



NORGE

(12) **PATENT**

(19) NO

(11) **303546**

(13) B1

(51) Int Cl⁶ C 12 N 15/82, 15/67, 15/32

Patentstyret

(21) Søknadsnr	19904585	(86) Int. inng. dag og søknadsnummer	13.02.1990, PCT/US90/00778
(22) Inng. dag	23.10.1990	(85) Videreføringsdag	23.10.1990
(24) Løpedag	13.02.1990	(30) Prioritet	24.02.1989, US, 315355
(41) Alm. tilgj.	27.12.1990		12.02.1990, US, 476661
(45) Meddelt dato	27.07.1998		
(73) Patenthaver	The Monsanto Co, 800 North Lindbergh Boulevard, St. Louis, MO 63167, US		
(72) Oppfinner	David Allen Fischhoff, Webster Groves, MO, US		
	Frederick Joseph Perlak, St. Louis, MO, US		
(74) Fullmektig	Tandbergs Patentkontor AS, 0306 Oslo		
(54) Benevnelse	Modifiserte, kinære gener som inneholder en promotor som virker i planteceller, samt plantetransformasjonsvektor som inneholder et plantegen som inneholder et slikt gen		
(56) Anførte publikasjoner	EP A3 142924 Adang et al. (1987), UCLA Symp., Mol. Cell. Biol., 48, 345-353		
(57) Sammendrag	Fremgangsmåte for modifikasjon av strukturgensekvenser for å fremme ekspresjonen av proteinproduktet er beskrevet. Også beskrevet er nye strukturgener som koder for insekticide proteiner fra B.t.k. HD-1, B.t.k. HD-73, B.t. tenebrionis, B.t. entomocidus, 2-protein fra B.t.k. HD-1 og kappeprotein til potetbladrollevirus.		

Oppfinnelsens bakgrunn

Modifiserte, kimære gener som inneholder en promoter som virker i planteceller, samt plantetransformasjonsvektor som inneholder et plantegen som inneholder et slikt gen.

5 Selv om det i de senere år er blitt gjort store fremskritt når det gjelder transgene planter som uttrykker fremmedproteiner, slik som herbicidresistente enzymer og viruskappeproteiner, er det kjent svært lite om hovedfaktorene som påvirker ekspresjon av fremmedgener i planter.

10 Flere potensielle faktorer kan i varierende grad være ansvarlig for nivået for proteinekspresjon av en bestemt kodende sekvens. Nivået av en bestemt mRNA i cellen er ganske visst en avgjørende faktor.

 De mulige årsakene for lave likevektsnivåer av mRNA

15 på grunn av naturen til den kodende sekvens er mange. For det første kan det tenkes at fullengde-RNA-syntese ikke oppstår ved høy hyppighet. Dette kan f.eks. være forårsaket av den altfor tidlige avslutning av RNA under transkripsjon, eller skyldes uventet mRNA-prosessering under transkripsjon.

20 For det andre kan fullengde-RNA bli fremstilt, men så prosesseres (spleises, polyA-addisjon) i kjernen på en måte som skaper en ikke-funksjonell mRNA. Dersom RNA syntetiseres, termineres og polyadenyleres korrekt, kan den så bevege seg til cytoplasmaet for translasjon. I cytoplasmaet har mRNA-er

25 entydige halveringstider som bestemmes av sekvensen deres og av celletypen som de uttrykkes i. Noen RNA-er har svært kortvarig levetid, og noen har mye lengre levetid. I tillegg er det en virkning hvis størrelse er uviss, av translasjonseffektivitet på mRNA-halveringstid. I tillegg folder hvert

30 RNA-molekyl seg til en bestemt struktur, eller muligens familie av strukturer, som er bestemt av dets sekvens. Den bestemte struktur til enhver RNA kan føre til større eller mindre stabilitet i cytoplasmaet. Struktur per se er sannsynligvis også en determinant for mRNA-prosessering i

35 kjernen. Dessverre er det umulig å forutsi, og nesten umulig å bestemme, strukturen til enhver RNA (bortsett fra tRNA) in vitro eller in vivo. Det er imidlertid sannsynlig at dramat-

iske endringer i sekvensen til RNA vil ha en stor effekt på dens foldede struktur. Det er sannsynlig at strukturen per se eller særlig strukturtrekkene, også spiller en rolle når det gjelder å bestemme RNA-stabilitet.

Noen bestemte sekvenser og signaler er blitt identifisert i RNA-er som har mulighet for å ha en bestemt effekt på RNA-stabilitet. Dette avsnittet oppsummerer det som er kjent om disse sekvensene og signalene. Disse identifiserte sekvenser er ofte A+T-rike og vil følgelig mer sannsynlig opptre i en A+T-rik, kodende sekvens, slik som et B.t.-gen. Sekvensen "motif"-ATTTA (eller AUUUA som den fremstår i RNA) har vært innblandet som en destabiliserende sekvens i pattedyr-celle-mRNA (Shaw and Kamen, 1986). Det har ikke vært gjort noen analyse av funksjonen til denne sekvensen i planter. Mange kortlivede mRNA-er har A+T-rike 3'-utranslaterte områder, og disse områdene har ofte ATTTA-sekvensen, noen ganger til stede i flere kopier eller som multimerer (f.eks. ATTTATTTA...). Shaw and Kamen viste at overføringen av 3'-enden av ustabil mRNA til en stabil RNA (globin eller VAL) reduserte halveringstiden til den stabile RNA dramatisk. De viste videre at en pentamer av ATTTA hadde en inngående destabiliserende effekt på et stabilt budskap, og at dette signalet kunne utøve sin effekt uansett om det var lokalisert i 3'-enden eller i den kodende sekvens. Antallet av ATTTA-sekvenser og/eller den sekvens-sammenhengen hvor de opptrådte, syntes imidlertid også å være viktig for bestemmelse av om de virker som destabiliserende sekvenser. Shaw and Kamen viste at en trimer av ATTTA hadde mye mindre effekt enn en pentamer på mRNA-stabilitet, og at en dimer eller en monomer ikke hadde noen effekt på stabilitet (Shaw and Kamen, 1987). Det bør legges merke til at multimerer av ATTTA, slik som en pentamer, automatisk skaper et A+T-rikt område. Dette ble vist å være en cytoplasmisk effekt, ikke nukleær. I andre ustabile mRNA-er kan ATTTA-sekvensen være til stede i bare én enkelt kopi, men den er ofte inneholdt i et A+T-rikt område. Ut fra dyrecelledataene som hittil er blitt samlet inn, synes det

- som om ATTTA i det minste i noen sammenhenger er viktig for stabilitet, men det er ennå ikke mulig å forutsi hvilke forekomster av ATTTA som er destabiliserende elementer, eller hvorvidt det er sannsynlig at noen av disse effektene er til stede i planter.

Noen undersøkelser vedrørende mRNA-nedbrytning i dyreceller indikerer også at RNA-nedbrytning i noen tilfeller kan begynne med nukleolytisk angrep i A+T-rike områder. Det er ikke klart om disse spaltningene oppstår i ATTTA-sekvenser. Det finnes også eksempler på mRNA-er som har forskjellig stabilitet avhengig av celletypen som de uttrykkes i, eller stadiet i cellesyklusen hvor de uttrykkes. Histon-mRNA-er er f.eks. stabile under DNA-syntese, men ustabile dersom DNA-syntese avbrytes. 3'-enden til noen histon-mRNA-er synes å være ansvarlig for denne effekten (Pandey and Marzluff, 1987). Den synes ikke å være mediert av ATTTA, det er heller ikke klart hva som regulerer den forskjellige stabilitet til denne mRNA. Et annet eksempel er den forskjellige stabilitet for IgG-mRNA i B-lymfocytter under B-cellemodning (Genovese and Milcarek, 1988). Et siste eksempel er ustabiliteten av en mutant beta-thallemisk globin-mRNA. I benmargceller hvor dette genet normalt uttrykkes, er den mutante mRNA ustabil, mens villtype-mRNA-en er stabil. Når mutantgenet er uttrykt i HeLa- eller L-celler in vitro, utviser den mutante mRNA ingen ustabilitet (Lim et al., 1988). Disse eksemplene gir alle bevis for at mRNA-stabilitet kan medieres av celletype- eller cellesyklusspesifikke faktorer. Denne type av ustabilitet er dessuten enda ikke assosiert med bestemte sekvenser. På bakgrunn av disse usikkerhetene er det ikke mulig å forutsi hvilke RNA-er som sannsynligvis vil være ustabile i en bestemt celle. Selv ATTTA-"motif" kan virke forskjellig avhengig av egenskapene til den celle hvor RNA-en er til stede. Shaw and Kamen (1987) har rapportert at aktivering av protein-kinase C kan blokkere nedbrytning mediert av ATTTA.

Addisjonen av en polyadenylatstreng til 3'-enden er felles for de fleste eukaryote mRNA-er, både hos planter og

• dyr. Det for tiden aksepterte syn på polyA-addisjon er at
det voksende transkript fortsetter forbi 3'-enden til
det fullstendige molekyl. Inneholdt i dette transkript er
signaler for polyadenylering og korrekt 3'-endedannelse.
5 Denne prosessering i 3'-enden omfatter spalting av mRNA-en
og addisjon av polyA til den ferdige 3'-ende. Ved å søke
etter konsensussekvenser nær polyA-området både i plante-
og dyre-mRNA-er, har det vært mulig å identifisere konsensus-
sekvenser som tilsynelatende er involvert i polyA-addisjon
10 og 3'-ende-spalting. De samme konsensussekvenser synes å
være viktige for begge disse prosessene. Disse signalene
er vanligvis en variasjon av sekvensen AATAAA. I dyre-
celler er noen varianter av denne sekvensen som er funk-
sjonelle, blitt identifisert; i planteceller synes det å
15 være et utvidet område med funksjonelle sekvenser (Wickens
and Stephenson, 1984; Dean et al., 1986). På grunn av at
alle disse konsensussekvensene er variasjoner av AATAAA,
har de alle A+T-rike sekvenser. Denne sekvensen finnes
vanligvis 15-20 bp før polyA-området i en ferdig mRNA.
20 Forsøk i dyreceller indikerer at denne sekvensen er invol-
vert både i polyA-addisjon og 3'-ferdigstillelse. Sted-
rettede mutasjoner i denne sekvensen kan ødelegge disse
funksjonene (Conway and Wickens, 1988; Wickens et al.,
1987). Det er imidlertid også blitt observert at sekvenser
25 opptil 50-100 bp 3' i forhold til det putative polyA-
signal også er påkrevet; dvs. et gen som har en normal
AATAAA, men er blitt erstattet eller avbrutt nedstrøms, ikke
blir korrekt polyadenylert (Gil and Proudfoot, 1984;
Sadofsky and Alwine, 1984; McDevitt et al., 1984). Dvs. at
30 polyA-signalet i seg selv ikke er tilstrekkelig for full-
stendig og korrekt prosessering. Det er ennå ikke kjent
hvilke spesifikke, nestrøms sekvenser som er påkrevet i til-
legg til polyA-signalet, eller om det er en bestemt sekvens
som har denne funksjonen. Sekvensanalyse kan derfor bare
35 identifisere potensielle polyA-signaler.

I naturlig forekommende mRNA-er som normalt er
polyadenylerte, er det blitt observert at avbrytelse av

denne prosessen, enten ved å endre polyA-signalet eller andre sekvenser i mRNA-en, kan det oppnås dyptgående effekter på nivået av funksjonell mRNA. Dette er blitt observert i flere naturlig forekommende mRNA-er, med resultater som så langt er genspesifikke. Ennå kan det ikke utledes noen generelle regler fra undersøkelsen av mutanter av disse naturlige genene, og ingen regler som kan anvendes på heterologe gener. Nedenfor er det fire eksempler:

1. I et globingen fører fravær av et korrekt polyA-sete til ukorrekt terminering av transkripsjon. Det er sannsynlig, men ikke vist, at den ukorrekt terminerte RNA er ikke-funksjonell og ustabil (Proudfoot et al., 1987).

2. I et globingen kan fravær av et funksjonelt polyA-signal føre til en 100-ganger reduksjon i nivået av mRNA-akkumulering (Proudfoot et al., 1987).

3. Et globingen-polyA-sete ble plassert inn i 3'-endene til to forskjellige histongener. Histongenene inneholder en sekundærstruktur (stammesløyfe) nær sine 3'-ender. Mengden av korrekt polyadenylert histon-mRNA produsert fra disse kimærene, avtok etter hvert som avstanden mellom stammesløyfen og polyA-setet økte. De to histongenene ga også svært forskjellige nivåer av korrekt polyadenylert mRNA. Dette tyder på en interaksjon mellom polyA-setet og andre sekvenser på mRNA-en som kan modulere mRNA-akkumulering (Pandy and Marzluff, 1987).

4. Soyabønne-leghemoglobingenet er blitt klonet inn i HeLa-celler, og det er blitt bestemt at dette plantegenet inneholder et "kryptisk" polyadenyleringssignal som er aktivt i dyreceller, men ikke utnyttes i planteceller. Dette fører til fremstillingen av en ny polyadenylert mRNA som er ikke-funksjonell. Dette viser igjen at analyse av et gen i én celletype ikke kan forutsi dets oppførsel i andre celletyper (Wiebauer et al., 1988).

Ut fra disse eksemplene er det klart at i naturlige mRNA-er er korrekt polyadenylering viktig i mRNA-akkumulering, og at avbrytelse av denne prosessen kan påvirke mRNA-

• nivåene betydelig. Det foreligger imidlertid utilstrekkelige kunnskaper til å forutsi effekten av endringer i et normalt gen. I et heterologt gen hvor vi ikke vet om de antatte polyA-setene (konsensussekvensene) er funksjonelle, er det enda vanskeligere å forutsi følgene. Det er imidlertid mulig at de identifiserte, antatte setene ikke er funksjonelle. Det vil si at disse setene ikke kan virke som korrekte polyA-seter, men i stedet virke som avvikende seter som gir opphav til ustabile mRNA-er.

I dyrecellesystemer er AATAAA langt det mest vanlige signal identifisert i mRNA-er oppstrøms for polyA, men minst fire varianter er også blitt funnet (Wickens and Stephenson, 1984). I planter har på langt nær så mange analyser blitt gjort, men det er klart at lignende, multiple sekvenser som AATAAA kan anvendes. Plantesetene nedenunder kalt viktige eller uviktige, henviser bare til undersøkelsen til Dean et al. (1986) som analyserte bare tre typer planter. Betegnelsen av polyadenyleringssetene som viktige eller mindre viktige, henviser bare til hyppigheten av deres forekomst som funksjonelle seter i naturlig forekommende gener som er blitt analysert. I tilfellet med planter er dette en svært begrenset database. Det er vanskelig å forutsi med noen sikkerhet om et sete betegnet viktig eller mindre viktig med større eller mindre sikkerhet vil virke delvis eller fullstendig når det finnes i et heterologt gen, slik som B.t.

	PA	AATAAA	Viktig konsensussete
	P1A	AATAAT	Viktig plantesete
30	P2A	AACCAA	Mindre viktig plantesete
	P3A	ATATAA	"
	P4A	AATCAA	"
	P5A	ATACTA	"
	P6A	ATAAAA	"
35	P7A	ATGAAA	"
	P8A	AAGCAT	"
	P9A	ATTAAT	"

	P10A	ATACAT	Mindre viktig plantesete
	P11A	AAAATA	"
	P12A	ATTAAA	Mindre viktig dyresete
	P13A	AATTAA	"
5	P14A	AATACA	"
	P15A	CATAAA	"

En annen type RNA-prosessering som opptrer i kjernen, er intronspleising. Nesten alt arbeidet vedrørende intron-prosessering er blitt gjort i dyreceller, men noen data skriver seg fra planter. Intronprosessering avhenger av korrekt 5'- og 3'-spleisestedsekvenser. Konsensussekvenser for disse spleisestedene er blitt utvunnet både fra dyre- og plante-mRNA-er, men bare noen få nukleotider er kjent for å være invariante. Det er derfor vanskelig å forutsi med noen sikkerhet hvorvidt et antatt spleisested er funksjonelt eller delvis funksjonelt bare basert på sekvensanalyse. Således er de eneste invariante nukleotider GT i 5'-enden til intronet og AG i 3'-enden til intronet. I planter kan ved enhver nærliggende posisjon, enten innenfor intronet eller i eksonet som flankerer intronet, alle fire nukleotidene finnes, selv om noen posisjoner utviser en viss nukleotidpreferanse (Brown, 1986; Hanley and Schuler, 1988).

Et planteintron er blitt flyttet fra et patatinngen til et GUS-gen. For å gjøre dette ble stedrettet mutagenese utført for å innføre nye restriksjonssteder, og denne mutagenesen endret flere nukleotider i intronet og eksonsekvensene som flankerer GT og AG. Dette intronet virker fortsatt korrekt, noe som viser viktigheten av GT og AG, og fleksibiliteten i andre nukleotidposisjoner. Det er selvsagt mange forekomster av GT og AG i alle gener som ikke virker som intronspleisesteder, så det må være noen andre sekvens- eller strukturetrekk som identifiserer spleisesteder. I planter synes ett slikt trekk å være basesammensetning per se. Wiebauer et al. (1988) og Goodall et al. (1988) har analysert planteintroner og -eksoner og funnet at eksoner har

• ca. 50% A+T, mens introner har ca. 70% A+T. Goodall et al. (1988) dannet også et kunstig planteintron som har konsensus-5'- og -3'-spleisesteder og en tilfeldig A+T-rik indre sekvens. Dette intronet ble spleiset korrekt i planter. Når det indre segment ble erstattet med en G+C-rik sekvens, ble spleiseeffektivitet drastisk redusert. Disse to eksemplene viser at introngjenkjennelse i planter kan avhenge av svært generelle trekk, spleisesteder som har en stor andel av sekvensforskjellighet, og A+T-rikelighet i selve intronet. Dette gjør det selvsagt vanskelig å forutsi ut fra sekvens alene hvorvidt det er sannsynlig at noen bestemt sekvens virker som et aktivt eller delvis aktivt intron for RNA-prosessering.

