L	I	G	N	S	Y	P	P	S	Q	G	S	F	R	N	D	A	R	W
	F	V/T	D	P	I	V	G	F	L	R	D	T	R	A	P	L	L	V	N	I
	Y	P	Y	F	S	Y	S	G	N	P	G	Q	I	S	L	P	Y	S	L	F
	T	A	P	N	V	V	V	Q	D	G	S	R	Q	Y	R	N	L	F	D	A
25	M	L	D	S	V	Y	A	A	L	E	R	S	G	G	A	S	V	G	I	V
	V	S	E	S	G	W	P	S	A	G	A	F	G	A	T	Y	D	N	A	A
	T	Y	L	R	N/Y	L	I	Q/P	H	A	K	E	G	S	P	R	K	P	G	P
	I	E	T	Y	I	F	A	M	F	D	E	N	N	K	N	P	E	L	E	K
	H	F	G	L	F	S	P	N	K	Q	P	K	Y	N	I/L	N	F	G		

30

hvor i bogstaver adskilt ved hjælp af en skråstreg repræsenterer den ene af de to aminosyrer.

35

20. Rekombinant DNA-sekvens ifølge krav 5, kendetegnet ved, at den koder for et polypeptid med formlen

										M	S	T	S	H	K	H	N	T	P
	Q	M	A	A	I	T	L	L	G	L	L	V	A	S	S/T	I	D/E	I	A
5	G	A	Q	S	I	G	V	C	Y	G	M	L	G	N	N	L	P	N	H
	E	V	I	Q	L	Y	K	S	R	N	I	G	R	L	R	L	Y	D	P
	H	G	A	L	Q	A	L	K	G	S	N	I	E	V	M	L	G	L	P
	S	D	V	K	H	I	A	S	G	M	E	H	A	R	W	W	V	Q	K
	V	K	D	F	W	P	D	V	K	I	K	Y	I	A	V	G	N	E	I
10	P	V	T	G	T	S	Y	L	T	S	F	L	T	P	A	M	V	N	I
	K	A	I	G	E	A	G	L	G	N	N	I	K	V	S	T	S	V	D
	T	L	I	G	N	S	Y	P	P	S	Q	G	S	F	R	N	D	A	R
	F	V/T	D	P	I	V	G	F	L	R	D	T	R	A	P	L	L	V	N
	Y	P	Y	F	S	Y	S	G	N	P	G	Q	I	S	L	P	Y	S	L
15	T	A	P	N	V	V	V	Q	D	G	S	R	Q	Y	R	N	L	F	D
	M	L	D	S	V	Y	A	A	L	E	R	S	G	G	A	S	V	G	I
	V	S	E	S	G	W	P	S	A	G	A	F	G	A	T	Y	D	N	A
	T	Y	L	R	N/Y	L	I	Q/P	H	A	K	E	G	S	P	R	K	P	G
	I	E	T	Y	I	F	A	M	F	D	E	N	N	K	N	P	E	L	E
20	H	F	G	L	F	S	P	N	K	Q	P	K	Y	N	I/L	N	F	G	V
	G	G	V	W	D	S	S	V	E	T	N	A	T	A	S	L	V/I	S	E

hvor i bogstaver adskilt med en skråstreg repræsenterer den ene af de to aminosyrer.

21. Signalpeptid, kendetegnet ved, at det er et signalpeptid for tobak- β -1,3-glucanase med formlen

	M	S	T	S	H	K	H	N	T	P	Q	M	A	A	I	T	L	L	G
30	L	L	V	A	S	S/T	I	D/E	I	A	G	A							

hvor i bogstaver adskilt med en skråstreg repræsenterer den ene af de to aminosyrer.

22. Rekombinant DNA-sekvens, kendetegnet ved, at den koder for et peptid med formelen

M S T S H K H N T P Q M A A I T L L G L
L L V A S S/T I D/E I A G A

hvor bogstaver adskilt af en skråstreg repræsenterer den ene af de to aminosyrer.