B.t.-gener som er A+T-rike, inneholder et stort antall strekninger med forskjellige lengder som har 70% eller mer A+T. Antallet av slike strekninger identifisert ved hjelp av sekvensanalyse, avhenger av lengden av sekvensen som er undersøkt.

Som for polyadenylering beskrevet ovenfor, er det komplikasjoner når det gjelder å forutsi hvilke sekvenser som kan benyttes som spleiseseter i et bestemt gen. For det første har mange naturlig forekommende gener alternative spleisespor som skaper alternative kombinasjoner av eksoner i slutt-mRNA-en (Gallega and Nadal-Ginard, 1988; Helfman and Ricci, 1988; Tsurushita and Korn, 1989). Det vil si at noen spleisesteder åpenbart gjenkjennes under visse omstendigheter eller i bestemte celletyper, men ikke i andre. Reglene som styrer dette, er ikke forstått. I tillegg kan det være en interaksjon mellom prosesseringsspor, slik at benyttelse av et bestemt polyadenyleringssete kan virke inn på spleising i et nærliggende spleisesete og vice versa (Adami and Nevins, 1988; Brady and Wold, 1988; Marzluff and Pandey, 1988). Igjen er det ikke tilgjengelig noen regler som forutsier resultatet. Sekvensendringer i et gen kan også endre utnyttelsen av bestemte spleisesteder drastisk. I et bovint veksthormongen forårsaker f.eks. små delesjoner i et ekson noen få hundre baser nedstrøms for et intron at

spleiseeffektiviteten til intronet faller fra mer enn 95% til mindre enn 2% (i det vesentlige ikke-funksjonelt).

Andre delesjoner har imidlertid praktisk talt ingen effekt (Hampson and Rottman, 1988). Endelig indikerer flere for-

skjellige in vitro og in vivo forsøk at mutasjoner som ødelegger normal spleising, fører til hurtig nedbrytning av RNA-en i kjernen. Spleising er en flertrinnsprosess i kjernen, og mutasjoner ved normal spleising kan føre til blokkeringer i prosessen på flere forskjellige trinn.

Hvilken som helst av disse blokkeringene kan så føre til en unormal og ustabil RNA. Undersøkelser av mutanter av normalt prosesserte (polyadenylering og spleising) gener er relevante for undersøkelsen av heterologe gener, slik som B.t. B.t.-gener kan inneholde funksjonelle signaler som fører til produksjonen av avvikende, ikke-funksjonelle mRNA-er, og disse mRNA-ene vil sannsynligvis være ustabile. Men det er kanskje enda mer sannsynlig at B.t.-genene inneholder signaler som er analoge med mutantsignalene i et naturlig gen. Som vist ovenfor, er det svært sannsynlig at disse mutantsignalene forårsaker defekter i prosesseringssporene som har som konsekvens at det produseres ustabile mRNA-er.

Det er ikke kjent med noen sikkerhet hva som signaliserer RNA-transkripsjonsterminering i plante- eller dyreceller. Noen undersøkelser på dyregener indikerer at sekvensstrekninger rike på T, forårsaker terminering ved hjelp av kalvetymus-RNA-polymerase II in vitro. Disse undersøkelsene har vist at 3'-endene i in vitro terminerte transkripter ofte ligger innenfor forekomster av T, slik som T5, T6 eller T7. Andre identifiserte seter har ikke vært sammensatt bare av T, men har hatt ett eller flere andre nukleotider også. Terminering er blitt funnet å forekomme innenfor sekvensene TATTTTTT, ATTCTC, TTCTT (Dedrick et al., 1987; Reines et al., 1987). I tilfellet med de to sistnevnte har den sammenhengen hvor sekvensen er funnet, også vært C+T-rik. Det er ikke kjent om dette er vesentlig. Andre undersøkelser har utpekt A-strekninger som potensielle

transkripsjonsterminatorer. Et interessant eksempel fra SV40 illustrerer uvissheten når det gjelder å definere terminatorer basert på sekvens alene. En potensiell terminator i SV40 ble identifisert til å være A-rik og ha et område med dyadsymmetri (potensiell stammesløyfe) 5' til den A-rike strekning. En andre terminator identifisert eksperimentelt nedstrøms i det samme gen, var imidlertid ikke A-rik og omfattet ingen potensiell sekundærstruktur (Kessler et al., 1988). På grunn av A+T-innholdet i B.t.-gener er de selvsagt rike på forekomster av A eller T som kunne virke som terminatorer. Viktigheten av terminering for stabilitet av mRNA-en er vist ved hjelp av globingen-eksemplet beskrevet ovenfor. Fravær av et normalt polyA-sete fører til en svikt i korrekt terminering med en reduksjon i mRNA som følge.

Det er også en effekt på mRNA-stabilitet på grunn av translasjonen av mRNA-en. For tidlig translasjonsterminering i human triosefosfatisomerase fører til ustabilitet hos mRNA-en (Daar et al., 1988). Et annet eksempel er beta-tallesemin-globin-mRNA-en beskrevet ovenfor som er spesifikt ustabil i benmargceller (Lim et al., 1988). Effekten i dette mutantgenet er en enkel basepardelesjon i kodon 44 som fører til translasjonsterminering (et nonsens-kodon) i kodon 60. Sammenlignet med korrekt translatert, normalt globin-mRNA er denne mutant-RNA svært ustabil. Disse resultatene indikerer at ukorrekt translatert mRNA er ustabil. Annet arbeide i gjær indikerer at korrekt, men dårlig translasjon, kan ha en effekt på mRNA-nivåer. Et heterologt gen ble modifisert for å omdanne bestemte kodoner til mer gjerne foretrukne kodoner. En totalt 10-gangers økning i proteinproduksjon ble oppnådd, men det var også en ca. 3-gangers økning i mRNA (Hoekema et al., 1987). Dette indikerer at mer virkningsfull translasjon kan føre til større mRNA-stabilitet, og at effekten av kodonbruk like gjerne kan være på RNA-nivået som på translasjonsnivået. Det er ikke klart fra kodonbrukundersøkelser hvilke kodoner som fører til dårlig translasjon, eller hvordan disse er forbundet til mRNA-stabilitet.

Det er derfor et formål ved foreliggende oppfinnelse å tilveiebringe modifiserte, kinnære gener som koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein. Videre er det et formål ved oppfinnelsen å tilveiebringe en plantetransformasjonsvektor som inneholder et plantegen som inneholder et slikt modifisert, kinnært gen.

Kort beskrivelse av tegningene

Figur 1 illustrerer trinnene som anvendes ved modifisering av et villtypegen for å øke ekspresjonseffektivitet i planter.

Figur 2 viser en sammenligning mellom endringene i den modifiserte B.t.k. HD-1-sekvens ifølge eksempel 1 (nedre linje) versus villtypesekvensen til B.t.k. HD-1 som koder for krystallproteintoksinet (øvre linje).

Figur 3 viser en sammenligning mellom endringene i den syntetiske B.t.k. HD-1-sekvens ifølge eksempel 2 (nedre linje) versus villtypesekvensen til B.t.k. HD-1 som koder for krystallproteintoksinet (øvre linje).

Figur 4 viser en sammenligning mellom endringene i den syntetiske B.t.k. HD-73-sekvens ifølge eksempel 3 (nedre linje) versus villtypesekvensen til B.t.k. HD-73 (øvre linje).

Figur 5 er et plasmidkart for intermediær plantetransformasjonsvektorkassett pMON893.

Figur 6 er et plasmidkart for intermediær plantetransformasjonsvektorkassett pMON900.

Figur 7 er et kart for den uskadeliggjorte T-DNA til *A. tumefaciens* ACO.

Figur 8 viser en sammenligning mellom endringene i det syntetiske, avkortede B.t.k. HD-73-gen (aminosyrene 29-615 med en N-terminal Met-Ala) ifølge eksempel 3 (nedre linje) versus villtypesekvensen til B.t.k. HD-73 (øvre linje).

Figur 9 viser en sammenligning mellom endringene i

- den syntetiske/villtype fullengde B.t.k. HD-73-sekvens ifølge eksempel 3 (nedre linje) versus villtype-fullengde-sekvensen til B.t.k. HD-73 (øvre linje).

5 Figur 10 viser en sammenligning mellom endringene i den syntetiske/modifiserte fullengde B.t.k. HD-73-sekvens ifølge eksempel 3 (nedre linje) versus villtype-fullengde-sekvensen til B.t.k. HD-73 (øvre linje).

10 Figur 11 viser en sammenligning mellom endringene i den helt syntetiske fullengde B.t.k. HD-73-sekvens ifølge eksempel 3 (nedre linje) versus villtype-fullengdesekvensen til B.t.k. HD-73 (øvre linje).

15 Figur 12 viser en sammenligning mellom endringene i den syntetiske B.t.t.-sekvens ifølge eksempel 5 (nedre linje) versus villtypesekvensen til B.t.t. som koder for krystallproteintoksinet (øvre linje).

20 Figur 13 viser en sammenligning mellom endringene i den syntetiske B.t. P2-sekvens ifølge eksempel 6 (nedre linje) versus villtypesekvensen til B.t.k. HD-1 som koder for P2-proteintoksinet (øvre linje).

25 Figur 14 viser en sammenligning mellom endringene i den syntetiske B.t. entomocidus-sekvens ifølge eksempel 7 (nedre linje) versus villtypesekvensen til B.t. entomocidus som koder for Btent-proteintoksinet (øvre linje).

30 Figur 15 viser et plasmidkart for planteekspresjons-kassettevektor pMON744.

Figur 16 viser en sammenligning mellom endringene i den syntetiske potetblad-"rulle"-virus- (PLRV) kappeprotein-sekvens ifølge eksempel 9 (nedre linje) versus villtype-kappeproteinsekvensen til PLRV (øvre linje).

Angivelse av oppfinnelsen

35 Foreliggende oppfinnelse tilveiebringer et modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylnukleotider til 3'-enden av RNA-en bevirkes. Genet er

kjennetegnet ved at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som at-

5 koder for *Bacillus thuringiensis*-proteinet, og har de følgende kjennetegn:

- 10 a) hvilke som helst 15 nukleotider i den strukturkodende sekvens som inneholder mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 15 nukleotidene to polyadenyleringssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- 15 c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens, eller
- a) den strukturkodende sekvens inneholder ikke noen polyadenyleringssekvens innenfor 10 nukleotider i hvilken som helst annen polyadenyleringssekvens,
- 20 b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens, eller
- 25 a) den strukturkodende sekvens som inneholder ikke noen polyadenyleringssekvens innenfor 10 nukleotider i hvilken som helst annen polyadenyleringssekvens,
- 30 b) hvilke som helst 30 nukleotider i den strukturkodende sekvens som inneholder: (1) mer enn ca. 80% adenin- og tyminrester, og (2) mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 30 nukleotidene noen polyadenyleringssekvens,
- 35 c) den strukturkodende sekvens inneholder ikke en

AATAAA- eller AATAAT-DNA-sekvens; og

- d) hvilke som helst 30 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 30 nukleotidene mer enn én ATTTA-sekvens, eller
- 5 a) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 15 nukleotidene to
- 10 polyadenyleringssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens innenfor 5'-halvdelen av den strukturkodende sekvens, og
- c) hvilke som helst 15 nukleotider i 5'-halvdelen
- 15 av den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens, eller
- a) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens
- 20 ikke noen polyadenyleringssekvens innenfor 10 nukleotider i hvilken som helst annen polyadenyleringssekvens,
- b) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens
- 25 ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens, eller
- 30 a) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke noen polyadenyleringssekvens innenfor 10 nukleotider i noen annen polyadenyleringssekvens,
- 35 b) hvilke som helst 30 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder:
- (1) mer enn ca. 80% adenin- og tyminrester; og
- (2) mer enn 4 på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester,

inneholder ikke innenfor de 30 nukleotidene noen polyadenyleringssekvens,

- c) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- d) hvilke som helst 30 nukleotider i 5'-halvdelen av den strukturkodende sekvens inneholder ikke innenfor de 30 nukleotidene mer enn én ATTTA-sekvens, eller

1) den strukturkodende sekvens har et område som er i det vesentlige homologt med den følgende sekvens:

```
TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC
1   5   10   15   20   25   30   35
```

hvor sekvensen ikke inneholder minst én av gruppen bestående av et plantepolyadenyleringssignal og en ATTTA-sekvens, eller

2) hvor den strukturkodende sekvens har et område som er i det vesentlige homologt med den følgende sekvens

```
TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC
1   5   10   15   20   25   30   35
```

med minst én modifikasjon valgt fra gruppen bestående av:

nukleotid 1 er cytosin (C),
 nukleotid 3 er cytosin (C),
 nukleotid 6 er cytosin (C),
 nukleotid 12 er guanin (G),
 nukleotid 18 er cytosin (C),
 nukleotid 24 er guanin (G), og
 nukleotid 36 er tymin (T), eller

3) hvor den naturlig forekommende DNA-sekvens omfatter et område med den følgende sekvens

```
TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC
1   5   10   15   20   25   30   35
```

og hvor den strukturkodende sekvens omfatter modifikasjoner slik at minst nevnte område inneholder minst én færre sekvens valgt fra gruppen bestående av en AACCAA- og en AATTAA-sek-

vens, eller

har minst ett kjennetegnvalgt fra gruppen bestående av:

- a) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder:
 - (1) mer enn ca. 80% adenin- og tyminrester, og
 - (2) mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 15 nukleotidene to plantepolyadenyleringssekvenser,
 - b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
 - c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens, eller
- den strukturkodende sekvens inneholder ikke mer enn 5 på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester.

Videre tilveiebringer oppfinnelsen en plantetransformasjonsvektor som inneholder et plantegen som inneholder et modifisert, kimært gen inneholdende en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes. Vektoren er videre kjennetegnet ved at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og har de følgende kjennetegn:

- a) hvilke som helst 15 nukleotider i den strukturkodende sekvens som inneholder mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 15 nukleotidene to polyadenyleringssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og

- c) hvilke som helst 15 nukleotider i den struktur-
kodende sekvens inneholder ikke innenfor de
15 nukleotidene mer enn én ATTTA-sekvens.

For korthet og klarhet i beskrivelsen, vil fore-
5 liggende beskrivelse primært bli beskrevet med hensyn til
fremstillingen av syntetiske plantegener som koder for
krystallproteintoksinet til *Bacillus thuringiensis* (B.t.).
Egnede B.t.-underarter omfatter, men er ikke begrenset til,
B.t. kurstaki HD-1, B.t. kurstaki HD-73, B.t. sotto,
10 B.t. berliner, B.t. thuringiensis, B.t. tolworthi,
B.t. dendrolimus, B.t. alesti, B.t. galleriae, B.t. aizawai,
B.t. subtoxicus, B.t. entomocidus, B.t. tenebrionis og
B.t. san diego. Fagfolk innen teknikken vil imidlertid
forstå, og det skal forstås, at foreliggende fremgangsmåte
15 kan anvendes til å fremstille syntetiske plantegener som
koder for andre ikke-planteproteiner enn krystallprotein-
toksinet til B.t., samt planteproteiner (se f.eks. eksempel
9).

Ekspresjonen av B.t.-gener i planter er problematisk.
20 Selv om ekspresjonen av B.t.-gener i planter er blitt rappor-
tert på insekticidnivåer, har ikke oppnåelsen av dette skjedd
lettvtint. Særlig har ekspresjonen av et fullengde lepi-
dopteran-spesifikt B.t.-gen (omfattende DNA fra et B.t.k.-
isolat) blitt rapportert å være mislykket når det gjelder å
25 gi insekticide nivåer av ekspresjon i noen plantearter
(Vaeck et al., 1987 og Barton et al., 1987).

Det er blitt rapportert at ekspresjon av full-
lengdegenet fra B.t.k. HD-1 var påvisbart i tomatplanter,
men at avkortede gener førte til en høyere frekvens av
30 insekticide planter med et totalt høyere ekspresjonsnivå.
Avkortede gener fra B.t. berliner førte også til en høyere
hyppighet av insekticide planter i tobakk (Vaeck et al.,
1987). På den annen side ble det tilveiebrakt insekticide
planter fra salattransformanter under anvendelse av et full-
35 lengdegen.

Det er også blitt rapportert at fullengdegenet fra
B.t.k. HD-73 ga en viss insekticid effekt i tobakk (Adang
et al., 1987). B.t. mRNA påvist i disse plantene, var
imidlertid bare 1,7 kb sammenlignet med de forventede 3,7 kb,

• noe som indikerer ukorrekt ekspresjon av genet. Det ble antatt at denne forkortede mRNA var for kort til å kode for et funksjonelt, avkortet toksin, men det må ha vært et lavt nivå av lengre mRNA i noen planter eller så ville ikke noen insekticid aktivitet blitt observert. Andre har rapportert i en publikasjon at de observerte en stor mengde kortere enn forventet mRNA fra et avkortet B.t.k.-gen, men noe mRNA av den forventede størrelse ble også observert. Faktisk ble det foreslått at ekspresjon av fullengdegenet er toksisk for tobakk-kallus (Barton et al., 1987). Det ovenfor nevnte illustrerer at B.t.-gener av sommerfugltype uttrykkes dårlig i planter sammenlignet med andre kimærgener som tidligere er uttrykt fra de samme promoter-kassetter.

15 Ekspresjonen av B.t.t. i tomat og potet skjer på lignende nivåer som i B.t.k. (dvs. lave). B.t.t.- og B.t.k.-gener har bare til felles begrenset sekvenshomologi, men de har mange fellestrekk når det gjelder basesammensetning og tilstedeværelsen av bestemte A+T-rike elementer.

20 Alle rapporter på området har bemerket den lavere enn forventede ekspresjon av B.t.-gener i planter. Vanligvis har insekticid effektivitet blitt målt ved å bruke insekter som er svært følsomme for B.t.-toksin, slik som tobakk-hornlarve. Selv om det har vært mulig å oppnå planter som er fullstendig beskyttet mot tobakkhornlarve, er det viktig å legge merke til at hornlarve er opptil 500 ganger mer sensitiv for B.t.-toksin enn noen agronomisk viktige skadeinsekter, slik som sukkerroehærklarve. Det er derfor av interesse å oppnå transgene planter som er beskyttet mot alle viktige sommerfugl-skadeinsekter (eller mot Colorado-potetbille i tilfellet med B.t. tenebrionis), og i tillegg ha et nivå av B.t.-ekspresjon som gir en ytterligere sikkerhetsmargin i forhold til og over det virkningsfulle beskyttelsesnivå. Det er også viktig å an vise plantegener som virker reproduserbart fra art til art, slik at insektresistente planter kan oppnås på en forutsigbar måte.

35 For å nå disse mål er det viktig å forstå naturen av

den dårligere enn forventede ekspresjon av B.t.-gener i planter. Nivået av stabil B.t.-mRNA i planter er mye lavere enn forventet. Dvs. at sammenlignet med andre kodende sekvenser som drives av den samme promoter, er nivået av B.t.-mRNA målt ved hjelp av Northern-analyse eller nukleasebeskyttelseforsøk, mye lavere. F.eks. ble tomatplante 337 (Fischhoff et al., 1987) valgt ut som den best uttrykkende plante med pMON9711 som inneholder B.t.k. HD-1 KpnI-fragmentet drevet av CaMV 35S-promoter og som inneholder det NOS-NPTII-NOS-selekterbare markørgen. I denne planten er nivået av B.t.-mRNA mellom 100 og 1000 ganger lavere enn nivået av NPTII-mRNA, selv om 35S-promoter er omtrent 50 ganger sterkere enn NOS-promoter (Sanders et al., 1987).

Nivået av B.t.-toksinprotein påvist i planter, er i overensstemmelse med det lave nivå av B.t.-mRNA. Den insekticide virkningsfullhet av de transgene planter korrelerer dessuten med B.t.-proteinnivået, noe som indikerer at toksinproteinet produsert i planter, er biologisk aktivt. Det lave nivå av B.t.-toksinekspresjon kan derfor være resultatet av de lave nivåer av B.t.-mRNA.

mRNA-nivåer bestemmes ved hjelp av syntese- og nedbrytningshastigheten. Det er balansen mellom disse to som bestemmer likevektsnivået for mRNA. Syntesehastigheten er blitt maksimalisert ved bruk av CaMV 35S-promoter, en sterk, konstitutiv, planteuttrykkbar promoter. Bruken av andre plantepromoterer, slik som nopalinsyntase (NOS), mannopinsyntase (MAS) og ribulose-bisfosfatcarboxylase-små underenhet (RUBISCO), har ikke ført til dramatiske endringer i nivåene av B.t.-toksinproteinekspresjon, noe som tyder på at effektene av å bestemme B.t.-toksinprotein-nivåer er promoteruavhengige. Disse dataene innebærer at de kodende sekvensene av DNA-gener som koder for B.t.-toksinproteiner, på en eller annen måte er ansvarlige for det lave ekspresjonsnivå, og at denne effekten manifesteres på et lavt nivå av akkumulert, stabil mRNA.

Lavere enn forventede nivåer av mRNA er blitt obser-

• vert hos fire forskjellige sommerfugl-spesifikke gener (to
fra B.t.k. HD-1; B.t. berliner og B.t.k. HD-73) samt
genet fra den bille-spesifikke B.t. tenebrionis. For
B.t.-gener av sommerfugltype synes det som om disse virk-
ningene er manifestert sterkere i de kodende sekvenser med
5 full lengde, enn i de avkortede, kodende sekvenser. Disse
virkningene ses over plantearter selv om størrelsen synes
å være større i noen plantearter, slik som tobakk.

Egenskapene til de kodende sekvenser i B.t.-gener
10 adskiller dem fra plantegener samt mange andre heterologe
gener uttrykt i planter. Særlig er B.t.-gener svært rike
(ca. 62%) på adenin (A) og thymin (T), mens plantegener og
de fleste bakteriegener som er blitt uttrykt i planter, er
i størrelsesorden 45-55% A+T. A+T-innholdet i genomene
15 (og derved genene) til enhver organisme er trekk ved organ-
ismen og gjenspeiler dens utviklingsmessige historie. Mens
gener innenfor én organisme har likt A+T-innhold, kan A+T-
innholdet variere veldig mye fra organisme til organisme.
F.eks. har noen Bacillus-arter blant de mest A+T-rike
20 genomer, mens noen Streptomyces-arter har blant de minst
A+T-rike genomer (ca. 30 til 35% A+T).

På grunn av den genetiske kodes degenererbarhet, og
det begrensede antall av kodonvalg for enhver aminosyre,
finnes mesteparten av det "overskytende" A+T i de struktur-
25 kodende sekvenser til noen Bacillus-arter i kodonenes
tredje posisjon. Det vil si at gener i noen Bacillus-arter
har A eller T som det tredje nukleotid i mange kodoner.
A+T-innhold kan således delvis bestemme kodonbruktilbøyelig-
het. I tillegg er det klart at gener utvikles for maksimal
30 funksjon i organismen som de utvikles i. Dette betyr at
særlig nukleotidsekvenser funnet i et gen fra én organisme
hvor de ikke behøver å spille noen annen rolle enn å kode
for en bestemt strekning av aminosyrer, har mulighet for å
bli gjenkjent som genkontrollelementer i en annen organisme
35 (slik som transkripsjonspromoterer eller -terminatorer,
polyA-addisjonsseter, intronspleiseseter eller spesifikke
mRNA-nedbrytningssignaler). Det er kanskje overraskende at

- slike feilleste signaler ikke er et mer vanlig trekk ved heterolog genekspresjon, men dette kan forklares delvis ved det relativt homogene A+T-innhold (ca. 50%) i mange organismer. Dette A+T-innhold pluss egenskapene til den genetiske kode setter klare begrensninger på forekomstsannsynligheten for enhver bestemt oligonukleotidsekvens. Det er således mye mindre sannsynlig at et gen fra *E. coli* med et 50% A+T-innhold skal inneholde noe bestemt A+T-rikt segment enn et gen fra *B. thuringiensis*.

10 Som beskrevet ovenfor, har ekspresjonen av B.t.-toksinprotein i planter vært problematisk. Selv om observasjonene gjort i andre systemer beskrevet ovenfor, gir et håp om en måte for å forhøye ekspresjonsnivået av B.t.-toksinproteiner i planter, er den suksess som er blitt oppnådd ved hjelp av foreliggende fremgangsmåte, helt uventet. Faktisk ville man, ettersom det nylig er blitt rapportert at ekspresjon av fullengde-B.t.k.-toksinproteinet i tobakk gjør kallusvev nekrotisk (Barton et al., 1987), med rimelighet forvente at høynivåekspresjon av B.t.-toksinprotein 15 ville være uoppnåelig på grunn av de rapporterte toksisitets-effekter.

I sin mest rigorøse anvendelse omfatter fremgangsmåten modifikasjon av en eksisterende, strukturkodende sekvens ("strukturen") som koder for et 25 bestemt protein, ved å fjerne ATTTA-sekvenser og antatte polyadenyleringssignaler ved stedrettet mutagenese av DNA-en omfattende strukturenet. Det er mest foretrukket at praktisk talt alle polyadenyleringssignalene og ATTTA-sekvensene fjernes, selv om forøkte ekspresjonsnivåer observeres med 30 bare delvis fjerning av en av de ovenfor identifiserte sekvenser. Dersom et syntetisk gen fremstilles som koder for ekspresjonen av det aktuelle protein, velges kodoner alternativt for å unngå ATTTA-sekvensen og de antatte polyadenyleringssignaler. For formålene ved foreliggende oppfinnelse 35 omfatter antatte polyadenyleringssignaler, men er ikke nødvendigvis begrenset til, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, AACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT,

- AAAATA, ATTTAA, AATTAA, AATACA og CATAAA. Ved utbygging av ATTTA-sekvensene og polyadenyleringssignalene benyttes fortrinnsvis kodoner slik at man unngår kodonene som sjelden finnes i plantegenomer.

5 Ved en annen utførelsesform vist i flytdiagrammet i figur 1 anvendes en fremgangsmåte for modifikasjon av et eksisterende strukturgren eller alternativt de novo syntese av et strukturgren, idet fremgangsmåten er noe mindre rigorøs enn den ovenfor først
10 beskrevne fremgangsmåte. Idet det vises til figur 1, gjennomføres først den utvalgte DNA-sekvens for å identifisere områder med mer enn fire på hverandre følgende adenin- (A) eller thymin- (T) nukleotider. A+T-områdene gjennomføres med hensyn på mulige plantepolyadenyleringssignaler. Selv om fraværet av fem eller flere på hverandre
15 følgende A- eller T-nukleotider eliminerer de fleste plantepolyadenyleringssignaler dersom mer enn ett av de mindre viktige polyadenyleringssignalene er identifisert innenfor ti nukleotider fra hverandre, så endres fortrinnsvis
20 nukleotidsekvensen i dette område for å fjerne disse signalene, mens den opprinnelige, innkodede aminosyresekvens beholdes.

Det andre trinn er å vurdere de 15-30 nukleotid-områdene som omgir det A+T-rike område identifisert i
25 trinn 1. Dersom A+T-innholdet i det omgivende område er mindre enn 80%, bør området undersøkes med hensyn på polyadenyleringssignaler. Endring av området basert på polyadenyleringssignaler er avhengig av (1) antallet polyadenyleringssignaler som er til stede, og (2) tilstedeværelse av
30 et viktig plante-polyadenyleringssignal.

Det utvidede område undersøkes med hensyn på tilstedeværelsen av plante-polyadenyleringssignaler. Polyadenyleringssignalene fjernes ved stedrettet mutagenese av DNA-sekvensen. Det utvidede område undersøkes også med
35 hensyn på multiple kopier av ATTTA-sekvensen som også fjernes ved mutagenese.

Det er også foretrukket at områder som omfatter

- mange på hverandre følgende A+T-baser eller G+C-baser, brytes opp ettersom disse områdene forutses å ha en høyere sannsynlighet for å danne hårnålstruktur på grunn av selv-komplementærhet. Innføyelse av heterogene basepar ville derfor redusere sannsynligheten for selv-komplementær sekundærstrukturdannelse som er kjent for å inhibere transkripsjon og/eller translasjon i noen organismer. I de fleste tilfeller kan de skadelige effektene minimaliseres ved å bruke sekvenser som ikke inneholder mer enn fem på hverandre følgende A+T eller G+C.

Syntetiske oligonukleotider for mutagenese

Oligonukleotidene som anvendes ved mutagenesen, utformes for å opprettholde den korrekte aminosyresekvens og avlesningsramme, og for fortrinnsvis ikke å innføre slike vanlige restriksjonsseter som BglIII, HindIII, SacI, KpnI, EcoRI, NcoI, PstI og SalI i det modifiserte gen. Disse restriksjonssetene finnes i multilinker-innføysesseter hos slike kloningsvektorer som plasmidene pUC118 og pMON7258. Innføringen av nye polyadenyleringssignaler, ATTTA-sekvenser eller på hverandre følgende strekninger av mer enn fem A+T eller G+C, bør selvsagt også unngås. Den foretrukne størrelse for oligonukleotidene er omkring 40-50 baser, men fragmenter i området fra 18 til 100 baser er blitt benyttet. I de fleste tilfeller opprettholdes et minimum på 5-8 basepar med homologi til templat-DNA-en på begge endene av det syntetiserte fragment for å sikre korrekt hybridisering av primeren til templatet. Oligonukleotidene bør unngå lengre sekvenser enn 5 basepar A+T eller G+C. Kodoner brukt ved erstatningen av villtype-kodoner, bør fortrinnsvis unngå TA- eller CG-dubletten når det er mulig. Kodoner velges fra en tabell over foretrukne kodoner for planter (slik som tabell I nedenunder) på en slik måte at man unngår kodoner som sjelden finnes i plante-genomer, og man bør være påpasselig med å velge kodoner slik at G+C-innholdet fortrinnsvis reguleres til ca. 50%.

Tabell IForetrukket kodonbruk i planter

	<u>Aminosyre</u>	<u>Kodon</u>	<u>Prosentvis anvendelse i planter</u>
5			
	ARG	CGA	7
		CGC	11
		CGG	5
10		CGU	25
		AGA	29
		AGG	23
	LEU	CUA	8
		CUC	20
		CUG	10
15		CUU	28
		UUA	5
		UUG	30
	SER	UCA	14
		UCC	26
		UCG	3
20		UCU	21
		AGC	21
		AGU	15
	THR	ACA	21
		ACC	41
		ACG	7
		ACU	31
25	PRO	CCA	45
		CCC	19
		CCG	9
		CCU	26
	ALA	GCA	23
30		GCC	32
		GCG	3
		GCU	41

Tabell I - forts.Foretrukket kodonbruk i planter

	<u>Aminosyre</u>	<u>Kodon</u>	<u>Prosentvis anvendelse i planter</u>
5			
	GLY	GGA	32
		GGC	20
		GGG	11
		GGU	37
10	ILE	AUA	12
		AUC	45
		AUU	43
	VAL	GUA	9
		GUC	20
		GUG	28
15		GUU	43
	LYS	AAA	36
		AAG	64
	ASN	AAC	72
		AAU	28
20	GLN	CAA	64
		CAG	36
	HIS	CAC	65
		CAU	35
	GLU	GAA	48
25		GAG	52
	ASP	GAC	48
		GAU	52
	TYR	UAC	68
		UAU	32
30	CYS	UGC	78
		UGU	22
	PHE	UUC	56
		UUU	44
35	MET	AUG	100
	TRP	UGG	100

• Områder med mange på hverandre følgende A+T-baser eller G+C-baser forutsies å ha en større sannsynlighet for å danne hårnålstrukturer på grunn av selv-komplementærheten. Avbrytelse av disse områdene med innføyelsen av heterogene basepar er foretrukket og bør redusere sannsynligheten for dannelsen av slike selv-komplementære sekundærstrukturer som hårnåler som i noen organismer er kjent for å inhibere transkripsjon (transkripsjonsterminatorer) og translasjon (attenuatorer). Det er imidlertid vanskelig å forutsi den biologiske effekten av et potensielt hårnåldannende område.

For fagfolk innen teknikken er det klart at selv om beskrivelsen ovenfor er rettet mot modifikasjonen av DNA-sekvensene i villtypegener, kan foreliggende fremgangsmåte anvendes til å konstruere et fullstendig syntetisk gen for en bestemt aminosyresekvens. Områder med fem eller flere på hverandre følgende A+T- eller G+C-nukleotider bør unngås. Kodoner bør velges slik at man unngår TA- og CG-dublettene i kodoner når dette er mulig. Kodonbruk kan normaliseres mot en tabell over foretrukket kodonbruk i planter (slik som tabell I) og G+C-innholdet fortrinnsvis reguleres til ca. 50%. Den resulterende sekvens bør undersøkes for å sikre at det er minimalt med antatte plante-polyadenyleringssignaler og ATTTA-sekvenser. Restriksjonsseter funnet i vanlig brukte kloningsvektorer, unngås også fortrinnsvis. Plassering av flere unike restriksjonsseter gjennom hele genet er imidlertid nyttig for analyse av gen-ekspresjon eller konstruksjon av genvarianter.

Plantegenkonstruksjon

Ekspresjonen av et plantegen som foreligger i dobbelt-trådet DNA-form, omfatter transkripsjon av "budbringer"-RNA (mRNA) fra en av DNA-trådene ved hjelp av RNA-polymerase-enzym, og den etterfølgende prosessering av mRNA-primærtranskriptet inne i kjernen. Denne prosessering omfatter et 3'-ikke-translatert område som adderer polyadenylatnukleotider til 3'-enden av RNA-en. Transkripsjon av DNA til mRNA reguleres ved hjelp av et område i DNA som vanligvis henvises

- til som "promoteren". Promoterområdet inneholder en sekvens av baser som gir RNA-polymerase signal om å binde seg til DNA-en og starte opp transkripsjonen av mRNA under anvendelse av én av DNA-trådene som et templat, for å lage en
5 tilsvarende tråd av RNA.