FIG. 1

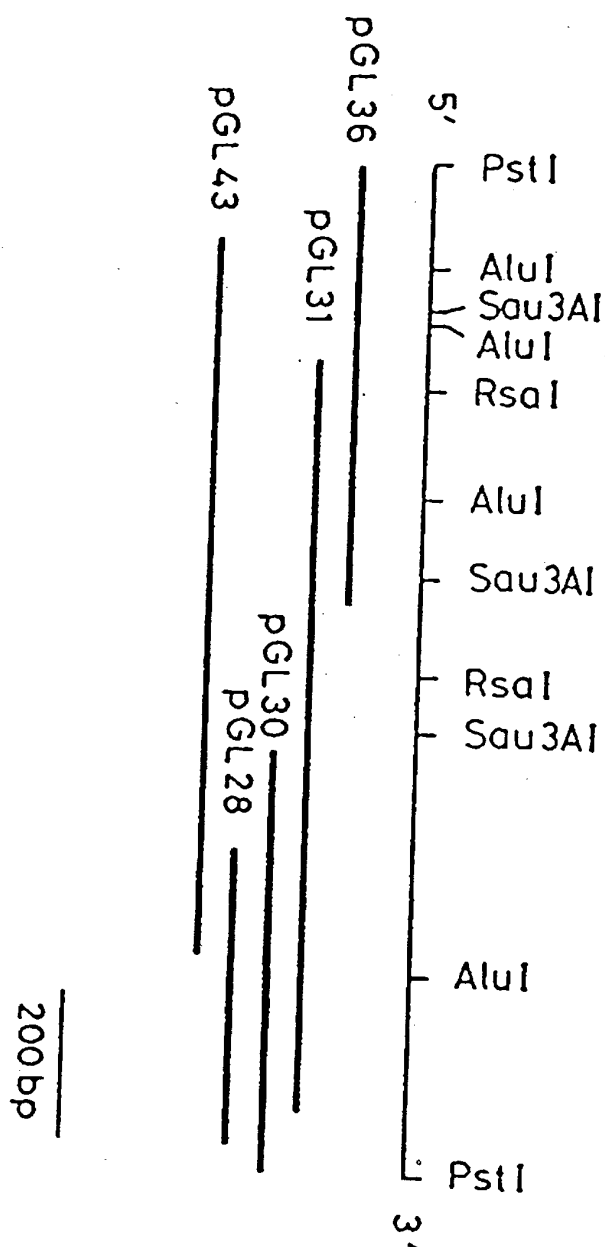


FIG. 2

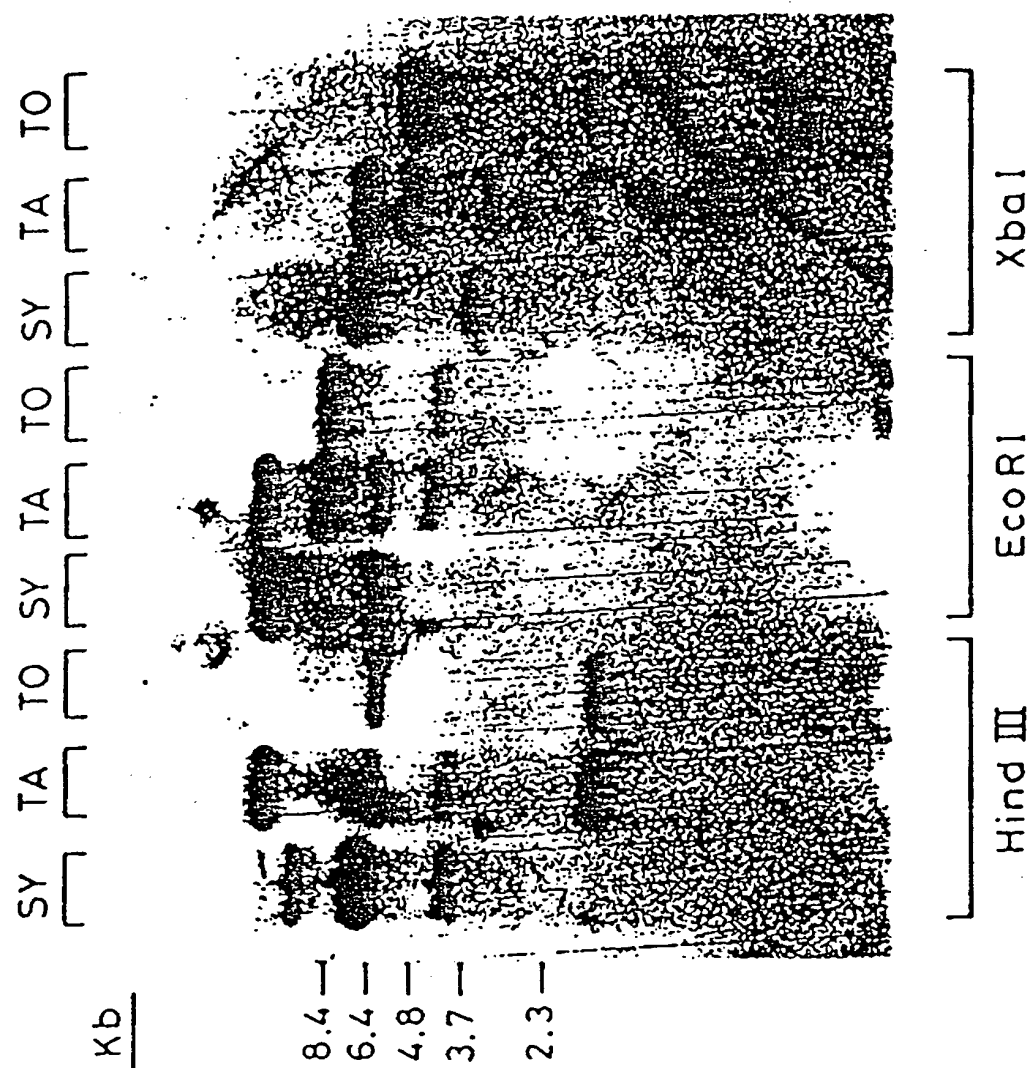
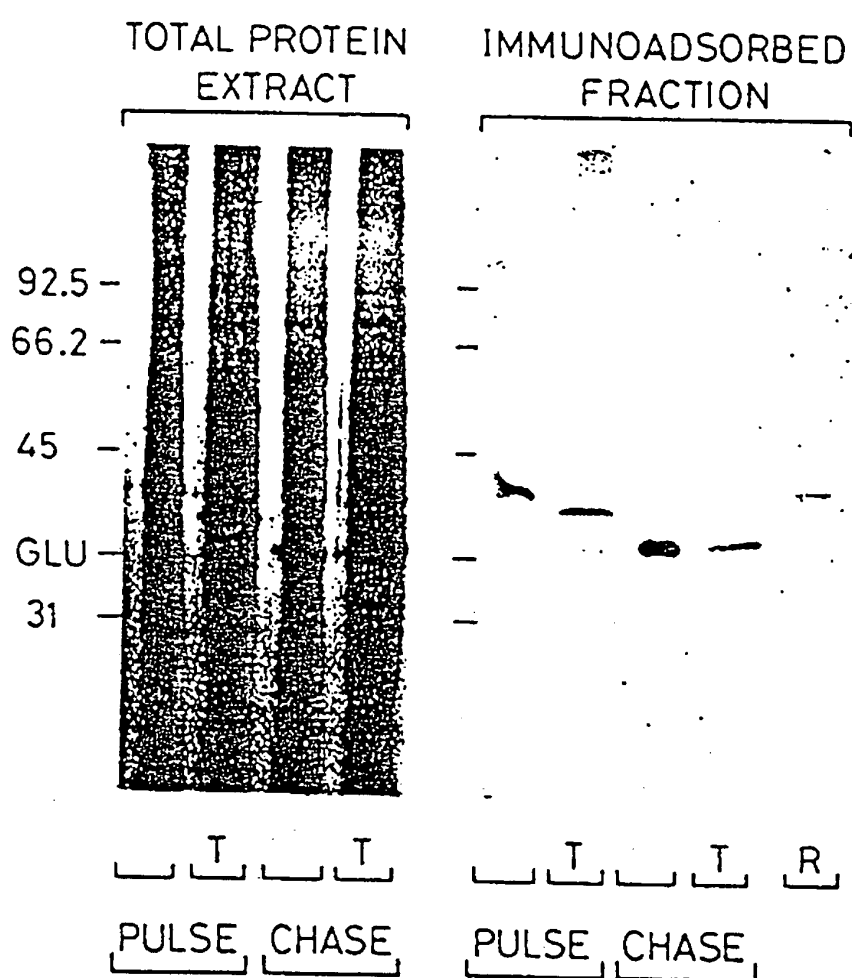


FIG. 3



SEQUENCE I

Fig. 4

Length: 1358

```

      10              30              50
1  CCTCAAATGGCTGCTATCACACTCCTAGGATTACTACTTGTTGCCAGCAGCATTGACATA 60
    M A A I T L L G L L L V A S S I D I
      (N)              90              110
61 GCAGGGGCTCAATCGATAGGTGTTTGCTATGGAATGCTAGGCAACAACCTGCCAAATCAT 120
    A G A Q S I G V C Y G M L G N N L P N H
      130              150              170
121 TGGGAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCA 180
    W E V I Q L Y K S R N I G R L R L Y D P
      190              210              230
181 AATCATGGAGCTTTACAAGCATTAAAAGGCTCAAATATTGAAGTTATGTTAGGACTTCCC 240
    N H G A L Q A L K G S N I E V M L G L P
      250              270              290
241 AATTCAGATGTGAAGCACATTGCTTCCGGAATGGAACATGCAAGATGGTGGGTACAGAAA 300
    N S D V K H I A S G M E H A R W W V Q K
      310              330              350
301 AATGTTAAAGATTTCTGGCCAGATGTTAAGATTAAGTATATTGCTGTTGGGAATGAAATC 360
    N V K D F W P D V K I K Y I A V G N E I
      370              390              410
361 AGCCCTGTCACTGGCACATCTTACCTAACCTCATTTCTTACTCCTGCTATGGTAAATATT 420
    S P V T G T S Y L T S F L T P A M V N I
      430              450              470
421 TACAAAGCAATTGGTGAAGCTGGTTTGGGAAACAACATCAAGGTCTCAACTTCTGTAGAC 480
    Y K A I G E A G L G N N I K V S T S V D
      490              510              530
481 ATGACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTTAGGAACGATGCTAGG 540
    M T L I G N S Y P P S Q G S F R N D A R
      550              570              590
541 TGGTTTACTGATCCAATTGTTGGGTTCTTAAGGGACACACGTGCACCTTTACTCGTTAAC 600
    W F T D P I V G F L R D T R A P L L V N
```