Et antall promoterer som er aktive i planteceller, er blitt beskrevet i litteraturen. Disse omfatter nopalinsyntase (NOS) og octopinsyntase- (OCS) promoterer (som bæres på tumorinduserende plasmider fra *Agrobacterium tumefaciens*), Cauliflower Mosaic Virus (CaMV) 19S- og 35S-promoterne den lett induserbare promoter fra små underenheten til ribulose-bis-fosfatcarboxylat (ssRUBISCO, et svært rikelig forekommende plantepolypeptid) og mannopinsyntase- (MAS) promoteren (Velten et al., 1984, og Velten & Schell, 1985). Alle disse promoterne er blitt brukt til å danne forskjellige typer av DNA-konstruksjoner som er blitt uttrykt i planter (se f.eks. PCT-publikasjon WO84/02913 (Rogers et al., Monsanto)).

Promoterer som er kjent for, eller er funnet, å forårsake transkripsjon av RNA i planteceller, kan anvendes i foreliggende oppfinnelse. Slike promoterer kan fås fra planter eller plantevirus og omfatter, men er ikke begrenset til, CaMV35S-promoteren og promoterer isolert fra plantegener, slik som ssRUBISCO-gener. Som beskrevet nedenunder, er det foretrukket at den bestemte promoter som velges ut, bør være i stand til å forårsake tilstrekkelig ekspresjon slik at det fås ved fremstilling av en effektiv mengde protein.

Promoterne som brukes i DNA-konstruksjonene (dvs. kimærplantegener) ifølge foreliggende oppfinnelse, kan, om ønsket, modifiseres for å påvirke deres kontrollkarakteristika. F.eks. kan CaMV35S-promoteren liggeres til den delen av ssRUBISCO-genet som undertrykker ekspresjonen av ssRUBISCO i fravær av lys, for å danne en promoter som er aktiv i blader, men ikke i røtter. Den resulterende kimærpromoter kan anvendes som beskrevet her. For formål ifølge denne beskrivelse omfatter således uttrykket "CaMV35S"-

- promoter variasjoner av CaMV35S-promoter, f.eks. promoterer avledet ved hjelp av ligering med operatorområder, tilfeldig eller kontrollert mutagenese, etc. Dessuten kan promoterer endres slik at de inneholder flere "enhancer-sekvenser" for å hjelpe til å forhøye genekspresjon.

Den RNA som fremstilles ved hjelp av en DNA-konstruksjon ifølge foreliggende oppfinnelse, inneholder også en 5'-ikke-translatert ledersekvens. Denne sekvensen kan avledes fra promoteren som er valgt til å uttrykke genet, og kan modifiseres spesifikt for å øke translasjon av mRNA-en. De 5'-ikke-translaterte områder kan også fås fra virus-RNA-er, fra egnede eukaryote gener eller fra en syntetisk gensekvens. Foreliggende oppfinnelse er ikke begrenset til konstruksjoner som er angitt i de følgende eksempler. Derimot kan den ikke-translaterte ledersekvens være en del av 5'-enden til det ikke-translaterte område i den kodende sekvens for viruskappeproteinet, eller del av promotersekvensen, eller kan avledes fra en ubeslektet promoter eller kodende sekvens. I alle tilfeller er det foretrukket at sekvensen som flankerer oppstartingsstedet, er i overensstemmelse med translasjonskonsensussekvensreglene for forøkt translasjonsinitiering rapportert av Kozak (1984).

DNA-konstruksjonen ifølge foreliggende oppfinnelse inneholder også en modifisert eller fullstendig syntetisk strukturkodende sekvens som er blitt endret for å øke virkningen av genet i planter. Ved en bestemt utførelsesform av foreliggende oppfinnelse er forøkningsmetoden blitt anvendt til å utforme modifiserte og fullstendig syntetiske gener som koder for krystalltoksinproteinet til *Bacillus thuringiensis*. Strukturgenene ifølge foreliggende oppfinnelse kan eventuelt kode for et fusjonsprotein som omfatter et aminoende-kloroplast-transittpeptid eller en sekretorisk signalsekvens (se f.eks. eksemplene 10 og 11).

DNA-konstruksjonen inneholder også et 3'-ikke-translatert område. Det 3'-ikke-translaterte område inneholder et polyadenyleringssignal som virker i planter til å forårsake addisjonen av polyadenylatnukleotider til 3'-

- enden i virus-RNA-en. Eksempler på egnede 3'-områder er (1) de 3'-transkriberte, ikke-translaterte områder som inneholder polyadenyleringssignalet til *Agrobacterium* tumor-induserende (Ti) plasmidgener, slik som nopalin-syntasegen (NOS), og (2) slike plantegener som genene for soyabønnelagringsprotein (7S) og små underenheten av genet for RuBP-carboxylase (E9). Et eksempel på et foretrukket 3'-område er det fra 7S-genet, beskrevet nærmere i eksemplene nedenunder.

10

Plantetransformasjon

Et kimærplantegen som inneholder en strukturkodende sekvens ifølge foreliggende oppfinnelse, kan innføres i genomet til en plante ved hjelp av hvilken som helst egnet metode. Egnede planter for anvendelse ved utøvelsen av foreliggende oppfinnelse, omfatter, men er ikke begrenset til, soyabønne, bomull, alfalfa, oljefrørap, lin, tomat, sukkerroe, solsikke, potet, tobakk, mais, ris og hvete. Egnede plantetransformasjonsvektorer omfatter de som er avledet fra et Ti-plasmid av *Agrobacterium tumefaciens*, samt de som er beskrevet f.eks. av Herrera-Estrella (1983), Bevan (1983), Klee (1985) og i EPO-publikasjon 120.516 (Schilperoort et al.). I tillegg til plantetransformasjonsvektorer avledet fra Ti- eller de rotinduserende (Ri) plasmider fra *Agrobacterium*, kan det anvendes alternative fremgangsmåter for å innføre DNA-konstruksjonene ifølge oppfinnelsen i planteceller. Slike fremgangsmåter kan f.eks. omfatte anvendelsen av liposomer, elektroporasjon, kjemikalier som øker opptak av fritt DNA, avlevering av fritt DNA via mikroprosjektilbombardement, og transformasjon under anvendelse av virus eller pollen.

En særlig anvendbar Ti-plasmidkassetvektor for transformasjon av tofrøbladede planter er vist i figur 5. Under henvisning til figur 5 består ekspresjonskassetten pMON893 av den forsterkede CaMV35S-promoter (EN 35S) og 3'-enden inkludert polyadenyleringssignaler fra et soyabønnegen som koder for alfa-prime-underenheten til beta-conglycinin.

35

- Mellom disse to elementene er det en multilinker som inneholder flere restriksjons seter for innføyelse av gener.

Den forsterkede CaMV35S-promoter ble konstruert på følgende måte. Et fragment av CaMV35S-promoterens som strakk seg mellom posisjon -343 og +9, ble tidligere konstruert i pUC13 av Odell et al. (1985). Dette segmentet inneholder et område identifisert av Odell et al. (1985) og som er nødvendig for maksimal ekspresjon av CaMV35S-promoterens. Det ble tatt ut som et ClaI-HindIII-fragment, gjort buttendet med DNA-polymerase I (Klenow-fragment) og innføyet i HincII-setet til pUC18. Dette oppstrøms område for 35S-promoterens ble tatt ut fra dette plasmidet som et HindIII-EcoRV-fragment (som strakk seg fra -343 til -90) og innføyet i det samme plasmid mellom HindIII- og PstI-setene. Den forsterkede CaMV35S-promoter inneholder således en fordobling av sekvenser mellom -343 og -90 (Kay et al., 1987).

3'-enden til 7S-genet er avledet fra 7S-genet som finnes på klonen betegnet 17.1 (Schuler et al., 1982). Dette 3'-endefragment som omfatter polyadenyleringssignalene, strekker seg fra AvaII-setet lokalisert ca. 30 bp oppstrøms for termineringskodonet for beta-conglyciningenet i klon 17.1, til et EcoRI-sete lokalisert ca. 450 bp nedstrøms for dette termineringskodonet.

Det gjenværende av pMON893 inneholder et segment av pBR322 som gir et opprinnelsessted for replikasjon i *E. coli*, og et område for homolog rekombinering med den uskadeliggjorte T-DNA i *Agrobacterium* stamme ACO (beskrevet nedenunder), oriV-området fra plasmidet RK1 med bredt vertsområde, streptomycin/spectinomycin-resistensgenet fra Tn7, og et kimært NPTII-gen som inneholder CaMV35S-promoterens og nopalinsyntase-3'-enden (NOS-3'-enden), som gir kanamycin-resistens i transformerte planteceller.

I figur 6 er transformasjonsvektorplasmid pMON900 et derivat av pMON893. Den forsterkede CaMV35S-promoter fra pMON893 er blitt erstattet med 1,5 kb mannopinsyntase-promoter (MAS-promoter) (Velten et al., 1984). De øvrige segmentene er de samme som i plasmid pMON893. Etter inkor-

porering av en DNA-konstruksjon i plasmidvektor pMON893 eller pMON900 innføres den intermediære vektor i *A. tumefaciens* stamme ACO som inneholder et uskadeliggjort Ti-plasmid. Samintegreerte Ti-plasmidvektorer velges ut og anvendes til å transformere tofrøbladede planter.

I figur 7 er *A. tumefaciens* ACO en lignende, ufarliggjort stamme som pTiB6SE beskrevet av Fraley et al. (1985). For konstruksjon av ACO var utgangsstammen av *Agrobacterium* stammen A208 som inneholder et Ti-plasmid av nopalin-type. Ti-plasmidet ble ufarliggjort på en lignende måte som beskrevet av Fraley et al. (1985), slik at hovedsakelig alt av den naturlig forekommende T-DNA ble fjernet, bortsett fra den venstre grense og noen få hundre basepar av T-DNA innenfor den venstre grense. Det gjenværende av T-DNA-en som strakk seg til et punkt like utenfor den høyre grense, ble erstattet med et nytt stykke DNA som omfatter (fra venstre mot høyre) et segment av pBR322, oriV-området fra plasmid RK2 og kanamycinresistensgenet fra Tn601. pBR322- og oriV-segmentene er like med segmentene i pMON893 og tilveiebringer et homologiområde for samintegreringsdannelse.

De følgende eksempler er gitt for bedre å belyse utøvelsen av foreliggende oppfinnelse.

Eksempel 1 -- Modifisert B.t.k. HD-1-gen

Under henvisning til figur 2, så er B.t.k. HD-1-villtypegenet kjent for å bli dårlig uttrykt i planter som et fullengdegen eller som et avkortet gen. G+C-innholdet i B.t.k.-genet er lavt (37%) og inneholder mange A+T-rike områder, potensielle polyadenyleringsseter (18 seter, se tabell II for listen over sekvenser) og et stort antall ATTTA-sekvenser.

Tabell IIListe over sekvenser for de potensielle
polyadenyleringssignaler

5

AATAAA*	AAGCAT
AATAAT*	ATTAAT
AACCAA	ATACAT
ATATAA	AAAATA
AATCAA	ATTAAA**
ATACTA	AATTAA**
ATAAAA	AATACA**
ATGAAA	CATAAA**

10

15

* angir et potensielt viktig plantepolyadenyleringssete.

** angir et potensielt, mindre viktig dyrepolyadenyleringssete.

20

Alle andre er potensielle, mindre viktige plantepolyadenyleringsseter.

25

Tabell III angir de syntetiske oligonukleotidene utformet og syntetisert for den stedrettede mutagenese av B.t.k. HD-1-genet.

30

35

Tabell IIIMutageneseprimere for B.t.k. HD-1-gen

	<u>Primer</u>	<u>Lengde (bp)</u>	<u>Sekvens</u>
5	BTK185	18	TCCCCAGATA ATATCAAC
	BTK240	48	GGCTTGATTCT CTAGCGAACT CTTCGATTCT CTGGTTGATG AGCTGTTC
10	BTK462	54	CAAACTGAG AGGTGGAGGT TGGCAGCTTG AACGTACACG GAGAGGAGAGGAAC
15	BTK669	48	AGTTAGTGTA AGCTCTCTTC TGAAGTGGTT GTACCTGATC CAATCTCT
20	BTK930	39	AGCCATGATC TGGTGACCGG ACCAGTAGTA TTCTCCTCT
	BTK1110	32	AGTTGTTGGT TGTTGATCCC GATGTTAAAA GG
25	BTK1380A	37	GTGATGAAGG GATGATGTTG TTGAACTCAG CACTACG
30	BTK1380T	100	CAGAAGTTCC AGAGCCAAGA TTAGTAGACT TGGTGAGTGG GATTTGGGTG ATTTGTGATG AAGGGATGAT GTTGTTGAAC TCAGCACTAC GATGTATCCA
35	BTK1600	27	TGATGTGTGG AACTGAAGGT TTGTGGT

• B.t.k. HD-1-genet (BglIII-fragment fra pMON9921 som koder for aminosyrene 29-607 med et Met-Ala i N-enden) ble klonet inn i pMON7258 (pUC118-derivat som inneholder et BglIII-sete i multilinker-kloningsområdet) i BglIII-setet, 5 hvorved man fikk pMON5342. Orienteringen til B.t.k.-genet ble valgt slik at den motsatte tråd (negative tråd) ble syntetisert i filament-fagpartikler for mutagenesen. Fremgangsmåten til Kunkle (1985) ble brukt for mutagenesen under anvendelse av plasmid pMON5342 som utgangsmateriale.

10 Områdene for mutagenese ble valgt på følgende måte. Alle områder i DNA-sekvensen til B.t.k.-genet som inneholdt fem eller flere på hverandre følgende basepar som var A eller T, ble identifisert. Disse ble rangert etter lengde og høyest prosentandel med A+T i den omgivende sekvens over 15 et område på 20-30 basepar. DNA-en ble så analysert med hensyn på områder som kunne inneholde polyadenylerings-seter (se tabell II ovenfor) eller ATTTA-sekvenser. Oligonukleotider som maksimaliserte elimineringen av A+T på hverandre følgende områder som inneholdt ett eller flere polyadenyleringsseter eller ATTTA-sekvenser, ble utformet. 20 To potensielle plante-polyadenyleringsseter ble vurdert som mer kritiske (se tabell II) basert på publiserte rapporter. Det ble valgt ut kodoner som økte G+C-innhold, ikke frembrakte restriksjonsseter for enzymer som er anvendbare for 25 kloning og sammensetning av det modifiserte gen (BamHI, BglIII, SacI, NcoI, EcoRV), og ikke inneholdt dublettene TA eller GC som er blitt rapportert å finnes lite hyppig i kodoner i planter. Oligonukleotidene var minst 18 bp lange, varierende opptil 100 basepar, og inneholdt minst 5-8 basepar med direkte homologi til naturlig forekommende sekvenser 30 i endene til fragmentene for effektiv hybridisering og priming i stedrettet mutagenesereaksjoner. Figur 2 sammenligner villtype B.t.k. HD-1-gensekvens med sekvensen som man fikk fra modifikasjonene ved hjelp av stedrettet mutagenese. 35

Sluttresultatet av disse endringene var å øke G+C-innholdet i B.t.k.-gen fra 37% til 41%, mens de potensielle

- plante-polyadenyleringssetene også ble redusert fra 18 til 7 og ATTTA-områdene ble redusert fra 13 til 7. Nærmere bestemt er mutageneseendringene fra amino- (5') ende til carboxy- (3') ende som følger:

5 BTK185 er en 18-mer anvendt til å fjerne et plante-polyadenyleringssete i midten av et A+T-område med ni basepar.

 BTK240 er en 48-mer. Syv basepar ble endret ved hjelp av dette oligonukleotid for å fjerne tre potensielle polyadenyleringsseter (2 AACCAA, 1 AATTAA). Et annet område nært opptil området endret ved hjelp av BTK240, som startet i bp 312, hadde et høyt A+T-innhold (13 av 15 basepar) og et ATTTA-område. Det inneholdt imidlertid ikke et potensielt polyadenyleringssete, og dets lengste streng av uavbrutt A+T var syv basepar.

 BTK462 er en 54-mer som innfører 13 baseparendringer. De første seks endringene var for å redusere A+T-rikheten i genet ved å erstatte villtypekodoner med kodoner som inneholdt G og C, samtidig som CG-dubletten ble unngått. De neste syv endringene gjort ved hjelp av BTK462, ble brukt til å fjerne et A+T-rikt område (13 av 14 basepar var A eller T) inneholdende to ATTTA-områder.

 BTK669 er en 48-mer som gjør ni individuelle baseparendringer som fjerner tre mulige polyadenyleringsseter (ATATAA, AATCAA og AATTAA) og et enkelt ATTTA-sete.

 BTK930 er en 39-mer utformet for å øke G+C-innholdet og fjerne et potensielt polyadenyleringssete (AATAAT - et viktig sete). Dette området inneholdt et område med ni basepar i en sekvens av på hverandre følgende A+T. En av disse baseparendringene var en G til A fordi en G i denne posisjon ville ha dannet et G+C-rikt område (CCGG(G)C). Ettersom sekvenseringsreaksjoner indikerer at det kan være vanskeligheter med å frembringe sekvens gjennom G+C på hverandre følgende baser, var det antatt å være viktig å unngå å frembringe potensielt problematiske områder selv om de var problematiske bare in vitro.

 BTK1110 er en 32-mer utformet for å innføre fem endringer i villtypegenet. Ett potensielt sete (AATAAT -

et viktig sete) ble fjernet i midten av et A+T-rikt område (19 av 22 basepar).

BTK1380A og BTK1380T er ansvarlige for 14 individuelle baseparendringer. Det første området (1380A) har 17 på hverandre følgende A+T-basepar. I dette området er et ATTTA og et potensielt polyadenyleringssete (AATAAT). 100-meren (1380T) inneholder alle endringene diktert av 1380A. Den store størrelsen til denne primeren var delvis et forsøk på å bestemme om det var mulig å utnytte store oligonukleotider for mutagenese (over 60 baser i lengde). En annen vurdering var at 100-meren ble brukt for å mutagenisere et templat som tidligere var blitt mutagenisert av 1380A. Den opprinnelige primer beregnet på å mutagenisere området nedstrøms for og grensende opp til 1380A, festet seg ikke effektivt til det ønskede sted, som indikert av en manglende evne til å oppnå ren sekvens som utnytter primeren. Det store homologiområde til 1380T sikret korrekt hybridisering. Den utvidede størrelse på 1380T var mer forårsaket av bekvemmelighetshensyn enn nødvendighet. Det andre området som grenser opp til 1380A og dekket av 1380T, har et høyt A+T-innhold (22 av 29 baser er A eller T).

BTK1600 er en 27-mer som er ansvarlig for fem individuelle baseparendringer. Et ATTTA-område og et plante-polyadenyleringssete ble identifisert og de passende endringer utført.

I alt 62 baser ble endret ved stedrettet mutagenese. G+C-innholdet økte med 55 basepar, de potensielle polyadenyleringsseter ble redusert fra 18 til 7, og ATTTA-sekvensene ble redusert fra 13 til 7. Endringene i DNA-sekvensen resulterte i endringer i 55 av de 579 kodonene i det avkortede B.t.k.-gen i pMON5342 (omtrent 9,5%).

Under henvisning til tabell IV ble modifiserte B.t.k. HD-1-gener konstruert som inneholdt alle de ovenfor nevnte modifikasjoner (pMON5370) eller forskjellige undersett av individuelle modifikasjoner. Disse genene ble innføyet i pMON893 for plantetransformasjon, og tobakkplanter som inneholdt disse genene, ble analysert. Analysen av

- tobakkplanter med de enkelte modifikasjonene ble gjennomført av flere grunner. Ekspresjon av det avkortede villtypegen i tobakk er svært dårlig, noe som resulterer i lite hyppig identifikasjon av planter som er toksiske for THW. Toksisitet defineres ved hjelp av analyser av nærings-tilførsel til blader med minst 60% dødelighet for nyfødte larver av tobakkhornorm med en skade som ble vurdert til 1 eller mindre (skala er 0-4, 0 er lik med total beskyttelse, 4 total ødeleggelse). Det modifiserte HD-1-gen (pMON5370) utviser en stor økning i ekspresjon (beregnet til å være omtrent 100 ganger, se tabell VIII) i tobakk. Økninger i ekspresjon av villtypegenet på grunn av individuelle modifikasjoner ville derfor være en tilsynelatende stor økning i hyppigheten av toksiske tobakksplanter og tilstedeværelsen av påvisbart B.t.k.-protein. Resultater er vist i den følgende tabell:

Tabell IV
Relative virkninger av regionale
modifikasjoner i B.t.k.-genet

	<u>Konstruksjon</u>	<u>Modifisert posisjon</u>	<u>Antall planter</u>	<u>Antall toksiske planter</u>
20	pMON5370	185,240,669,930, 1110,1380a+b,1600	38	22
25	pMON10707	185,240,462,669	48	19
	pMON10706	930,1110,1380a+b,1600	43	1
30	pMON10539	185	55	2
	pMON10537	240	57	17
	pMON10540	185,240	88	23
35	pMON10705	462	47	1

Virkningene av endringene av hver enkelt oligonukleotid på ekspresjon avslørte ikke noen almenngyldige tilbøyeligheter. 6 forskjellige konstruksjoner ble frembrakt som var utformet for å identifisere nøkkelområdene. De 9 forskjellige oligonukleotidene ble delt i to etter deres posisjon i genet. Endringer i den N-terminale halvdel ble inkorporert i pMON10707 (185, 240, 462, 669). Endringer i C-ende halvdel ble inkorporert i pMON10706 (930, 1110, 1380a+b, 1600). Resultatene av analyser av planter med disse to konstruksjonene indikerer at pMON10707 gir et betydelig antall toksiske planter (19 av 48). Protein fra disse plantene er påvisbart ved hjelp av ELISA-analyse. pMON10706-planter ble sjeldent identifisert som insekticide (1 av 43), og nivåene av B.t.k. var bare påvisbare ved hjelp av immunologisk analyse. Nærmere undersøkelse av N-endeendringene ble gjort med 4 pMON-konstruksjoner; 10539 (185 alene), 10537 (240 alene), 10540 (185 og 240) og 10705 (462 alene). Resultatene indikerer at tilstedeværelsen av endringene i 240 var påkrevet for å frembringe et vesentlig antall av toksiske planter (pMON10540; 23 av 88, pMON10537; 17 av 57). Fraværet av de 240 endringene resulterte i en lav hyppighet av toksiske planter med lave B.t.k.-proteinnivåer, identisk med resultatene med villtypegenet. Disse resultatene indikerer at endringene i 240 er ansvarlige for en vesentlig økning i B.t.k.-ekspresjonsnivåer i forhold til en analog villtypekonstruksjon i tobakk. Endringer i ytterligere områder (185, 462, 669) sammen med 240 kan resultere i økninger i B.t.k.-ekspresjon (>2 ganger). Endringer i 240-området til N-endedelen av genet resulterer imidlertid ikke i dramatiske økninger i ekspresjon.

Til tross for viktigheten av endringen i 240-området når det gjelder ekspresjon av modifiserte gener, kan forøkt ekspresjon oppnås ved endring av andre områder. Hybridgener, delvis av villtype, delvis syntetiske, ble frembrakt for å bestemme effektene av syntetiske gensegmenter på nivåene av B.t.k.-ekspresjon. Et hybridgen ble frembrakt med en syntetisk N-endetredjedel (basepar 1-590 i figur 2:

til XbaI-setet) med C-ende-villtype B.t.k. HD-1 (pMON5378). Planter transformert med denne vektoren, var like toksiske som planter transformert med det modifiserte HD-1-gen (pMON5370). Dette er i overensstemmelse med endringen i 240-området. pMON10538, et hybrid med en villtype-N-ende-tredjedel (villtypegenen for de første 600 basepar, til det andre XbaI-sete) og en syntetisk C-ende siste to tredjedeler (basepar 590-1845 i figur 3) ble brukt til å transformere tobakk og resulterte i en dramatisk økning i ekspresjon. Ekspresjonsnivåene synes ikke å være like høye som de som ses med det syntetiske gen, men er sammenlignbare med nivåene til det modifiserte gen. Disse resultatene indikerer at modifikasjon av 240-segmentet ikke er av avgjørende betydning for økt ekspresjon ettersom pMON10538 har et intakt 240-område. Et fullstendig syntetisk gen er i de fleste tilfeller bedre for ekspresjonsnivåer for B.t.k. (se eksempel 2).

Eksempel 2 -- Fullstendig syntetisk B.t.k. HD-1-gen

Et syntetisk B.t.k. HD-1-gen ble utformet under anvendelse av de foretrukne plantekodoner angitt i tabell V nedenunder. Tabell V angir kodonene og brukshyppigheten i plantegenet i tofrøbladede planter sammenlignet med hyppigheten av deres bruk i B.t.k. HD-1-genet til villtypen (aminosyrene 1-615) og det syntetiske gen ifølge dette eksempel. Det totale antall av hver aminosyre i dette segmentet av genet er angitt i parentes under aminosyreangivelsen.

Tabell VKodon i bruk i syntetisk B.t.k. HD-1-gen

5	Amino- <u>syre</u>	<u>Kodon</u>	Prosentvis anvendelse i planter/villtype- <u>B.t.k./syntetisk</u>		
10	ARG (43)	CGA	7	11	2
		CGC	11	5	5
		CGG	5	2	0
		CGU	25	14	27
		AGA	29	55	41
		AGG	23	14	25
15	LEU (49)	CUA	8	16	4
		CUC	20	0	20
		CUG	10	2	6
		CUU	28	22	24
		UUA	5	50	0
		UUG	30	10	45
20	SER (64)	UCA	14	27	5
		UCC	26	9	28
		UCG	3	8	0
		UCU	21	19	31
		AGC	21	6	32
		AGU	15	31	5
25	THR (42)	ACA	21	31	14
		ACC	41	19	53
		ACG	7	14	0
		ACU	31	36	33
30	PRO (34)	CCA	45	35	53
		CCC	19	6	12
		CCG	9	21	3
		CCU	26	38	32

Tabell V - forts.Kodon i bruk i syntetisk B.t.k. HD-1-gen

5	Amino- syre	Kodon	Prosentvis anvendelse i planter/villtype- B.t.k./syntetisk		
10	ALA (31)	GCA	23	38	26
		GCC	32	9	29
		GCG	3	3	0
		GCU	41	50	45
15	GLY (46)	GGA	32	52	45
		GGC	20	17	15
		GGG	11	15	6
		GGU	37	15	34
20	ILE (46)	AUA	12	39	2
		AUC	45	11	67
		AUU	43	50	30
25	VAL (38)	GUA	9	45	3
		GUC	20	5	16
		GUG	28	11	37
		GUU	43	39	45
	LYS (3)	AAA	36	100	33
		AAG	64	0	67
30	ASN (44)	AAC	72	27	80
		AAU	28	73	20
	GLN (31)	CAA	64	77	61
		CAG	36	23	39
35	HIS (10)	CAC	65	0	80
		CAU	35	100	20

Tabell V - forts.Kodon i bruk i syntetisk B.t.k. HD-1-gen

5	Amino- syre	Kodon	Prosentvis anvendelse i planter/villtype- B.t.k./syntetisk		
10	GLU	GAA	48	87	50
	(30)	GAG	52	13	50
15	ASP	GAC	48	17	65
	(23)	GAU	52	83	35
20	TYR	UAC	68	20	72
	(25)	UAU	32	80	28
25	CYS	UGC	78	50	100
	(2)	UGU	22	50	0
30	PHE	UUC	56	17	83
	(36)	UUU	44	83	17
35	MET	AUG	100	100	100
	(9)				
40	TRP	UGG	100	100	100
	(9)				

Det resulterende, syntetiske gen mangler ATTTA-sekvenser, inneholder bare ett potensielt polyadenyleringssete og har et G+C-innhold på 48,5%. Figur 3 er en sammenligning av villtype-HD-1-sekvensen med den syntetiske gen-sekvens for aminosyrene 1-615. Det er omtrent 77% DNA-homologi mellom det syntetiske gen og villtypegenet, og 356 av de 615 kodonene er blitt endret (omtrent 60%).

• Eksempel 3 -- Syntetisk B.t.k. HD-73-gen

Krystallproteintoksinet fra B.t.k. HD-73 utviser en høyere enhetsaktivitet mot noen viktige skadeinsekter i landbruket. Toksinproteinet til HD-1 og HD-73 har betydelig
5 homologi (ca. 90%) i de 450 aminosyrene i N-enden, men at-
skiller seg betydelig i aminoområdet 451-615. Fusjons-
proteiner som omfatter aminosyrene 1-450 i HD-1 og 451-615
i HD-73, har de insekticide egenskapene til villtype HD-73.
Den strategi som ble anvendt, var å bruke 5'-to tredje-
10 delene av det syntetiske HD-1-gen (de første 1350 baser,
opptil SacI-setet) og modifisere de siste 590 basene dramat-
isk (forbi aminosyre 645) i HD-73 på en måte som er i over-
ensstemmelse med den algoritme som er brukt for å betegne
det syntetiske HD-1-gen. Tabell VI nedenunder angir oligo-
15 nukleotidene som ble brukt til å modifisere HD-73-genet i
den anvendte rekkefølge i genet fra 5'- til 3'-ende. Det
ble brukt 9 oligonukleotid i et 590 basepar stort område,
idet hvert nukleotid varierte i størrelse fra 33 til 60 baser.
De eneste områder som forble uendret, var områder hvor det
20 ikke var noen lange, fortløpende strenger av A- eller T-
baser (lengre enn seks). Alle polyadenylerings seter og
ATTTA-seter ble fjernet.

25

30

35

Tabell VIMutageneseprimere for B.t.k. HD-73

5			
	<u>Primer</u>	<u>Lengde (bp)</u>	<u>Sekvens</u>
10	73K1363	51	AATACTATCG GATGCGATGA TGTTGTTGAA CTCAGCACTA CGGTGTATCC A
	73K1437	33	TCCTGAAATG ACAGAACCGT TGAAGAGAAA GTT
15	73K1471	48	ATTTCCTACTG CTGTTGAGTC TAACGAGGTC TCCACCACTG AATCCTGG
20	73K1561	60	GTGAATAGGG GTCACAGAAG CATACCTCAC ACGAACTCTA TATCTGGTAG ATGTTGGATGG
25	73K1642	33	TGTAGCTGGA ACTGTATTGG AGAAGATGGA TGA
	73K1675	48	TTCAAAGTAA CCGAAATCGC TGGATTGGAG ATTATCCAAG GAGGTAGC
30	73K1741	39	ACTAAAGTTT CTAACACCCA CGATGTTACC GAGTGAAGA
	73K1797	36	AACTGGAATG AACTCGAATC TGTCGATAAT CACTCC
35	73KTERM	54	GGACACTAGA TCTTAGTGAT AATCGGTCAC ATTTGTCTTG AGTCCAAGCT GGTT

• Det resulterende gen har to potensielle polyadenyleringssetter (sammenlignet med 18 i villtypen) og ingen ATTTA-sekvens (12 i villtypen). G+C-innholdet har økt fra 37 til 48%. I alt 59 individuelle baseparendringer ble gjort ved
5 å bruke primerne i tabell VI. Totalt er det 90% DNA-homologi mellom området i HD-73-genet modifisert ved hjelp av stedrettet mutagenese og villtype-sekvensen i det analoge område i HD-73. Den syntetiske HD-73 er en hybrid av de første 1360 basene fra den syntetiske HD-1 og de neste 590
10 basene av slik modifisert HD-73-sekvens. Figur 4 er en sammenligning mellom den ovenfor beskrevne, syntetiske B.t.k. HD-73 og villtypen av B.t.k. HD-73 som koder for aminosyrene 1-645. I det modifiserte område av HD-73-genet ble 44 av de 170 kodonene (25%) endret som et resultat av
15 endringene av stedrettet mutagenese som skrev seg fra oligonukleotidene som finnes i tabell VI. Totalt adskiller omtrent 50% av kodonene i den syntetiske B.t.k. HD-73 seg fra det analoge segment av villtype- og HD-73-genet.

En én-basepar-delesjon i det syntetiske HD-73-gen
20 ble påvist under sekvenseringen av 3'-enden i basepar 1890. Dette resulterer i en rammeskiftmutasjon i aminosyre 625 med et for tidlig stoppkodon ved aminosyre 640 (pMON5379). Tabell VII nedenunder sammenligner kodonbruken av villtype-genet til B.t.k. HD-73 versus det syntetiske gen ifølge
25 dette eksemplet for aminosyrene 451-645, og kodonbruk hos naturlig forekommende gener fra tofrøbladede planter. Det totale antall av hver aminosyre kodet for i dette segmentet av genet, finnes i parentesene under aminosyrebetegnelsen.

30

35

Tabell VIIKodonbruk i syntetisk B.t.k. HD-73-gen

5	Amino- syre	Kodon	Prosentvis anvendelse i planter/villtype- HD-73/syntetisk		
10	ARG (10)	CGA	7	10	0
		CGC	11	0	8
		CGG	5	10	0
		CGU	25	20	23
		AGA	29	60	62
		AGG	23	0	8
15	LEU (12)	CUA	8	25	8
		CUC	20	17	58
		CUG	10	17	8
		CUU	28	8	0
		UUA	5	33	8
		UUG	30	0	17
20	SER (21)	UCA	14	24	18
		UCC	26	10	27
		UCG	3	10	0
		UCU	21	24	18
		AGC	21	0	14
		AGU	15	33	23
30	THR (15)	ACA	21	47	38
		ACC	41	13	31
		ACG	7	13	0
		ACU	31	27	31
35	PRO (7)	CCA	45	71	71
		CCC	19	0	0
		CCG	9	14	0
		CCU	26	14	29

Tabell VII - forts.Kodonbruk i syntetisk B.t.k. HD-73-gen

5	<u>Amino-</u> <u>syre</u>	<u>Kodon</u>	Prosentvis anvendelse i planter/villtype- <u>HD-73/syntetisk</u>		
10	ALA (14)	GCA	23	29	31
		GCC	32	7	8
		GCG	3	21	15
		GCU	41	43	46
15	GLY (15)	GGA	32	33	43
		GGC	20	0	0
		GGG	11	27	14
		GGU	37	40	43
20	ILE (15)	AUA	12	33	7
		AUC	45	7	40
		AUU	43	60	53
25	VAL (15)	GUA	9	40	7
		GUC	20	0	7
		GUG	28	20	36
		GUU	43	40	50
30	LYS (3)	AAA	36	67	100
		AAG	64	33	0
35	ASN (20)	AAC	72	20	53
		AAU	28	80	47

Tabell VII - forts.Kodonbruk i syntetisk B.t.k. HD-73-gen

5	<u>Amino- syre</u>	<u>Kodon</u>	<u>Prosentvis anvendelse i planter/villtype- HD-73/syntetisk</u>		
	GLN	CAA	64	60	67
	(5)	CAG	36	40	33
10	HIS	CAC	65	67	100
	(3)	CAU	35	33	0
	GLU	GAA	48	86	57
	(7)	GAG	52	14	43
15	ASP	GAC	48	40	50
	(5)	GAU	52	60	50
20	TYR	UAC	68	0	20
	(5)	UAU	32	100	80
	CYS	UGC	78	0	0
	(0)	UGU	22	0	0
25	PHE	UUC	56	8	67
	(13)	UUU	44	92	33
	MET	AUG	100	100	100
	(2)				
30	TRP	UGG	100	100	100
	(2)				

Et annet avkortet, syntetisk HD-73-gen ble konstruert. Sekvensen til dette syntetiske HD-73-gen er identisk med sekvensen til det ovenfor nevnte, syntetiske HD-73-gen i området hvor de overlapper (aminosyrene 29-615), og det koder også for Met-Ala i N-enden. Figur 8 viser en sammenligning med dette avkortede, syntetiske HD-73-gen og N-ende-Met-Ala versus villtype-HD-73-genet.