Continuation Fig. 4

	610	630	650	
601	ATTTACCCCTATTTTCAGCTATTCTGGTAATCCAGGGCAGATTCTCTCCCTATTCTCTT			660
	I Y P Y F S Y S G N P G Q I S L P Y S L			
	670	690	710	
661	TTTACAGCACCAATGTGGTAGTACAAGATGGTTCAGCCAATATAGGAACCTATTGAT			720
	F T A P N V V V Q D G S R Q Y R N L F D			
	730	750	770	
721	GCAATGCTGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATT			780
	A M L D S V Y A A L E R S G G A S V G I			
	790	810	830	
781	GTTGTGTCCGAGAGTGGCTGGCCATCTGCTGGTGCATTGGAGCCACATATGACAATGCA			840
	V V S E S G W P S A G A F G A T Y D N A			
	850	870	890	
841	GCAACTTACTTGAGGAACCTAATTCAACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGA			900
	A T Y L R N L I Q H A K E G S P R K P G			
	910	930	950	
901	CCTATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCCCTGAACTGGAG			960
	P I E T Y I F A M F D E N N K N P E L E			
	970	990	1010 (C)	
961	AAACATTTTGGATTGTTTCCCCCAACAAGCAGCCCAAATATAATCAACTTTGGGGTC			1020
	K H F G L F S P N K Q P K Y N I N F G V			
	1030	1050 (N-glyc)	1070	
1021	TCTGGTGGAGTTTGGGACAGTTTCAGTTGAACTAATGCTACTGCTTCTCTCGTAAGTGAG			1080
	S G G V W D S S V E T N A T A S L V S E			
	1090	1110	1130	
1081	ATGTGAGCTGATGAGACACTTGAAATCTCTTTACATACGTATTCCTTGGATGGAACCT			1140
	M *			
	1150	1170	1190	
1141	AGTAAAAACAAGAGAAATTTTTCTTTATGCAAGATACTAAATAACATTGCATGTCTCTG			1200
	1210	1230	1250	
1201	TAAGTCCTCATGGATTGTTATCCAGTGACGATGCAACTCTGAGTGGTTTTAAATTCCTTT			1260

Continuation Fig. 4

	1270	1290	1310	
1261	TCTTTGTGATATTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGCATC	1320		
	1330	1350		
1321	<u>ATTAAT</u> TATACATCAGTTCCATGTTTGATCAAAAAAA	1358		

SEQUENCE II

Fig. 5

Length: 1358

```

      10              30              50
1  CCTCAAATGGCTGCTATCACACTCCTAGGATTACTACTTGTGGCCAGCAGCATTGACATA 60
    M A A I T L L G L L L V A S S I D I

      (N)              90              110
61 GCAGGGGCTCAATCGATAGGTGTTTGCTATGGAATGCTAGGCAACAACCTTGCCAAATCAT 120
    A G A Q S I G V C Y G M L G N N L P N H

      130              150              170
121 TGGGAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCA 180
    W E V I Q L Y K S R N I G R L R L Y D P

      190              210              230
181 AATCATGGAGCTTTACAAGCATTAAAAGGCTCAAACATTGAAGTTATGTTAGGACTTCCC 240
    N H G A L Q A L K G S N I E V M L G L P

      250              270              290
241 AATTCAGATGTCAAGCACATTGCTCCGGAATGGAACATGCAAGATGGTGGGTACAAAAA 300
    N S D V K H I A S G M E H A R W W V Q K

      310              330              350
301 AATGTTAAAGATTTCTGGCCAGATGTTAAGATTAAGTATATTGCTGTTGGGAATGAAATC 360
    N V K D F W P D V K I K Y I A V G N E I

      370              390              410
361 AGCCCTGTACAGGCACATCTTACCTTACCTCATTCTTACTCCTGCCATGGTAAATATT 420
    S P V T G T S Y L T S F L T P A M V N I

      430              450              470
421 TACAAAGCAATTGGTGAAGCTGGTTTAGGAAACAACATCAAGGTCTCAACTTCTGTAGAC 480
    Y K A I G E A G L G N N I K V S T S V D

      490              510              530
481 ATGACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTTAGGAACGATGCTAGG 540
    M T L I G N S Y P P S Q G S F R N D A R

      550              570              590
541 TGGTTTACTGATCCAATTGTTGGGTTCTTAAGGGACACACGTGCACCTTTACTCGTTAAC 600
    W F T D P I V G F L R D T R A P L L V N

```

Continuation Fig. 5

	610	630	650	
601	ATTTACCCCTATTTTCAGTTATTCTGGTAATCCAGGCCAGATTTCTCTCCCCTATTCTCTT			660
	I Y P Y F S Y S G N P G Q I S L P Y S L			
	670	690	710	
661	TTTACAGCACCAAATGTGGTGGTACAAGATGGTTCGCGCAATATAGGAACTTATTTGAT			720
	F T A P N V V V Q D G S R Q Y R N L F D			
	730	750	770	
721	GCAATGCTGGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATT			780
	A M L D S V Y A A L E R S G G A S V G I			
	790	810	830	
781	GTTGTGTCCGAGAGTGGCTGGCCATCTGCTGGTGCATTTGGAGCCACATATGACAATGCA			840
	V V S E S G W P S A G A F G A T Y D N A			
	850	870	890	
841	GCAACTTACTTGAGGAACTTAATTCAACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGA			900
	A T Y L R N L I Q H A K E G S P R K P G			
	910	930	950	
901	CCTATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCCTGAACTGGAG			960
	P I E T Y I F A M F D E N N K N P E L E			
	970	990	1010 (C)	
961	AAACATTTTGGATTGTTTTCCCCCAACAAGCAGCCCAATATAATCTCAACTTTGGGGTC			1020
	K H F G L F S P N K Q P K Y N L N F G V			
	1030	1050 (N-glyc)	1070	
1021	TCTGGTGGTGTGTTGGGACAGTTCAGTTGAACTAATGCTACTGCTTCTCTCGTAAGTGAG			1080
	S G G V W D S S V E T N A T A S L V S E			
	1090	1110	1130	
1081	ATGTGAGCTGATGAGACACTTGAAATCTCTTTACATACGTATTCCTTGGATGGAAAACCT			1140
	M *			
	1150	1170	1190	
1141	AGTAAAAACAAGAGAAATTTTTCTTTATGCAAGATACTAAATAACATTGCATGTCTCTG			1200
	1210	1230	1250	
1201	TAAGTCCTCATGGATTGTTATCCAGTGACGATGCAACTCTGAGTGGTTTTAAATTCCTTT			1260

Continuation Fig. 5
1310

1270 1290
1261 TCTTTGTGATATTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGCATC 1320
1330 1350
1321 ATTAATTTATACATCAGTTCCATGTTTGATCAAAAAAAAA 1358

SEQUENCE III

Fig. 6

Length: 1358

```

      10              30              50
1  CCTCAAATGGCTGCTATCACACTCCTAGGATTACTACTTGTGGCCAGCAGCATTGACATA 60
   M A A I T L L G L L L V A S S I D I

      (N)              90              110
61 GCAGGGGCTCAATCGATAGGTGTTTGCTATGGAATGCTAGGCAACAACCTTGCCAAATCAT 120
   A G A Q S I G V C Y G M L G N N L P N H

      130              150              170
121 TGGGAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCA 180
   W E V I Q L Y K S R N I G R L R L Y D P

      190              210              230
181 AATCATGGAGCTTTACAAGCATTAAAAGGCTCAAATATTGAAGTTATGTTAGGACTTCCC 240
   N H G A L Q A L K G S N I E V M L G L P

      250              270              290
241 AATTCAGATGTGAAGCACATTGCTTCCGGAATGGAACATGCAAGATGGTGGGTACAGAAA 300
   N S D V K H I A S G M E H A R W W V Q K

      310              330              350
301 AATGTTAAAGATTTCTGGCCAGATGTTAAGATTAAGTATATTGCTGTTGGGAATGAAATC 360
   N V K D F W P D V K I K Y I A V G N E I

      370              390              410
361 AGCCCTGTCACTGGCACATCTTACCTAACCTCATTCTTACTCCTGCTATGGTAAATATT 420
   S P V T G T S Y L T S F L T P A M V N I

      430              450              470
421 TACAAAGCAATTGGTGAAGCTGGTTTGGGAAACAACATCAAGGTCTCAACTTCTGTAGAC 480
   Y K A I G E A G L G N N I K V S T S V D

      490              510              530
481 ATGACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTTAGGAACGATGCTAGG 540
   M T L I G N S Y P P S Q G S F R N D A R

      550              570              590
541 TGGTTTGTGATCCCATTGTTGGCTTCTTAAGGGACACACGTGCACCTTTACTCGTTAAC 600
   W F V D P I V G F L R D T R A P L L V N