Mens de tidligere eksempler har vært rettet mot fremstillingen av syntetiske og modifiserte gener som koder for avkortede B.t.k.-proteiner, kan det også fremstilles syntetiske eller modifiserte gener som koder for toksinproteiner med full lengde.

Et fullengde B.t.k.-gen består av den syntetiske HD-73-sekvens ifølge figur 4 fra nukleotid 1-1845 pluss villtype-HD-73-sekvens som koder for aminosyrene 616 til C-enden i det naturlig forekommende protein. Figur 9 viser en sammenligning mellom dette syntetiske/villtype-fullengde-HD-73-gen versus villtype-fullengde-HD-73-genet.

Et annet fullengde B.t.k.-gen består av den syntetiske HD-73-sekvens ifølge figur 4 fra nukleotid 1-1845 pluss en modifisert HD-73-sekvens som ender i aminosyrene 616 til C-enden i det naturlig forekommende protein. C-ende-delen er blitt modifisert ved stedrettet mutagenese for å fjerne antatte polyadenyleringssignaler og ATTTA-sekvenser ifølge algoritmen i figur 1. Figur 10 viser en sammenligning av dette syntetiske/modifiserte fullengde-HD-73-gen versus villtype-fullengde-HD-73-genet.

Et annet fullengde B.t.k.-gen består av en fullstendig syntetisk HD-73-sekvens som omfatter den syntetiske HD-73-sekvens ifølge figur 4 fra nukleotid 1-1845 pluss en syntetisk sekvens som koder for aminosyrene 616 til C-enden i det naturlig forekommende protein. Den C-terminale, syntetiske del er blitt utformet for å fjerne antatte polyadenyleringssignaler og ATTTA-sekvenser, og for å omfatte planteforetrukne kodoner. Figur 11 viser en sammenligning mellom dette fullstendig syntetiske fullengde-HD-73-gen versus villtype-fullengde-HD-73-genet.

Alternativt består et annet fullengde-B.t.k.-gen av en fullstendig syntetisk sekvens som omfatter baseparene 1-1830 i B.t.k. HD-1 (figur 3) og baseparene 1834-3534 i B.t.k. HD-73 (figur 11).

Eksempel 4 -- Ekspresjon av modifisert og syntetisk B.t.k. HD-1 og syntetisk HD-73

En del plantetransformasjonsvektorer for ekspresjonen av B.t.k.-gener ble konstruert ved å inkorporere de strukturkodende sekvenser av de tidligere beskrevne gener inn i plantetransformasjonskassetvektor pMON893. Den respektive intermediære transformasjonsvektor innføres i en egnet, uskadeliggjort Agrobacterium-vektor, slik som *A. tumefaciens* ACO, se ovenfor. Veveksplantater ble dyrket sammen med den uskadeliggjorte Agrobacterium-vektor og planter regenerert under utvelgelse for kanamycinresistens under anvendelse av kjente fremgangsmåter: tobakk (Horsch et al., 1985), tomat (McCormick et al., 1986) og bomull (Trolinder et al., 1987).

a) Tobakk.

Nivået av B.t.k. HD-1-protein i transgene tobakksplanter inneholdende pMON9921 (villtype, avkortet), pMON5370 (modifisert HD-1, eksempel 1, figur 2) og pMON5377 (syntetisk HD-1, eksempel 2, figur 3), ble analysert ved hjelp av Western-analyse. Bladvev ble nedfrysset i flytende nitrogen, oppmalt til et fint pulver og så oppmalt i en 1:2 (vekt:volum) SDS-PAGE-prøvebuffer. Prøver ble nedfrysset på tørris, så inkubert i 10 minutter i et kokende vannbad og mikrosentrifugert i 10 minutter. Proteinkonsentrasjonen i supernatanten ble bestemt ved hjelp av metoden til Bradford (Anal. Biochem., 72, s. 248-254). 50 µg protein ble anvendt pr. felt på 9% SDS-PAGE-geler, proteinet ble overført til nitrocellulose og B.t.k. HD-1-proteinet visualisert under anvendelse av antistoffer fremstilt mot B.t.k. HD-1-protein som det primære antistoff og alkalisk fosfatase-konjugerte, andre antistoff som beskrevet av produsenten. Renset HD-1

• tryptisk fragment ble brukt som kontroll. Mens B.t.k.-
 proteinet fra tobakksplanter inneholdende pMON9921 var
 under påvisningsnivået, ble B.t.k.-proteinet fra planter
 som inneholdt det modifiserte (pMON5370) og syntetiske
 5 (pMON5377) genet, lett påvist. B.t.k.-proteinet fra planter
 inneholdende pMON9921, forble upåvisbare, selv med 10 ganger
 lengre inkubasjonstider. De relative nivåene av B.t.k.
 HD-1-protein i disse plantene er beregnet i tabell VIII.
 Ettersom proteinet fra planter inneholdende pMON9921 ikke
 10 ble observert, ble proteinnivået i disse plantene beregnet
 ut fra de relative mRNA-nivåer (se nedenunder). Planter
 inneholdende det modifiserte gen (pMON5370), uttrykte omtrent
 100 ganger mer B.t.k.-protein enn planter som inneholdt
 villtypegenet (pMON9921). Planter inneholdende det full-
 15 stendig syntetiske B.t.k. HD-1-gen (pMON5377), uttrykte
 omtrent fem ganger mer protein enn planter som inneholdt det
 modifiserte gen. Det modifiserte gen bidrar med hoveddelen
 av økningen i B.t.k.-ekspresjon observert. Plantene som
 ble brukt til å frembringe dataene ovenfor, er de beste
 20 representanter fra hver konstruksjon basert enten på en
 biologisk analyse på tobakkshornlarve eller på data avledet
 fra tidligere Western-analyse.

Tabell VIII

Ekspresjon av B.t.k. HD-1-protein
i transgen tobakk

<u>Gen- beskrivelse</u>	<u>Vektor</u>	<u>B.t.k.-protein* konsentrasjon</u>	<u>Ganger økning i B.t.k.- ekspresjon</u>
Villtype	pMON9921	10	1
Modifisert	pMON5370	1000	100
Syntetisk	pMON5377	5000	500

* B.t.k.-proteinkonsentrasjoner er uttrykt i ng/mg av totalt
 35 oppløselig protein. B.t.k.-proteinnivået for planter
 inneholdende villtypegenet, er beregnet ut fra mRNA-
 nivåer.

• Planter som inneholdt disse genene, ble testet med hensyn på biologisk aktivitet for å bestemme hvorvidt de økte proteinmengder som ble observert ved Western-analyse, resulterte i en tilsvarende økning i biologisk aktivitet.

5 Blader fra de samme plantene som var brukt til Western-dataene i tabell 1, ble testet med hensyn på biologisk aktivitet mot to insekter. En biologisk analyse på fraskilt blad ble først gjort under anvendelse av tobakkshornlarve, et svært følsomt sommerfuglinsekt. Blader fra alle tre

10 transgene tobakksplanter ble fullstendig beskyttet, og 100% dødelighet hos tobakkshornlarve ble observert (se tabell IX nedenunder). Et mye mindre følsomt insekt, sukkerroe-hærlarve, ble så brukt i en annen biologisk analyse på adskilt blad. Sukkerroe-hærlarve er omtrent 500 ganger

15 mindre følsom overfor B.t.k. HD-1-protein enn det tobakshornlarve er. Forskjellen i følsomhet mellom disse to insektene ble bestemt ved å bruke rensset HD-1-protein i en diettinkorporeringsanalyse (se nedenunder). Planter inneholdende villtypegenet (pMON9921), utviste bare minimal

20 beskyttelse mot sukkerroe-hærlarve, mens planter som inneholdt det modifiserte gen, utviste nesten fullstendig beskyttelse, og planter som inneholdt det fullstendig syntetiske gen, var totalt beskyttet mot sukkerroe-hærlarveskade. Resultatene av disse biologiske analysene bekrefter

25 nivåene av B.t.k. HD-1-ekspressjon observert i Westernanalysen og viser at de økte nivåer av B.t.k. HD-1-protein korrelerer med økt insekticid aktivitet.

30

35

Tabell IX

Beskyttelse av tobakksplanter mot
tobakkshornlarve og sukkerroe-hærlarve

5	Gen- beskrivelse	Vektor	Tobakkshornlarve-	Sukkerroe-
			skade*	hærlarve- skade*
	Ingen	Ingen	NL	NL
	Villtype	pMON9921	0	3
10	Modifisert	pMON5370	0	1
	Syntetisk	pMON5377	0	0

* Omfang av insektskade ble gradert: 0: ingen skade;
 1: lite; 2: middels; 3: voldsomt; eller NL: ingen
 15 blad tilbake.

Den biologiske aktivitet til B.t.k. HD-1-proteinet fremstilt ved hjelp av disse transgene planter, ble ytterligere undersøkt for mer nøyaktig å kvantifisere de relative virkningene. Bladvev fra tobakksplanter inneholdende villtype-, de modifiserte og de syntetiske genene, ble oppmalt i 100 mM natriumcarbonatbuffer, pH 10, ved et forhold på 1:2 (vekt:volum). Partikkelformet materiale ble fjernet ved sentrifugering. Supernatanten ble inkorporert i en syntetisk diett som var lik med den som er beskrevet av Marrone et al. (1985). Diettmediet ble fremstilt på dagen for testen med planteekstraktopløsningene inkorporert istedenfor den 20% vannbestanddel. 1 ml av dietten ble plassert i brønnene i 96-brønners plater.

Etter at dietten hadde tørket, ble det tilsatt en nyfødt tobakksknopplarve til hver brønn. 16 insekter ble testet med hver planteprøve. Plantene ble inkubert ved 27°C. Etter 7 dager ble larvene fra hver behandling slått sammen og veid på en analytisk skålvekt. Gjennomsnittsvekten pr. insekt ble beregnet og sammenlignet med en standardkurve for B.t.k.-proteinkonsentrasjoner for å finne

gjennomsnittlig larvevekt. Insektvekt var omvendt proporsjonal (på logaritmisk måte) med den relative økning i B.t.k.-proteinkonsentrasjon. Mengden av B.t.k. HD-1-protein, basert på omfanget av larvevekstinhibering, ble bestemt for to forskjellige planter inneholdende hvert av de tre genene. Den spesifikke aktivitet (ng B.t.k. HD-1 pr. mg plante-protein) ble bestemt for hver plante. Planter inneholdende det modifiserte HD-1-gen (pMON5370), hadde et gjennomsnitt på omtrent 1400 ng (1200 og 1600 ng) av B.t.k. HD-1 pr. mg plantekstraktprotein. Denne verdien stemmer godt overens med de 1000 ng av B.t.k. HD-1-protein pr. mg planteekstraktprotein bestemt ved hjelp av Western-analyse (tabell I). B.t.k. HD-1-konsentrasjoner for plantene som inneholdt det syntetiske HD-1-gen, hadde et gjennomsnitt på omtrent 8200 ng (7200 og 9200 ng) av B.t.k. HD-1-protein pr. mg planteekstraktprotein. Dette tallet stemmer godt overens med de 5000 ng av HD-1-protein pr. mg planteekstraktprotein beregnet ved hjelp av Western-analyse. På samme måte utviste planter som inneholdt det syntetiske gen, omtrent en seks ganger høyere spesifikk aktivitet enn de tilsvarende planter som inneholdt det modifiserte gen for disse biologiske analysene. I Western-analysen var forholdet omtrent 10 ganger, igjen er begge deler i god overensstemmelse. Nivået av B.t.k.-protein i planter inneholdende villtype-HD-1-genet (pMON9921), var for lavt til å gi en signifikant reduksjon i larvevekt og var følgelig under et nivå som kunne kvantifiseres i denne analysen. Endelig stemte nivåene av B.t.k. HD-1-protein bestemt ved begge disse biologiske analysene og Western-analysen for de plantene som inneholdt det modifiserte og syntetiske gen, overens med hverandre, noe som viser at B.t.k. HD-1-proteinet fremstilt ved hjelp av disse plantene, er biologisk aktivt.

Nivåene av mRNA ble bestemt i plantene som inneholdt villtype B.t.k. HD-1-genet (pMON9921) og det modifiserte gen (pMON5370) for å fastslå hvorvidt de økte nivåer av proteinproduksjon skriver seg fra økt transkripsjon eller translasjon. mRNA fra planter inneholdende det syntetiske

• gen, kunne ikke analyseres direkte med den samme DNA-probe som ble brukt for villtype- og det modifiserte gen på grunn av det store antall endringer som var gjort i den kodende sekvens. mRNA ble isolert og hybridisert med en enkelt-
5 trådet DNA-probe som var homolog med omtrent de 5' 90 bp av de kodende sekvenser fra villtype eller modifisert gen. Hybridene ble oppløst med S1-nuklease og de beskyttede probefragmenter analysert ved gelelektroforese. Ettersom fremgangsmåten gjorde bruk av et stort overskudd probe og
10 lang hybridiseringstid, er mengden av beskyttet probe proporsjonal med mengden av B.t.k.-mRNA som er til stede i prøven. To planter som uttrykte det modifiserte gen (pMON5730), ble funnet å gi opptil 10 ganger mer RNA enn en plante som uttrykker villtypegenet (pMON9921).

15 Det økte mRNA-nivå fra det modifiserte gen er i overensstemmelse med resultatet som var forventet ut fra modifikasjonene innført i dette genet. Denne 10-ganger økning i mRNA hos det modifiserte gen sammenlignet med villtypegenet, er imidlertid i motsetning til den 100-ganger
20 økning i B.t.k.-protein fra disse genene i tobakksplanter. Dersom de to mRNA-ene ble like godt translatert, så ville en 10-ganger økning i stabil mRNA forventes å gi en 10-ganger økning i protein. Den høyere økning i protein indikerer at den modifiserte gen-mRNA translateres ved om-
25 trent en 10 ganger høyere effektivitet en villtype. Omtrent halvparten av den totale effekt for genekspressjon kan således forklares ved endringer i mRNA-nivåer, og omtrent halvparten med endringer i translasjonseffektivitet. Denne økning i translasjonseffektivitet er slående ved at bare
30 9,5% av kodonene er blitt endret i det modifiserte gen; dvs. at denne effekten klart ikke skyldes endringer i bruken av store deler av kodonene. Den økte translasjonseffektivitet kunne skyldes endringer i mRNA-sekundærstruktur som påvirker translasjon, eller fjerningen av bestemte transla-
35 sjonsblokkeringer på grunn av bestemte kodoner som ble endret.

Den økte ekspresjon som ses med det syntetiske HD-1-gen, ble også sett med et syntetisk HD-75-gen i tobakk.

- B.t.k. HD-73 var upåvist i ekstrakter av tobakksplanter som inneholdt det avkortede HD-73-gen av villtype (pMON5367), mens B.t.k. HD-73-protein lett ble påvist i ekstrakter i tobakksplanter som inneholdt det syntetiske HD-73-gen ifølge figur 4 (pMON5383). Omtrent 1000 ng B.t.k. HD-73-protein ble påvist pr. mg av alt oppløselig planteprotein.

Som beskrevet i eksempel 3 ovenfor, inneholder B.t.k. HD-73-proteinet som pMON5383 koder for, en liten C-terminal forlengelse av aminosyrer som det ikke kodes for i villtype-HD-73-proteinet. Disse ekstra aminosyrene hadde ingen effekt på insekttoksisitet eller på økt planteekspresjon. Et andre syntetisk HD-73-gen ble konstruert som beskrevet i eksempel 3 (figur 8) og brukt til å transformere tobakk (pMON5390). Analyse av planter som inneholdt pMON5390, viste at dette genet ble uttrykt på nivåer som er sammenlignbare med nivået til pMON5383, og at disse plantene hadde lignende insekticid virkningsfullhet.

I tobakksplanter ble det syntetiske HD-1-gen uttrykt ved omtrent et 5 ganger høyere nivå enn det syntetiske HD-73-gen. Dette syntetiske HD-73-gen ble imidlertid fortsatt uttrykt minst 100 ganger bedre enn villtype-HD-73-genet. HD-73-proteinet er omtrent 5 ganger mer toksisk for mange skadeinsekter enn HD-1-proteinet, så både syntetisk HD-1- og HD-73-gener gir omtrent sammenlignbar insekticid virkningsfullhet i tobakk.

Fullengde-B.t.k.-HD-73-genene beskrevet i eksempel 3, ble også inkorporert i plantetransformasjonsvektoren pMON893 slik at de ble uttrykt fra En 35S-promoter. Det syntetiske/villtype-fullengde-HD-73-gen i figur 9 ble inkorporert i pMON893 for å danne pMON10505. Det syntetiske/modifiserte fullengde-HD-73-gen i figur 10 ble inkorporert i pMON893, hvorved pMON10526 ble dannet. Det fullstendig syntetiske HD-73-gen ifølge figur 11 ble inkorporert i pMON893, hvorved pMON10518 ble dannet. Disse vektorene ble brukt til å oppnå transformerte tobakksplanter, og plantene ble analysert med hensyn på insekticid virkningsfullhet og med hensyn på B.t.k. HD-73-proteinnivåer ved

- "Western blot" eller ELISA immunologisk analyse.

Tobakksplanter inneholdende alle tre av disse full-lengde-B.t.k.-gener, produserte påvisbart B.t.k.-protein og ga 100% mortalitet på tobakkshornlarver. Dette resultatet er overraskende i lys av tidligere rapporterte forsøk på å uttrykke fullengde-B.t.k.-genene i transgene planter. Vaeck et al. (1987) rapporterte at et fullengde-B.t.k. berliner-gen som er likt med vårt HD-1-gen, ikke kunne bli påvisbart uttrykt i tobakk. Barton et al. (1987) rapporterte et lignende resultat for et annet fullengdegen fra B.t.k. HD-1 (det såkalte 4,5 kb gen), og indikerte videre at tobakk-kallus inneholdende dette genet, ble nekrotisk, noe som tyder på at fullengdegenproduktet var toksisk for planteceller. Fischhoff et al. (1987) rapporterte at fullengde-B.t.k.-HD-1-genet i tomat ble dårlig uttrykt sammenlignet med et avkortet gen, og ingen planter som var fullstendig toksiske for tobakkshornlarve, kunne utvinnes. I alle de tre ovenfor nevnte rapporter ble det angitt mye høyere ekspressionsnivåer og utvinning av toksiske planter dersom de respektive B.t.k.-gener var avkortet. Adang et al. rapporterte at fullengde-HD-73-genet ga noen få tobakksplanter med noe biologisk aktivitet (ingen var høytoksiske) mot hornlarve og knapt nok påvisbart B.t.k.-protein. Det ble også lagt merke til av dem at hoved-B.t.k.-mRNA i disse plantene var en avkortet 1,7 kb art som ikke ville kode et funksjonelt toksin. Dette indikerte ukorrekt ekspresjon av genet i tobakk. I motsetning til alle disse rapportene fører de tre fullengde-B.t.k.-HD-73-genene beskrevet ovenfor, alle til forholdsvis høye nivåer av protein og høye nivåer av insekttoksisitet.

B.t.k.-protein- og mRNA-nivåer i tobakksplanter er vist i tabell X for disse tre vektorene. Som det vil ses av tabellen, gir syntetisk/villtypegenet (pMON10506) B.t.k.-protein som omtrent 0,01% av totalt oppløselig protein; syntetisk/modifisert-genet gir B.t.k. som omtrent 0,02% av totalt oppløselig protein; og det fullstendig syntetiske gen gir B.t.k. som omtrent 0,2% av totalt oppløselig protein.

- B.t.k.-mRNA ble analysert i disse plantene ved "Northern blot"-analyse under anvendelse av den felles 5' syntetiske halvdel av genene som en probe. Som vist i tabell X, kan de forøkte proteinnivåer stort sett tilskrives økte mRNA-nivåer. Sammenlignet med de avkortede, modifiserte og syntetiske gener kunne dette indikere at de viktigste bidragsyterne til økt translasjonell virkningsfullhet er i 5'-halvdelen av genet, mens 3'-halvdelen av genet inneholder stort sett determinanter for mRNA-stabilitet. De økte proteinnivåer indikerer også at ved å øke mengden av fullengdegenet som er syntetisk eller modifisert, øker B.t.k.-proteinnivåene. Sammenlignet med de avkortede, syntetiske B.t.k. HD-73-genene (pMON5383 eller pMON5390), gir det fullstendig syntetiske gen (pMON10518) like mye eller litt mer B.t.k.-protein, noe som viser at fullengdegenene kan uttrykkes ved høye nivåer i planter. Disse tobakksplantene med høye nivåer av fullengde-HD-73-protein viser ingen tegn på unormalitet og er fullstendig fertile. B.t.k.-proteinnivåene i disse plantene gir også de forventede nivåer av insekttoksisitet basert på fôringsundersøkelser med sukkerroe-hærlarve eller diett-inkorporeringsanalyser av planteekstrakter med tobakksknopplarve. B.t.k.-proteinet påvist ved "Western blot"-analyse i disse tobakksplantene, inneholder ofte en varierende mengde protein med ca. 80 kDa som åpenbart er et proteolytisk fragment av fullengdeproteinet. C-ende-halvdelen av fullengdeproteinet er kjent for å være proteolytisk følsomt, og lignende proteolytiske fragmenter ses fra fullengdegenet i E. coli og B.t. selv. Disse fragmentene er fullstendig insekticide. Northern-analyse indikerte at praktisk talt all mRNA fra disse fullengdegenene hadde den forventede fullengdestørrelse. Det er ingen tegn på avkortede mRNA-er som kunne gi opphav til det 80 kDa proteinfragment. I tillegg er det mulig at fragmentet ikke er til stede i intakte planteceller og bare skyldes proteolyse under ekstraksjon for immunologisk analyse.

Tabell XB.t.k. HD-73-fullengdeprotein og mRNA-nivåer
i transgene tobakksplanter

5	Gen- <u>beskrivelse</u>	<u>Vektor</u>	B.t.k. protein- <u>konsentrasjon</u>	Relativt B.t.k. <u>mRNA-nivå</u>
	Syntetisk/villtype	pMON10506	>100	0,5
10	Syntetisk/modifisert	pMON10526	400	1
	Fullstendig syntetisk	pMON10518	>2000	40

Det er således ikke noe alvorlig hinder for å prod-
 15 usere høye nivåer av B.t.k. HD-73-protein i planter fra
 syntetiske gener, og dette forventes å være gyldig for
 andre fullengde, tofrøbladede, aktive gener, slik som
 B.t.k. HD-1 eller B.t. entomocidus. Det fullstendig syntet-
 iske B.t.k. HD-1-gen ifølge eksempel 3 er blitt satt sammen
 20 i plantetransformasjonsvektorer, slik som pMON893.

Det fullstendig syntetiske gen i pMON10518 ble også
 benyttet i en annen plantevektor og analysert i tobakks-
 planter. Selv om CaMV35S-promoterer generelt er en høy-
 nivå, konstitutiv promoter i de fleste plantevev, er eks-
 25 presjonsnivået for gener drevet av CaMV35S-promoterer, lavt
 i blomstervev i forhold til nivåene som ses i bladvev.
 Ettersom de økonomisk viktige mål som skades av noen insekter,
 er blomsterdelene eller avledet fra blomsterdelene (f.eks.
 bomullsterninger og -frøkapsler, tobakksknopper, tomat-
 30 knopper og frukt), kan det være fordelaktig å øke ekspresjonen
 av B.t.-protein i disse vevene i forhold til det som ble
 oppnådd med CaMV35S-promoterer.

35S-promoterer i Figwort mosaikkvirus (FMV) er ana-
 log med CaMV35S-promoterer. Denne promoterer er blitt iso-
 35 lert og bygget inn i en plantetransformasjonsvektor analog
 med pMON893. I forhold til CaMV-promoterer uttrykkes FMV
 35S-promoterer sterkt i blomstervevet, mens den fortsatt gir

• lignende høye nivåer av genekspresjon i andre vev, slik som blad. Det ble konstruert en plantetransformasjonsvektor, pMON10517, hvor det fullengde, syntetiske B.t.k. HD-73-gen ifølge figur 11 ble drevet av FMV 35S-promoter. Denne vektoren er identisk med pMON10518 i eksempel 3, bortsett fra at FMV-promoter har erstattet CaMV-promoter. Tobakksplanter transformert med pMON10517 og pMON10518, ble erholdt og sammenlignet for ekspresjon av B.t.k.-proteinet ved "Western blot" eller ELISA immunologisk analyse i blad- og blomstervev. Denne analysen viste at pMON10517 inneholdende FMV-promoter, uttrykte fullengde-HD-73-proteinet ved høyere nivåer i blomstervev enn pMON10518 inneholdende CaMV-promoter. Ekspresjon av fullengde-B.t.k.-HD-73-protein fra pMON10517 i bladvev er sammenlignbar med det som ses med de sterkest uttrykkende planter inneholdende pMON10518. Når blomstervev analyseres, hadde imidlertid tobakksplanter inneholdende pMON10518 som hadde høye nivåer av B.t.k.-protein i bladvev, ikke påvisbart B.t.k.-protein i blomstene. På den annen side hadde blomster på tobakksplanter inneholdende pMON10517, nivåer av B.t.k.-protein som var nesten like høye som nivåene i blader, ved omtrent 0,05% av totalt oppløselig protein. Denne analysen viste at FMV-promoter kunne brukes til å fremstille forholdsvis høye nivåer av B.t.k.-protein i blomstervev sammenlignet med CaMV-promoter.

b) Tomat.

Villtype, de modifiserte og syntetiske B.t.k. HD-1-gener testet i tobakk, ble innført i andre planter for å vise den brede utnyttbarhet av denne oppfinnelse. Det ble fremstilt transgene tomater som inneholder disse tre genene. Data viser at den økte ekspresjon som ble observert med det modifiserte og det syntetiske gen i tobakk, også gjelder for tomat. Mens B.t.k. HD-1-proteinet bare er såvidt påvisbart i planter som inneholder villtype-HD-1-genet (pMON9921), ble B.t.k. HD-1 lett påvist og nivåene bestemt for planter som inneholder de modifiserte (pMON5370) eller syntetiske

- (pMON5377) gener. Ekspresjonsnivåer for plantene som inneholder villtype-, de modifiserte og de syntetiske HD-1-gener, var omtrent 10, 100 og 500 ng pr. mg totalt planteekstrakt (se tabell XI nedenunder). Økningen i B.t.k. HD-1-protein for det modifiserte gen var ansvarlig for hoveddelen av den observerte økning; 10 ganger høyere enn plantene som inneholdt villtypegenet, sammenlignet med bare en ytterligere fem ganger økning for planter som inneholder det syntetiske gen. Igjen er de stedrettede endringer gjort i det modifiserte gen, de viktigste bidragsyterne til den økte ekspresjon av B.t.k. HD-1.

Tabell XI

<u>B.t.k. HD-1-ekspresjon i transgene tomatplanter</u>			
<u>Gen-</u>		<u>B.t.k. protein-*</u>	<u>Ganger økning i B.t.k.-</u>
<u>beskrivelse</u>	<u>Vektor</u>	<u>konsentrasjon</u>	<u>ekspresjon</u>
Villtype	pMON9921	10	1
Modifisert	pMON5370	100	10
Syntetisk	pMON5377	500	50

* B.t.k. HD-1-proteinkonsentrasjoner er uttrykt i ng/mg av totalt oppløselig planteprotein. Data for planter som inneholder villtypegenet, er estimater fra mRNA-nivåer og proteinnivåer bestemt ved hjelp av ELISA.

Disse forskjellene i B.t.k. HD-1-ekspresjon ble bekreftet med biologiske analyser mot tobakkhornlarve og sukkerroe-hærlarve. Blader fra tomatplanter inneholdende hvert av disse genene, kontrollerte tobakkhornlarveskade og ga 100% dødelighet. Med sukkerroe-hærlarve utviste blader fra planter som inneholdt villtype-HD-1-genet (pMON9921), betydelig skade, blader fra planter inneholdende det modifiserte gen (pMON5370), utviste mindre skade, og bladene fra

- planter inneholdende det syntetiske gen (pMON5377), var fullstendig beskyttet (se tabell XII nedenunder).

Tabell XII

Beskyttelse av tomatplanter mot
tobakkshornlarve og sukkerroe-hærlarve

10	<u>Gen- beskrivelse</u>	<u>Vektor</u>	<u>Tobakkshorn- larveskade*</u>	<u>Sukkerroe- hærlarveskade*</u>
	Ingen	Ingen	NL	NL
	Villtype	pMON9921	0	3
	Modifisert	pMON5370	0	1
15	Syntetisk	pMON5377	0	0

* Skade ble gradert som vist i tabell IX.

20 Anvendeligheten av fremgangsmåten med syntetisk gen ble utvidet til tomat med et syntetisk B.t.k. HD-73-gen.

I tomat utviste ekstrakter fra planter inneholdende det villtype-avkortede HD-73-gen (pMON5367), ikke noe påvisbart HD-73-protein. Ekstrakter fra planter inneholdende
25 det syntetiske HD-73-gen (pMON5383), utviste høye nivåer av B.t.k. HD-73-protein, omtrent 2000 ng pr. mg plantekstraktprotein. Disse dataene viser klart at endringer som ble gjort i det syntetiske HD-73-gen, førte til dramatiske økninger i ekspresjonen av HD-73-proteinet i tomat, samt i
30 tobakk.

I motsetning til tobakk uttrykkes det syntetiske HD-73-gen i tomat ved omtrent 4 ganger til 5 ganger høyere nivåer enn det syntetiske HD-1-gen. Ettersom HD-73-proteinet er omtrent 5 ganger mer aktivt enn HD-1-proteinet mot mange
35 skadeinsekter, inkludert Heliothis-arter, svarer den økte ekspresjon av syntetisk HD-73 sammenlignet med syntetisk HD-1, til ca. en 25 ganger forøkt insekticid virkningsfullhet i tomat.

For å bestemme mekanismene som er involvert i den økte ekspresjon av modifiserte og syntetiske B.t.k. HD-1-gener i tomat, ble det utført S1-nuklease-analyse av mRNA-nivåer fra transformerte tomatplanter. Som angitt ovenfor, var en lignende analyse blitt utført med tobakksplanter, og denne analysen viste at det modifiserte gen ga opptil 10 ganger mer mRNA enn villtypegenet. Ved analysen i tomat ble det benyttet en annen DNA-probe som muliggjorde analysen av villtype- (pMON9921), det modifiserte (pMON5370) og det syntetiske (pMON5377) HD-1-genet med den samme probe. Denne proben var avledet fra det 5'-avkortede område i CaMV35S-promoterens i pMON893 som var felles for alle tre vektorene (pMON9921, pMON5370 og pMON5377). Denne S1-analyse indikerte at B.t.k.-mRNA-nivåer fra det modifiserte gen var 3 til 5 ganger høyere enn for villtypegenet, og at mRNA-nivåer for det syntetiske gen var ca. 2 til 3 ganger høyere enn for det modifiserte gen. Tre uavhengige transformanter ble analysert for hvert gen. Sammenlignet med ganger økninger i B.t.k. HD-1-protein fra disse genene i tomat vist i tabell XI, kan disse mRNA-økninger forklare ca. halvparten av den totale proteinøkning som ble sett i tobakk for villtype- og det modifiserte gen. For tomat er den totale mRNA-økning fra villtype til syntetisk ca. 6 til 15 ganger sammenlignet med en proteinøkning på ca. 50 ganger. Dette resultatet er likt med det som ses for tobakk når villtype- og modifiserte gener sammenlignes, og det gjelder også for det syntetiske gen. Dvs. at omtrent halvparten av de totale ganger økning i B.t.k.-protein fra villtype til modifiserte gener kan forklares ved mRNA-økninger, og ca. halvparten med forøkt translasjonseffektivitet. Det samme gjelder også når det modifiserte gen sammenlignes med det syntetiske gen. Selv om det er en tilleggsøkning i RNA-nivåer, kan denne mRNA-økning bare forklare ca. halvparten av den totale proteinøkning.

Fullengde-B.t.k.-genene beskrevet ovenfor, ble også anvendt til å transformere tomatplanter, og disse plantene ble analysert med hensyn på B.t.k.-protein og insekticid

- virkningsfullhet. Resultatene av denne analyse er vist i tabell XIII. Planter inneholdende det syntetiske/villtype-gen (pMON10506), produserer B.t.k. HD-73-proteinet ved nivåer på ca. 0,01% av deres totale oppløselige protein.
- 5 Planter inneholdende det syntetiske/modifiserte gen (pMON10526), produserer ca. 0,04% B.t.k.-protein, og planter inneholdende det fullstendig syntetiske gen (pMON10518), produserer ca. 0,2% B.t.k.-protein. Disse resultatene er svært like med tobakksplanteresultatene for de samme genene.
- 10 mRNA-nivåer beregnet ved hjelp av "Northern blot"-analyse i tomat, øker også parallelt med proteinnivåøkningen. Som for tobakk med disse tre genene, kan mesteparten av proteinøkningen tilskrives økt mRNA med en liten komponent av translasjonseffektivitetsøkning indikert for det fullstendig
- 15 syntetiske gen. De høyeste nivåene av fullengde-B.t.k.-protein (fra pMON10518) er sammenlignbare med eller bare litt lavere enn de høyeste nivåene som ble observert for de avkortede HD-73-genene (pMON5383 og pMON5390). Tomatplanter som uttrykker disse fullengdegenene, har den
- 20 insekticide aktivitet som forventes for de observerte proteinnivåer, slik de bestemmes ved fôringsanalyser med sukkerroe-hærlarve eller ved diettinkorporering av planteekstrakter med tobakkshornlarve.

25

Tabell XIII

B.t.k. HD-73 fullengdeprotein og mRNA-nivåer i transgene tomatplanter

30	<u>Gen- beskrivelse</u>	<u>Vektor</u>	<u>B.t.k.-protein- konsentrasjon</u>	<u>Relatvt B.t.k.- mRNA-nivå</u>
	Syntetisk/villtype	pMON10506	100	1
	Syntetisk/modifisert	pMON10526	400	2-4
35	Fullstendig syntetisk	pMON10518	2000	10

c) Bomull.