```

Continuation Fig. 6

610 630 650
 601 ATTTACCCCTATTTTCAGCTATTCTGGTAATCCAGGGCAGATTCTCTCCCTATTCTCTT 660
 I Y P Y F S Y S G N P G Q I S L P Y S L
 670 690 710
 661 TTTACAGCACCAAATGTGGTAGTACAAGATGGTTCACGCCAATATAGGAACCTATTGAT 720
 F T A P N V V V Q D G S R Q Y R N L F D
 730 750 770
 721 GCAATGCTGGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATT 780
 A M L D S V Y A A L E R S G G A S V G I
 790 810 830
 781 GTTGTGTCCGAGAGTGGCTGGCCATCTGCTGGTGCATTGGAGCCACATATGACAATGCA 840
 V V S E S G W P S A G A F G A T Y D N A
 850 870 890
 841 GCAACTTACTTGAGGAACTTAATTCAACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGA 900
 A T Y L R N L I Q H A K E G S P R K P G
 910 930 950
 901 CCTATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCTGAACTGGAG 960
 P I E T Y I F A M F D E N N K N P E L E
 970 990 1010 (C)
 961 AAACATTTTGGATTGTTTTCCCCCAACAAGCAGCCCAATATAATATCAACTTTGGGGTC 1020
 K H F G L F S P N K Q P K Y N I N F G V
 1030 1050 (N-glyc) 1070
 1021 TCTGGTGGAGTTTGGGACAGTTCAGTTGAACTAATGCTACTGCTTCTCTCGTAAGTGAG 1080
 S G G V W D S S V E T N A T A S L V S E
 1090 1110 1130
 1081 ATGTGAGCTGATGAGACACTTGAAATCTCTTTACATACGTATTCCTTGATGGAAAACCT 1140
 M *
 1150 1170 1190
 1141 AGTAAAAACAAGAGAAATTTTTCTTTATGCAAGATACTAAATAACATTGCATGTCTCTG 1200
 1210 1230 1250
 1201 TAAGTCCTCATGGATTGTTATCCAGTGACGATGCAACTCTGAGTGGTTTTAAATTCCTTT 1260

Continuation Fig. 6

1270 1290 1310
 1261 TCTTTGTGATATTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGCATC 1320

1330 1350
 1321 ATTAATTATACATCAGTTCCATGTTTGATCAAAAAAAAA 1358

SEQUENCE IV

Fig. 7

Length: 1358

```

      10              30              50
1  CCTCAAATGGCTGCTATCACA...CTAGGATTACTACTTGTGCCAGCAGCATTGACATA 60
   M A A I T L L G L L L V A S S I D I
      (N)              90              110
61 GCAGGGGCTCAATCGATAGGTGTTTGC...TATGGAATGCTAGGCAACA...CTTGCCAAATCAT 120
   A G A Q S I G V C Y G M L G N N L P N H
      130              150              170
121 TGGGAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCA 180
   W E V I Q L Y K S R N I G R L R L Y D P
      190              210              230
181 AATCATGGAGCTTTACAAGCATTAAAAGGCTCAAATATTGAAGTTATGTTAGGACTTCCC 240
   N H G A L Q A L K G S N I E V M L G L P
      250              270              290
241 AATTCAGATGTGAAGCACATTGCTTCCGGAATGGAACATGCAAGATGGTGGGTACAGAAA 300
   N S D V K H I A S G M E H A R W W V Q K
      310              330              350
301 AATGTTAAAGATTTCTGGCCAGATGTTAAGATTAAGTATATTGCTGTTGGGAATGAAATC 360
   N V K D F W P D V K I K Y I A V G N E I
      370              390              410
361 AGCCCTGTCACTGGCACATCTTACCTAACCTCATTTCTTACTCCTGCTATGGTAAATATT 420
   S P V T G T S Y L T S F L T P A M V N I
      430              450              470
421 TACAAAGCAATTGGTGAAGCTGGTTTGGGAAACAACATCAAGGTCTCAACTTCTGTAGAC 480
   Y K A I G E A G L G N N I K V S T S V D
      490              510              530
481 ATGACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTTAGGAACGATGCTAGG 540
   M T L I G N S Y P P S Q G S F R N D A R
      550              570              590
541 TGGTTTACTGATCCAATTGTTGGGTTCTTAAGGGACACACGTGCACCTTTACTCGTTAAC 600
   W F T D P I V G F L R D T R A P L L V N

```

Continuation Fig. 7

	610	630	650	
601	ATTTACCCCTATTTTCAGCTATTCTGGTAATCCAGGGCAGATTTCTCTCCCCTATTCTCTT			660
	I Y P Y F S Y S G N P G Q I S L P Y S L			
	670	690	710	
661	TTTACAGCACCAAATGTGGTAGTACAAGATGGTTCACGCCAATATAGGAACCTATTTGAT			720
	F T A P N V V V Q D G S R Q Y R N L F D			
	730	750	770	
721	GCAATGCTGGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATT			780
	A M L D S V Y A A L E R S G G A S V G I			
	790	810	830	
781	GTTGTGTCGAGAGTGGCTGGCCATCTGCTGGTGCATTGGAGCCACATATGACAATGCA			840
	V V S E S G W P S A G A F G A T Y D N A			
	850	870	890	
841	GCAACTTACTTGAGGTACTTAATTCCACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGA			900
	XXX XXX			
	A T Y L R Y L I P H A K E G S P R K P G			
	910	930	950	
901	CCTATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCCTGAACTGGAG			960
	P I E T Y I F A M F D E N N K N P E L E			
	970	990	1010 (C)	
961	AAACATTTTGGATTGTTTCCCCCAACAAGCAGCCCAAATATAATATCAACTTTGGGGTC			1020
	K H F G L F S P N K Q P K Y N I N F G V			
	1030	1050 (N-glyc)	1070	
1021	TCTGGTGGAGTTTGGGACAGTTCAGTTGAAACTAATGCTACTGCTTCTCTCGTAAGTGAG			1080
	S G G V W D S S V E T N A T A S L V S E			
	1090	1110	1130	
1081	ATGTGAGCTGATGAGACACTTGAAATCTCTTTACATACGTATTCCTTGGATGGAAAACCT			1140
	M *			
	1150	1170	1190	
1141	AGTAAAAACAAGAGAAATTTTTCTTTATGCAAGATACTAAATAACATTGCATGTCTCTG			1200
	1210	1230	1250	
1201	TAAGTCCTCATGGATTGTTATCCAGTGACGATGCAACTCTGAGTGGTTTTAAATTCCTTT			1260

Continuation Fig. 7

	1270		1290		1310	
1261	TCTTTGTGATATTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGCATC					1320
	1330		1350			
1321	<u>ATTAATT</u> TATACATCAGTTCCATGTTTGATCAAAAAAAAA					1358

SEQUENCE V

Fig. 8

Length: 4483

1 GAGCTCCTCT AGGGGGCCAA GGCAAACCTT TTTGCTAATG GGAAAAAAT
51 ATTAGACCAA GAGTTTGTA TAGCTACTCA ATTTCAATTT ACAAAGGGGA
101 AAATTTAACT GTTTTGCCCT TATATCTTTT GGTCCCGAAA CATAAAATAT
151 CCCATCCGAA ATTCCAAATG GTCCATTATC GGCAAGTAGC TTTCTTTTAT
201 TATAGTTAGT TGACAAAACA CTATCGAGAT ATCATTAGTA TAATAATAAC
251 TTCAAAGTCC ATCTTCTTAG CTGCCTCCTC ATTAGAGCCG CCACTAAAAT
301 AAGACCGATC AAATAAAAGC CGCCATTAAA ATAATGAATT TTAGGACTCT
351 CAATTGTCAC GTAAGTGCCA AACTCTTCC AATACTTTGC TGCAACTTGG
401 GGCTGCTAGC TTCTGAGCTT CCTGGGATAT TTCTATGTTT ATCTCTTAAT
451 TTACATCTCA ACTAATATCA AGAAATTAAA CAGGTGCAGC AAATCATAAA
501 ATTTTCCTCT AAAGAAGAAA ATGACTCCGG TTAGTGATTC ATTGGCCTTT
551 TCAGAGTCTG CGTGCCATAT TCACTAATGG GTCGTTTGGT ACAAGAAATA
601 ATGATAATAA TTTCGAAATA GAATTTGGGA TTTCATTTAT TTCATATTTA
651 ATTATAAATA TTAGCTAATT TCGAAATAAA TTTTACATTA AAATAGTGAA
701 ATCAACTATC TCATATATAA GGTGGAATAG CTAATCTCAT AGCCACCTCA
751 GTTAGAATCC AGTTTCCTCT AATAAATGCA GCGAAATATT AGAAGACTTT
801 CATTAAATCA ATTCATATAA TTTAAAATA CTAGACATGG AAAAAAAAAA
851 CGATTGAGA CTGTTATGGA AGGCGTTGCC TTCGATGTAG ATTCTCATCC
901 ATTGCTTTTCG TGCAATAGCA ATATGACATC TTATCTTAG AACTACTTTT
951 AAATGAAAGT CATATAGAAC TTTAAATCT CTCAACTAGT TTTAAGGGAA
1001 TTCAAATAC GACCAATATT TATTACTTAC TTATGGATAA ATTTAAATAA
1051 TATGTATTTT ATCTTGAAAT TGAATTGAAA ATATTAAATT ACTTGTTTAA
1101 ATATAAACAA TAGATATCGT TAAGTATTTA CCACAAACAT TTTATTAGTT
1151 GTAACGATGA TTAAGCAGGA ATTCCTCTGG TTGTGCAGGA TGAAAGAAAC
1201 TAACTAGCTA TAATTTCTTT TGTAAGTCA AGATAGTACG GCACCTTATG

Continuation Fig. 8

1251 GAGAAAATAA ATAACCTTAC ATCATCAAAC TCATCTTCTT TTTTCCACAA
1301 AATGATCAAG TTGACATGTT AATAGCCAGG TCACCGGGGG CGGCTCTTAA
1351 CTTCAATTAGC CTACGAATAA CAAATCCAAT ATTATATTTA CACAAGGCTA
1401 TATATATATC TCAATATAAA TAGCTCGTTG TTCATCTTAA TTCTCCCAAC
1451 AAGTCTTCCC ATCATGTCTA CCTCACATAA ACATAATACT CCTCAAATGG
1501 CTGCTATCAC ACTCCTAGGA TTACTACTTG TTGCCAGCAG CATTGACATA
1551 GCAGGTTTCT GGTCAAATAT TTGAACCTCC CAGCCAAAAA TATTGTCTTA
1601 TAATTTTGTG TGCGCAAAAT TTTAATTTAG TTGATAGTTA TTTGCTTATT
1651 TTTCTTTTCA AATTGCTTGT GTTTTTTCT CAAATTAAC TGCACCGTAT
1701 TCATTTAGCG ATAGTTATTT GCTCTATTTT TGTGTAACAC TCACTCACAA
1751 ACTTTTCAAT TTGAGGGGAG GACAGTGAAT CTAAGATTGA AATTTATGAG
1801 TTTAATTAGA CTAATTCCCA TTTGATTTAT TGGCTAGAAG TCAATTATTT
1851 GCATAGTGAG TCTTTTAACA CACAGATTTG AGTTAAAGCT ACTACGTTCTG
1901 TATTAACCCA TAACATATAC ACCTTCTGTT CTAATTTCTT TGACACTTTT
1951 TGTTAGTTTG TTCCAAAAAG GACGGACATA TTTGATATTT GAGAATACTT
2001 TACCTTAACC TTAATAGAAT TTTTATGAC ATCACATATA TTATGGAATA
2051 TATACGACCA TAATTTTCAA ATATCTTATA GTCGTACAAA TATTATAGCA
2101 TGTTTAATAC CACAACCTTC AAATTCTTCT TTCCTTAAA AACAAAATAT
2151 GTCACATAAA TTAAAATAGA GGAAGTATAC TACATCAATC AGCCCCTAGT
2201 GGAGGGGACC CTACTGTAAG TTTTAAAGTT TTCAAGAATT CAGTAATTGA
2251 TTAGGAGCCC GTCTGGACAT AAAAAAAAT TCCTTTTTTT CCAAAAAATG
2301 CCCACTAAAT TTCTAACACT ATTTTGTAAT TCTTATTGAG CAGGGGGCTC
2351 AATCGATAGG TGTTTGCTAT GGAATGCTAG GCAACAACCT GCCAAATCAT
2401 TGGGAAGTTA TACAGCTCTA CAAGTCAAGA AACATAGGAA GACTGAGGCT
2451 TTATGATCCA AATCATGGAG CTTTACAAGC ATTAAAAGGC TCAAATATTG
2501 AAGTTATGTT AGGACTTCCC AATTCAGATG TGAAGCACAT TGCTCCGGA
2551 ATGGAACATG CAAGATGGTG GGTACAGAAA AATGTTAAAG ATTTCTGGCC

Continuation Fig. 8

2601 AGATGTTAAG ATTAAGTATA TTGCTGTTGG GAATGAAATC AGCCCTGTCA
2651 CTGGCACATC TTACCTAACC TCATTTCTTA CTCCTGCTAT GGTAATATT
2701 TACAAAGCAA TTGGTGAAGC TGGTTTGGGA AACAAATCA AGGTCTCAAC
2751 TTCTGTAGAC ATGACCTTGA TTGGAACTC TTATCCACCA TCACAGGGTT
2801 CGTTTAGGAA CGATGCTAGG TGGTTTGTTG ATCCCATTTG TGGCTTCTTA
2851 AGGGACACAC GTGCACCTTT ACTCGTTAAC ATTTACCCCT ATTTCAAGTTA
2901 TTCTGGTAAT CCAGGCCAGA TTTCTCTCCC CTATTCTCTT TTTACAGCAC
2951 CAAATGTGGT GGTACAAGAT GGTCCCGCC AATATAGGAA CTTATTTGAT
3001 GCAATGCTGG ATTCTGTGTA TGCTGCCCTC GAGCGATCAG GAGGGGCATC
3051 TGTAGGGATT GTTGTGTCCG AGAGTGGCTG GCCATCTGCT GGTGCATTTG
3101 GAGCCACATA TGACAATGCA GCAACTTACT TGAGGAAGTT AATTCAACAC
3151 GCTAAAGAGG GTAGCCCAAG AAAGCCTGGA CCTATTGAGA CCTATATATT
3201 TGCCATGTTT GATGAGAACA ACAAGAACCC TGAAGTGGAG AACATTTTG
3251 GATTGTTTTT CCCCACAAG CAGCCCAAAT ATAATATCAA CTTTGGGGTC
3301 TCTGGTGGAG TTTGGGACAG TTCAGTTGAA ACTAATGCTA CTGCTTCTCT
3351 CGTAAGTGAG ATGTGAGCTG ATGAGACACT TGAATCTCT TTACATACGT
3401 ATTCCTTGGA TGGAAAACCT AGTAAAAACA AGAGAAATTT TTTCTTTATG
3451 CAAGATACTA AATAACATTG CATGTCTCTG TAAGTCCTCA TGGATTGTTA
3501 TCCAGTGACG ATGCAACTCT GAGTGGTTTT AAATTCCTTT TCTTTGTGAT
3551 ATTGGTAATT TGGCAAGAAA CTTTCTGTAA GTTGTGAAT TTCATGCATC
3601 ATTAATTATA CATCAGTTCC ATGTTTGATC AGATTGGGAT TTGGTAACTT
3651 CAATGTTTAG TTATTTATTA ATTAGTGTCT TTATCATTTG ACTATCAATT
3701 AATCTTTATT TGGCAAGGCT TGATATATTT GAGTTACTCT TAGGTATTTG
3751 CAAGCAACTG ATCTTTCTTT TATCCCGTTT CTCCGTAAA CCTCATTAGA
3801 AATATATTAT AATGTCACCT ACTCTGTGGT TTAAGACATT CCCTTACATT
3851 ATAAGGTATT TCACGTCGTA TCAGGTCGAA AAAAATAATG GTACGCTCTT

Continuation Fig. 8

3901 TCTTATCACA AATTTCTCTC AACTTCTAGA CCAATTGAAT CTTGTCTCCA
3951 ATAAGCATTG CTTTACTGT ATGTTTCTCT CTATCAATTC AAGGCTCAAG
4001 CCATCAAATC GCTTGGTATT TCTCGCTCCC ATTTAACCBA TCGAGCTGTT
4051 AGCTCTGCTA AAGTCTCATT CTTGAGCTTC TCAAACCACT CAGAGCTTCT
4101 CCAACCAAAA ATGTAGAAAA AATCTTTGAT CCACATGTAA AACCCAGAA
4151 TTGTCATTCG AGTGAAGGAA GGTGCTGGAA ACGCGACATC AAACACGGAT
4201 TTGCCAGCTG ATTGTCATAG GAAAAAGAAA AGTTGATCAG ATAGTTCGTG
4251 CGGGGTATTT CCTACCTCGC CTTTGATCAT AGCTCTAGGC TTGCGAACT
4301 AAGTTATGCA CGAGGCCCCC TGCGAATCCA TAAATCTAAG AAGCATGCCA
4351 CAACAAACGG GATAACTTCC AGCTAAGAGA TCTATTTTGA TCTTACCCAC
4401 AAAGTAGCTC CGTGTGAATT GTCCATTTCA TCAACTCCAA AGGCAGGAAA
4451 GAAGACTCTA TTCAGAGAGC AGCAAAAGAG CTC

SEQUENCE VI

Fig. 9

Length: 4699

```
1  GAGCTCCCTT GGGGGGCAAG GGCAAACTT TTTGCTAAAT GGAAAAATAT
51  TATACCAAGT GTTTGTAATA GTTACTCAAT TTGAATTAAC AAAGGGGCAA
101 ATTTGACTAT TTTGCCCTTA TATCTTTTGG TCACAAAAAC ATAAAATATC
151 CCATCCGAAA TTCCAAATGG TCCATTATCG GCAAGTAGCT TTCTTTAATT
201 ATAGTTAGTT GACAAAACAC TATCAAGATA TCATTATTAT AATAATAACT
251 TCAAAGTCCA TCATCTTAGC TGCCTCCTCA GTAGAGCCGC CAGTAAAATA
301 AGACCGATCA AATAAAAGCC GCCATTAAAA TAATGAATTT TAGGACTCTC
351 GATTGGCACG TAAGTGCCAA AACTCTTCCA ATACTTTGCT GCAACTTGGG
401 GCTGCTAGGT TCTGAGCTTC CAGATATGGG ATATTTCTAA GTTTATCTCC
451 TAATTTACAT CTCAACTAAT ATTAAGAAAT TAAACAGGTA CAGCAAATCA
501 TAAAATTTTC CTCTAAAGAA GACAATGAAT CCGGTTACTG ATTCATTGGC
551 CTTTTCAGAG TCTGCATGCC ATATTCACTA AGGGGTCGTT TGGTACAAGA
601 AATAATAATA ATAATTTTCGG GATAGAATTT GAGATTGCAT TTATCTTGTG
651 TTAAATTATA AGTATTAGCT AATTTCAGAA TAAATTTTAC ACTAAAATAG
701 TAAAATCAAC TATCACATGT AGAAGGTGGA ATGGAATAGC TAATCCCATA
751 GCCACTCACA TAGAATATCC TTATTTATCT CACTATTTTA CCAAATGATC
801 GGTTAGTCTT CATGAGAATC CAGTATCCTC AATAAATGCA GTAAGAAGTT
851 AGAAAATTTT CATTAAATCA ATTCATATAA TTTAAAAATA TTAGATATGG
901 AGCACTTAAG ATACAATAAA AGATGTACCG TTAATAATAA AAGATAAGAT
951 AGAGTTTTAA ATAGGAAAAA AAAAACGGTT CGAGACACTC TTATGGAAGG
1001 CGTTGTCTTC AAAGTAGATT CTCATTCATT GCTCTGGTGC AATAGCAAAA
1051 TGACATCTTA CTCTTAAGAT ACAGCGAGCC ACTCTACAAT CTTCTATTGT
1101 ATACTCAAAT GAAAGTTTTA GAGAACTTTC AAATCTCTCA ACTACTTTTA
1151 AGGGAATTCA AAATACGACC AATATTTATT ACTTACTTAC TTATAGTTAA
1201 ATGATATGAA TTTTATTTTA AATTTGAATT GAAAATATTA AATTACTTGA
```

Continuation Fig. 9

1251 TTTAATATAA ACAATAGATA TCGCTAAGTA TTTACCACAA ACATGGAGAT
1301 ACTACAGAAG ATTTTATTAT TTGTAACGAT GATTAAGCAG CTATTCATCT
1351 GGTGTGCAG GATGAAAGAA AGTAACTAGC TATAATTTCT TTTGTAAAGT
1401 CAAGATAGTA CGGCACCTTT GGANNAAATA TAAACTTGT ACATCATCAA
1451 ACTCATTTTC TTTTTCAC AAAATGATCA AGTTGACATG TTAATAGCCA
1501 GGTGAGCCGG GCGGCTCTT AACTTCATTA GCCTACGAAT AACAAATACT
1551 CCAATAATAT TCCACTCGAA TATTTACATT TACACAAGGC TATCTCAATA
1601 TAAATAGCTC ATTGTTTCATC TTAATTCCTC CAACAAGTCT TCTCATCATG
1651 TCTACCTCAG ATAAACATAA TACTCCTCAA ATGGCTGCTA TCACACTGCT
1701 AGGATTACTA CTTGTTGCCA GCACCATGA GATAGCAGGT TTCTGGTCAA
1751 ATATTTGAAC TTCCGAGCCA AAAATATTGT CTTATAATTT TGTGGTGCA
1801 AAATCTTAAT TTAGTTGATA GTTATTGCT TATTTTCTT ATCAAATTGC
1851 TTGTGTTTTT TTCTCAAATT TACTGCACT ATATTCATT AGCGATAGTT
1901 ATTTGCTCTT TTTTCGGTA AACTCACTC ACAAGCTTT CAAATTTGAG
1951 GGGAGGGCGG TGAATCTAAA ATTTGAAATT TATGAGTTA GACTAGTGTC
2001 CATTTGATTT ATTGGCTAGA CGTCTATTAG TTGTATAGTA AATCTTTTAA
2051 CACATACACC GACCTGAGTC AAAGCTATTA GGTTCGTATT AACACATAAC
2101 ACATATTCCC TCTGTTCTAA TTTATGTGAC AACTTTCTG TTAGTTTCTT
2151 CCAAAGAGAA TGACATATTT GATATTTAAA AATATTTTAA CTTTAAACTT
2201 TTTATTCTCA ACCTTTTATA ACATCACAAA TATTATGGAA CATATTAGAC
2251 TACAAGTTCC AAATATCTTA TAGTCTGTAC AAATATTATA GCATGTTTAA
2301 TAACACAATT TTCACATTCT TCTTTTCTT AAACTTTGTG CCGAATTAAA
2351 TTATGTCACA TAAATTAAAA TGGTTACATC ATCCCCTAGT GGAGGGACCT
2401 ACCATGTCTA CTGTAAGTTT TTAACTTTC AAGAATTACA TAATTGATTT
2451 AGTTTCTAAC ACTAATTCTA ATTCTTATTG AGCAGGGGCT CAATCAATAG
2501 GTGTTTGCTA TGGAATGCTA GGCAACAAC TGCCAAATCA TTGGGAAGTT
2551 ATACAGCTCT ACAAGTCAAG AACATAGGA AGACTGAGGC TTTATGATCC

Continuation Fig. 9

2601 AAATCATGGA GCTTTACAAG CATTAAAAGG CTCAAACATT GAAGTTATGT
2651 TAGGACTTCC CAATTCAGAT GTCAAGCACA TTGCTTCCGG AATGGAACAT
2701 GCAAGATGGT GGGTACAAAA AAATGTTAAA GATTTCTGGC CAGATGTTAA
2751 GATTAAGTAT ATTGCTGTTG GGAATGAAAT CAGCCCTGTC ACAGGCACAT
2801 CTTACCTTAC CTCATTTCTT ACTCCTGCCA TGGTAAATAT TTACAAAGCA
2851 ATTGGTGAAG CTGGTTTAGG AAACAACATC AAGGTCTCAA CTTCTGTAGA
2901 CATGACCTTG ATTGGAAACT CTTATCCACC ATCACAGGGT TCGTTTAGGA
2951 ACGATGCTAG GTGGTTTACT GATCCAATTG TTGGGTTCTT AAGGGACACA
3001 CGTGCACCTT TACTCGTTAA CATTTACCCC TATTTAGCT ATTCTGGTAA
3051 TCCAGGGCAG ATTTCTCTCC CCTATTCTCT TTTTACAGCA CCAATGTGG
3101 TAGTACAAGA TGGTTCACGC CAATATAGGA ACTTATTTGA TGCAATGCTG
3151 GATTCTGTGT ATGCTGCCCT CGAGCGATCA GGAGGGGCAT CTGTAGGGAT
3201 TGTTGTGTCC GAGAGTGGCT GGCCATCTGC TGGTGCATTT GGAGCTACAT
3251 ATGACAATGC AGCAACTTAC TTGAGGAACT TAATTCAACA CGCTAAAGAG
3301 GGTAGCCCAA GAAAGCCTGG ACCTATTGAG ACCTATATAT TTGCCATGTT
3351 TGATGAGAAC AACAAGAACC CTGAACTGGA GAAACATTTT GGATTGTTTT
3401 CCCCCAACAA GCAGCCCAA TATAATCTCA ACTTTGGGGT CTCTGGTGGT
3451 GTTTGGGACA GTTCAGTTGA AACTAATGCT ACTGCTTCTC TCATAAGTGA
3501 GATGTGAGAT GAGACACTTG AAATCTCTTT ACATAAGTAT TGCTTAGATG
3551 GAAAGCTTAG TAAAAACAAG AGAAATTTAT TCTTCATGCA AGACACTAAA
3601 TAACATTGCA CGTCTCTGTA AGTCCTCATG GATTGTTATC CAGAGACGAT
3651 GCAACTCTGA GTGGTTTTAA ATCCCTTTTC TTTGTGATGT TGGTAATTTG
3701 GCAAGAAACT TTCTGTAAGT TTGTGAATTT CATGACTTTC TGTAAGTTTG
3751 TGAATTTTCA GCACCATCAA TTATACATCT TTTCCATGTT TGATCACATT
3801 AGGATTTGGT AATTGCAAAG TTTAGTTATT TATTAATTAG TGTCTTTATC
3851 ATTTGACTCG ATCAATTAAT CTTTAATTGG TAAGGCTTGA TATATCGGAG

Continuation Fig. 9

3901 CGACTCTTAG GTAGTGGCAT TCAACTGATC TTTCTGTTAT CCCATGTCTC
3951 CGTTAACCCT CATTAGAAAT ATATTATAAT GTCACCTAAA TCAGAGGTTT
4001 TAGAGCTTCA AAATCGATTG ACAACAGTTT TGGAGTTACA CTGTGGTTTA
4051 GGACATTCTG TTACATTATA AGGTATTTCA CGTCGTATCA AGGTCGAATA
4101 AAATAATGGT ACGCTCTTTC TTATCACAAA TTTCTCTCAA CTTCTAGACC
4151 AATTGAATCT TGTCTCCAAT AAGTATTGCT TTTACTCTAT GTTTCTCTCT
4201 ATCAATTCAA GGCTCAAGCC ATCAAATCGC TTGGTATTTT TCGCTCTCAA
4251 TTAACCAATC GAGCTGTAA CTCTGCTAAA GTCTCATTCT TCAGCTTCTC
4301 AAACCACTCA AAGCTTCTCC AACAAAAAAA GTAGAAAAAA ACTTGATCC
4351 ACATGTAAAA CCCATAACC ATGTCATTCT AGTGAAGGAA GGTGCTTGAA
4401 ATGCAACGAA CAAACACGGC TTGCCACTG CTTGTGATAG GTAAAAGAAA
4451 ACTTGATTAG ATAGTTCGTG CGGGGCATTT CCTACCTCGC CTTTGATCTT
4501 AGCTTTAGGC TTGCGAACT AAGTAGAACC TAAGGCCCCA GCGAATCCGT
4551 AAATCTGAGG AGCATGCCAC AGCAAAAAGG TTAGCTTTCA GATAAGAGAT
4601 CTATTTGATC TTACCCACTA ACTAGCTCTG TGTGAATTGT CCATTTTCATC
4651 AACTCCAAAG GCAGGAAAGA AGACTATTGA GAGAGCAGCA AAAGAGCTC

SEQUENCE VII

Fig. 10

Length: 1407

10	30	50	
1	ATCTTAATTCTCCCAACAAGTCTTCCCATCATGTCTACCTCACATAAACATAATACTCCT	60	
	M S T S H K H N T P		
70	90	110	
61	CAAATGGCTGCTATCACACTCCTAGGATTACTACTTGTGCCAGCAGCATTGACATAGCA	120	
	Q M A A I T L L G L L L V A S S I D I A		
130	150	170	
121	GGGGCTCAATCGATAGGTGTTTGTCTATGGAATGCTAGGCAACAACCTTGCCAAATCATTGG	180	
	G A Q S I G V C Y G M L G N N L P N H W		
190	210	230	
181	GAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCAAAT	240	
	E V I Q L Y K S R N I G R L R L Y D P N		
250	270	290	
241	CATGGAGCTTTACAAGCATTAAAAGGCTCAAATATTGAAGTTATGTTAGGACTTCCCAAT	300	
	H G A L Q A L K G S N I E V M L G L P N		
310	330	350	
301	TCAGATGTGAAGCACATTGCTTCCGGAATGGAACATGCAAGATGGTGGGTACAGAAAAAT	360	
	S D V K H I A S G M E H A R W W V Q K N		
370	390	410	
361	GTTAAAGATTTCTGGCCAGATGTTAAGATTAAGTATATTGCTGTTGGGAATGAAATCAGC	420	
	V K D F W P D V K I K Y I A V G N E I S		
430	450	470	
421	CCTGTCACTGGCACATCTTACCTAACCTCATTTCTTACTCCTGCTATGGTAAATATTTAC	480	
	P V T G T S Y L T S F L T P A M V N I Y		
490	510	530	
481	AAAGCAATTGGTGAAGCTGGTTTGGGAAACAACATCAAGGTCTCAACTTCTGTAGACATG	540	
	K A I G E A G L G N N I K V S T S V D M		
550	570	590	
541	ACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTAGGAACGATGCTAGGTGG	600	
	T L I G N S Y P P S Q G S F R N D A R W		

Continuation Fig. 10

	610	630	650	
601	TTTGTGATCCCATTTGTTGGCTTCTTAAGGGACACACGTGCACCTTTACTCGTTAACATT			660
	F V D P I V G F L R D T R A P L L V N I			
	670	690	710	
661	TACCCCTATTTTCAGTTATTCTGGTAATCCAGGCCAGATTTCTCTCCCCTATTCTCTTTT			720
	Y P Y F S Y S G N P G Q I S L P Y S L F			
	730	750	770	
721	ACAGCACCAAATGTGGTGGTACAAGATGGTTCCCGCCAATATAGGAACCTATTTGATGCA			780
	T A P N V V V Q D G S R Q Y R N L F D A			
	790	810	830	
781	ATGCTGGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATTGTT			840
	M L D S V Y A A L E R S G G A S V G I V			
	850	870	890	
841	GTGTCCGAGAGTGGCTGGCCATCTGCTGGTGCATTGGAGCCACATATGACAATGCAGCA			900
	V S E S G W P S A G A F G A T Y D N A A			
	910	930	950	
901	ACTTACTTGAGGAACCTTAATTCAACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGACCT			960
	T Y L R N L I Q H A K E G S P R K P G P			
	970	990	1010	
961	ATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCCCTGAACTGGAGAAA			1020
	I E T Y I F A M F D E N N K N P E L E K			
	1030	1050	1070	
1021	CATTTTGGATTGTTTTCCCCCAACAAGCAGCCCAAATATAATATCAACTTTGGGGTCTCT			1080
	H F G L F S P N K Q P K Y N I N F G V S			
	1090	1110	1130	
1081	GGTGGAGTTTGGGACAGTTCAGTTGAACTAATGCTACTGCTTCTCTCGTAAGTGAGATG			1140
	G G V W D S S V E T N A T A S L V S E M			
	1150	1170	1190	
1141	TGAGCTGATGAGACACTTGAAATCTCTTTACATACGTATTCCTTGGATGGAAAACCTAGT			1200
	*			
	1210	1230	1250	
1201	AAAAACAAGAGAAATTTTTCTTTATGCAAGATACTAAATAACATTGCATGTCTCTGTAA			1260

Continuation Fig. 10

	1270	1290	1310	
1261	GTCCTCATGGATTGTTATCCAGTGACGATGCAACTCTGAGTGGTTTTAAATTCCTTTTCT			1320
	1330	1350	1370	
1321	TTGTGATATTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGCATCATT			1380
	1390			
1381	AATTATACATCAGTTCCATGTTTGATC	1407		

SEQUENCE VIII

Fig. 11

Length: 1431

```

      10          30          50
1  ATCTTAATTCTTCCAACAAGTCTTCTCATCATGTCTACCTCAGATAAACATAATACTCCT 60
      M S T S D K H N T P

      70          90          110
61 CAAATGGCTGCTATCACACTGCTAGGATTACTACTTGTTGCCAGCACCATTGAGATAGCA 120
      Q M A A I T L L G L L L V A S T I E I A

      130          150          170
121 GGGGCTCAATCAATAGGTGTTTGCTATGGAATGCTAGGCAACAACCTGCCAAATCATTGG 180
      G A Q S I G V C Y G M L G N N L P N H W

      190          210          230
181 GAAGTTATACAGCTCTACAAGTCAAGAAACATAGGAAGACTGAGGCTTTATGATCCAAAT 240
      E V I Q L Y K S R N I G R L R L Y D P N

      250          270          290
241 CATGGAGCTTTACAAGCATTAAAAGGCTCAAACATTGAAGTTATGTTAGGACTTCCCAAT 300
      H G A L Q A L K G S N I E V M L G L P N

      310          330          350
301 TCAGATGTCAAGCACATTGCTTCCGGAATGGAACATGCAAGATGGTGGGTACAAAAAAT 360
      S D V K H I A S G M E H A R W W V Q K N

      370          390          410
361 GTTAAAGATTTCTGGCCAGATGTTAAGATTAAAGTATATTGCTGTTGGGAATGAAATCAGC 420
      V K D F W P D V K I K Y I A V G N E I S

      430          450          470
421 CCTGTCACAGGCACATCTTACCTTACCTCATTCTTACTCCTGCCATGGTAAATATTTAC 480
      P V T G T S Y L T S F L T P A M V N I Y

      490          510          530
481 AAAGCAATTGGTGAAGCTGGTTTAGGAAACAACATCAAGGTCTCAACTTCTGTAGACATG 540
      K A I G E A G L G N N I K V S T S V D M

      550          570          590
541 ACCTTGATTGGAAACTCTTATCCACCATCACAGGGTTCGTTTAGGAACGATGCTAGGTGG 600
      T L I G N S Y P P S Q G S F R N D A R W

```

Continuation Fig. 11

	610	630	650	
601	TTTACTGATCCAATTGTTGGGTTCTTAAGGGACACACGTGCACCTTTACTCGTTAACATT			660
	F T D P I V G F L R D T R A P L L V N I			
	670	690	710	
661	TACCCCTATTTTCAGCTATTCTGGTAATCCAGGGCAGATTTCTCTCCCCTATTCTCTTTTT			720
	Y P Y F S Y S G N P G Q I S L P Y S L F			
	730	750	770	
721	ACAGCACCAAATGTGGTAGTACAAGATGGTTCACGCCAATATAGGAACTTATTTGATGCA			780
	T A P N V V V Q D G S R Q Y R N L F D A			
	790	810	830	
781	ATGCTGGATTCTGTGTATGCTGCCCTCGAGCGATCAGGAGGGGCATCTGTAGGGATTGTT			840
	M L D S V Y A A L E R S G G A S V G I V			
	850	870	890	
841	GTGTCCGAGAGTGGCTGGCCATCTGCTGGTGCATTTGGAGCTACATATGACAATGCAGCA			900
	V S E S G W P S A G A F G A T Y D N A A			
	910	930	950	
901	ACTTACTTGAGGAACTTAATTCAACACGCTAAAGAGGGTAGCCCAAGAAAGCCTGGACCT			960
	T Y L R N L I Q H A K E G S P R K P G P			
	970	990	1010	
961	ATTGAGACCTATATATTTGCCATGTTTGATGAGAACAACAAGAACCCTGAACTGGAGAAA			1020
	I E T Y I F A M F D E N N K N P E L E K			
	1030	1050	1070	
1021	CATTTTGGATTGTTTTCCCCCAACAAGCAGCCCAATATAATCTCAACTTTGGGGTCTCT			1080
	H F G L F S P N K Q P K Y N L N F G V S			
	1090	1110	1130	
1081	GGTGGTGTGTTGGGACAGTTCAGTTGAACTAATGCTACTGCTTCTCTCATAAGTGAGATG			1140
	G G V W D S S V E T N A T A S L I S E M			
	1150	1170	1190	
1141	TGAGATGAGACACTTGAAATCTCTTTACATAAGTATTGCTTAGATGGAAAGCTTAGTAAA			1200
	*			
	1210	1230	1250	
1201	AACAAGAGAAATTTATTCTTCATGCAAGACACTAAATAACATTGCACGTCTCTGTAAGTC			1260

Continuation Fig. 11

	1270	1290	1310	
1261	CTCATGGATTGTTATCCAGAGACGATGCAACTCTGAGTGGTTTTAAATCCCTTTTCTTTG			1320
	1330	1350	1370	
1321	TGATGTTGGTAATTTGGCAAGAACTTTCTGTAAGTTTGTGAATTCATGACTTTCTGTA			1380
	1390	1410	1430	
1381	AGTTTGTGAATTCATGCACCATCAATTATACATCTTTCCATGTTTGATC			1431