Almengyldigheten av den økte ekspresjon av B.t.k. HD-1 og B.t.k. HD-73 ved bruk av de modifiserte og syntetiske gener ble utvidet til bomull. Det ble fremstilt transgene kalli som inneholder villtypegenet (pMON9921) og det syntetiske HD-1-gen (pMON5377). Her ble igjen B.t.k. HD-1-proteinet produsert fra kalli inneholdende villtypegenet, ikke påvist, mens kalli inneholdende det syntetiske HD-1-gen, uttrykte HD-1-proteinet ved lett påvisbare nivåer. HD-1-proteinet ble fremstilt ved omtrent 1000 ng/mg av plante-kalliekstraktprotein. For igjen å sikre at proteinet fremstilt ved hjelp av de transgene bomullskalli var biologisk aktivt og at den økte ekspresjon som ble iakttatt med det syntetiske gen translatert til økt biologisk aktivitet, ble det igjen laget ekstrakter av bomullskalli på lignende måte som beskrevet for tobakksplanter, bortsett fra at kalliene først ble tørket mellom Whatman filterpapir for å fjerne så mye som mulig av vannet. De tørkede kalli ble så oppmalt i flytende nitrogen og oppmalt i 100 mM natriumcarbonat-buffer, pH 10. Omtrent 0,5 ml aliquoter av dette materiale ble tilført tomatblader med en malingspensel. Etter at bladet hadde tørket, ble fem tobakkshornlarver påført hver av de to bladprøvene. Blader malt med ekstrakt fra kontrollkalli, ble fullstendig ødelagt. Blader malt med ekstrakt fra kalli inneholdende villtype-HD-1-genet (pMON9921), hadde alvorlig skade. Blader malt med ekstrakt fra kalli inneholdende det syntetiske HD-1-gen (pMON5377), hadde ingen skade (se tabell XIV nedenunder).

Tabell XIV

Beskyttelse mot tobakkshornlarve ved tomatblader
malt med ekstrakter fremstilt fra bomullskalli
 inneholdende en kontroll, villtype-B.t.k.-HD-1-genet,
 det syntetiske HD-1-gen eller det syntetiske HD-73-gen

Gen- beskrivelse	Vektor	Tobakkshornlarve- skade*
Kontroll	Kontroll	NL
Villtype-HD-1	pMON9921	3
Syntetisk HD-1	pMON5377	0
Syntetisk HD-73	pMON5383	0

* Skade ble gradert som vist i tabell VIII.

Det ble også fremstilt bomullskalli inneholdende et annet syntetisk gen, et gen som koder for B.t.k. HD-73. Fremstillingen av dette genet er beskrevet i eksempel 3. Kalli inneholdende det syntetiske HD-73-gen, produserte det tilsvarende HD-73-protein med enda høyere nivåer enn kalliene som inneholdt det syntetiske HD-1-gen. Ekstrakter som ble laget fra kalli inneholdende det syntetiske HD-73-gen (pMON5383), utviste fullstendig kontroll med tobakkshornlarver når de ble malt på tomatblader som beskrevet ovenfor for ekstrakter inneholdende HD-1-proteinet. (Se tabell XIV).

Transgene bomullsplanter inneholdende det syntetiske B.t.k. HD-1-gen (pMON5377) eller det syntetiske B.t.k. HD-73-gen (pMON5383), er også blitt undersøkt. Disse plantene produserer HD-1- eller HD-73-proteinene ved nivåer som er sammenlignbare med det som ses i bomullskallus med de samme genene, og er sammenlignbare med tomat- og tobakksplanter med disse genene. Både for syntetisk, avkortet HD-1- og HD-73-gen ble bomullsplanter som uttrykte B.t.k.-protein ved 1000 til 2000 ng/mg totalt protein (0,1-0,2%), utvunnet ved

• en høy frekvens. Insektsføringsanalyser ble utført med blader fra bomullsplanter som uttrykker det syntetiske HD-1- eller HD-73-gen. Disse bladene hadde ingen skade (gradering på 0) når de ble utfordret med larver av kål-
måler (*Trichoplusia ni*), og bare liten skade når de ble ut-
fordret med sukkerroe-hærlarver (*Spodoptera exigua*). Skade-
graderinger er som definert i tabell VIII ovenfor. Dette
viser at bomullsplanter samt kalli uttrykte de syntetiske
HD-1- eller HD-73-gener ved høye nivåer, og at disse plantene
ble beskyttet mot skade av sommerfuglinsektlarver.

Transgene bomullsplanter inneholdende enten det syntetiske, avkortede HD-1-gen (pMON5377) eller det syntetiske, avkortede HD-73-gen (pMON5383), ble også bedømt med hensyn på beskyttelse mot bomullsfrøkapsellarve ved hel-
plantenivået i drivhus. Dette er en mer realistisk test på
disse plantenes evne til å gi et landbruksmessig akseptabelt
kontrollnivå. Bomullsfrøkapsellarven (*Heliothis zea*) er et
viktig skadeinsekt for bomull som gir økonomisk skade ved
å ødelegge ender, terninger og frøkapsler, og beskyttelse
av disse fruktlegemene samt bladvevet vil være viktig for
effektiv insektkontroll og adekvat avlingsbeskyttelse. For
å teste beskyttelsen gitt til helplanter, ble R1-avkom av
bomullsplanter som uttrykker høye nivåer av enten B.t.k.
HD-1 (pMON5377) eller B.t.k. HD-73 (pMON5383), analysert ved
å tilføre 10-15 egg av bomullsfrøkapsellarver pr. frøkapsel
eller terning til de 20 øverste terninger eller frøkapsler
på hver plante. Minst 12 planter ble analysert pr. behand-
ling. Utklekkingsgraden for eggene var omtrent 70%. Dette
svarer til svært høyt insekttrykk sammenlignet med antall
larver pr. plante som ses under typiske frilandsbetingelser.
Under disse betingelsene ble 100% av frøkapslene på kon-
trollbomullsplanter ødelagt ved insektskade. For de trans-
gene plantene ble det observert signifikant frøkapsel-
beskyttelse. Planter inneholdende pMON5377 (HD-1), hadde
70-75% frøkapseloverlevelse etter det sterke trykk ved
denne analyse. Planter inneholdende pMON5383 (HD-73),
hadde 80-90% frøkapselbeskyttelse. Dette er sannsynligvis

- en følge av den høyere aktivitet av HD-73-protein mot bomullsfrøkapsellarver sammenlignet med HD-1-protein. I tilfeller hvor de transgene plantene ble skadet av insektene, ble de overlevende larver forsinket i sin utvikling med minst ett stadium.

Den økte ekspresjon som ble oppnådd med de modifiserte og syntetiske gener, er derfor ikke begrenset til den ene avling; tobakk, tomat og bomullskalli og bomullsplanter utviste alle drastiske økninger i B.t.k.-ekspresjon når plantene/kalliene ble fremstilt inneholdende de modifiserte eller syntetiske gener. Anvendbarheten av endringer gjort for å fremstille det modifiserte og syntetiske B.t.k. HD-1-gen, er likeledes ikke begrenset til HD-1-genet. Det syntetiske HD-73-gen utviste i alle tre artene også drastiske økninger i ekspresjon.

Det er således blitt vist at: (1) de genetiske endringene som er gjort i det HD-1-modifiserte gen, fører til svært betydelige økninger i B.t.k. HD-1-ekspresjon; (2) fremstilling av et totalt syntetisk gen fører til en ytterligere 5-gangers økning i B.t.k. HD-1-ekspresjon; (3) endringene inkorporert i det modifiserte HD-1-gen, var ansvarlige for mesteparten av den økte B.t.k.-ekspresjon som ble observert med det syntetiske gen; (4) den økte ekspresjon ble demonstrert i tre forskjellige planter: tobakksplanter, tomatplanter og bomullskalli og bomullsplanter; (5) den økte ekspresjon som observert med Western-analyse, korrelerte også med lignende økninger i biologisk aktivitet, noe som viser at B.t.k. HD-1-proteinene som ble fremstilt, var sammenlignbart aktive; (6) når fremgangsmåten brukt til å utforme det syntetiske HD-1-gen ble anvendt til å utforme et syntetisk HD-73-gen, ble det også uttrykt ved mye høyere nivåer i tobakk, tomat og bomull enn villtype-ekvivalentgenet med økninger i biologisk aktivitet som følge derav; (7) et fullstendig syntetisk fullengde-B.t.k.-gen ble uttrykt ved nivåer som var sammenlignbare med nivåene til syntetiske, avkortede gener.

• Eksempel 5 -- Syntetisk B.t. tenebrionis-gen i tobakk, tomat og potet

Ifølge figur 12 fremstilles et syntetisk gen som koder for et sommerfuglaktivt toksin ved å fremstille de angitte endringer i villtypegenet til B.t. tenebrionis, eller de novo syntese av det syntetiske strukturgenet. Det syntetiske gen innføres i en intermediær plantetransformasjonsvektor slik som pMON893. Plasmid pMON893 som inneholder det syntetiske B.t.t.-gen, innføres så i en egnet, ufarliggjort Agrobacterium-stamme, slik som A. tumefaciens ACO.

Transformasjon og regenerering av potet

Sterile skuddkulturer av Russet Burbank oppbevares i ampuller som inneholder 10 ml PM-medium (Murashige and Skoog (MS) uorganiske salter, 30 g/l sukrose, 0,17 g/l $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$, 0,4 mg/l thiamin-HCl og 100 mg/l myo-inositol, stivnet med 1 g/l "Gelrite" ved pH 6,0). Når skuddene nådde en omtrentlig lengde på 5 cm, ble stengelleddstykker på 7-10 mm skåret ut og påsmurt i kuttendene med en uskadeliggjort Agrobacterium tumefaciens-vektor som inneholdt det syntetiske B.t.t.-gen fra en fire dager gammel platekultur. Stengeleksplantatene ble samdyrket i tre dager ved 23°C på et sterilt filterpapir plassert over 1,5 ml tobakkcelleføringslag lagt oppå 1/10 P medium (1/10 styrke MS uorganiske salter og organiske tilsetninger uten kasein som i Jarret et al. (1980), 30 g/l sukrose og 8,0 g/l agar). Etter samdyrking ble eksplantatene overført til P-1-medium med full styrke for kallusinduksjon, sammensatt av MS uorganiske salter, organiske tilsetninger som i Jarret et al. (1980), med unntak av kasein, 3,0 mg/l benzyladenin (BA) og 0,01 mg/l naphthalenediksyre (NAA) (Jarret et al., 1980). Carbenicillin (500 mg/l) er inkludert for å inhibere bakterievekst, og 100 mg/l kanamycin tilsettes for å selektere for transformerte celler. Etter fire uker overføres eksplantatene til medium med den samme sammensetning, men med 0,3 mg/l gibberellinsyre (GA3) som erstatter BA og NAA

(Jarret et al., 1981) for å fremme skuddannelse. Skudd begynner å utvikles omtrent to uker etter overføring til skuddinduksjonsmedium; disse skjæres bort og overføres til ampuller med PM-medium for rotdannelse. Skudd testes med
5 hensyn på kanamycinresistens tilført ved hjelp av enzymet neomycinfosfotransferase II, ved å plassere et snitt av stengelen på kallusinduksjonsmedium inneholdende MS organiske og uorganiske salter, 30 g/l sukrose, 2,25 mg/l BA, 0,186 mg/l NAA, 10 mg/l GA3 (Webb et al., 1983) og 200 mg/l
10 kanamycin for å selektere for transformerte celler.

Det syntetiske B.t.t.-gen beskrevet i figur 12, ble plassert på en planteekspresjonsvektor som beskrevet i eksempel 5. Plasmidet har de følgende karakteristika; et syntetisk BglIII-fragment med omtrent 1800 basepar ble inn-
15 føyet i pMON893 på en slik måte at den forsterkede 35S-promoter ville uttrykke B.t.t.-genet. Denne konstruksjonen, pMON1982, ble brukt til å transformere både tobakk og tomat. Tobakksplanter valgt ut som kanamycinresistente planter, ble undersøkt med kanin-anti-B.t.t.-antistoff. Kryssreaktivt
20 materiale ble påvist ved nivåer som var forutsagt å være egnet til å forårsake dødelighet for CPB. Disse mål-insekter vil ikke innta næring på tobakk, men de transgene tobakksplantene viser at det syntetiske gen forbedrer ekspresjon av dette protein til påvisbare nivåer.

25 Tomatplanter med pMON1982-konstruksjonen ble bestemt å produsere B.t.t.-protein ved nivåer som er insekticide for CPB. I innledende undersøkelser viste bladene fra fire planter (5190, 5225, 5328 og 5133) liten eller ingen skade når de ble eksponert mot CPB-larver (skadegradering 0-1 på
30 en skala fra 0 til 4 med 4 som ikke noe blad tilbake). Under disse betingelsene ble kontrollbladene fullstendig oppspist. Immunologisk analyse av disse planter bekreftet tilstedeværelsen av betydelig kryssreaktivitet med anti-B.t.t.-antistoff. Nivåer av proteinekspresjon i disse
35 plantene ble beregnet ved omtrent 1-5 ng B.t.t.-protein i 50 µg av alt ekstraherbart protein. Det er i alt blitt identifisert 17 tomatplanter (17 av 65 testede) som har be-

• skyttelse av bladvev mot CPB (gradering på 0 eller 1) og utviser god insektdødelighet.

Lignende resultater som de som ses på tobakk og tomat med pMON1982, kunne også ses med pMON1984 på de samme plantearter. pMON1984 er identisk med pMON1982 bortsett fra at den syntetiske proteaseinhibitor (CMTI) er fusjonert oppstrøms fra det naturlig forekommende, proteolytiske spaltingssted. Ekspresjonsnivåer i tobakk ble beregnet til å være like med pMON1982, mellom 10-15 ng pr. 50 µg av alt oppløselig protein.

Det er blitt identifisert tomatplanter som uttrykker pMON1984 som beskytter bladene mot inntak av CPB. Skadegraderingen var 0 med 100% insektdødelighet.

Potet ble transformert som beskrevet i eksempel 5 med en lignende vektor som pMON1982 inneholdende det forsterkede CaMV35S/syntetiske B.t.t.-gen. Blader fra potetplanter transformert med denne vektoren, ble undersøkt ved hjelp av biologisk insektanalyse av CPB. Av de 35 plantene som ble testet, var blader fra 4 planter, 16a, 13c, 13d og 23a, totalt beskyttet når de ble utfordret. Biologiske insektanalyser med blader fra tre andre planter, 13e, 1a og 13b, målte skadenivåer på 1 på en skala fra 0 til 4, idet 4 var total ødeleggelse av bladmaterialet. Immunologisk analyse bekreftet tilstedeværelsen av B.t.t.-kryssreaktivt materiale i bladvevet. Nivået av B.t.t.-protein i bladvev fra plante 16a (skadegradering på 0) ble beregnet til 20-50 ng av B.t.t.-protein/50 µg av alt oppløselig protein. Nivåene av B.t.t.-protein som ses i 16a-vev, var i overensstemmelse med dets biologiske aktivitet. Immunologiske analyser av 13e og 13b (vev som scoret 1 i skadegradering), avslører mindre protein (5-10 ng/50 µg totalt oppløselig protein) enn i plante 16a. Avkutt fra plante 16a ble utfordret med 50-200 egg CPB i en helplantanalyse. Under disse betingelsene viste 16a ingen skade og 100% dødelighet for insekter, mens kontrollpotetplanter ble svært skadet.

• Eksempel 6 -- Syntetisk B.t.k. P2-proteingen

P2-proteinet er et særskilt, insekticid protein fremstilt av noen stammer av B.t. inkludert B.t.k. HD-1. Det er kjennetegnet ved sin aktivitet mot både sommerfugl- og tovingede insekter (Yamamoto and Iizuka, 1983). Gener som koder for P2-proteinet, er blitt isolert og karakterisert (Donovan et al., 1988). P2-proteinene som kodes for av disse genene, er omtrent 600 aminosyrer lange. Disse proteinene har bare begrenset homologi til felles med de sommerfugl-spesifikke P1-type proteiner, slik som B.t.k. HD-1- og HD-73-proteinene beskrevet i tidligere eksempler.

P2-proteinene har vesentlig aktivitet mot flere forskjellige sommerfugllarver, inkludert kålmåler, tobakkshornlarver og tobakksknopplarver. Ettersom de er aktive mot landbruksmessig viktige skadeinsekter, er P2-proteinene en ønskelig kandidat ved fremstillingen av insekttolerante, transgene planter, enten alene eller i kombinasjon med de øvrige B.t.-toksinene beskrevet i eksemplene ovenfor. I noen planter kan ekspresjon av P2-proteinet alene være tilstrekkelig til å gi beskyttelse mot skadeinsekter. I tillegg kan P2-proteinene gi beskyttelse mot landbruksmessig viktige, tovingede skadeinsekter. I andre tilfeller kan ekspresjon av P2 sammen med B.t.k. HD-1- eller HD-73-proteinet være foretrukket. P2-proteinene bør gi i det minste et tilleggsnivå av insekticid aktivitet når de kombineres med krystallproteintoksinet fra B.t.k. HD-1 eller HD-73, og kombinasjonen kan til og med gi en synergistisk virkning. Selv om virkningsmåten til P2-proteinet er ukjent, tyder dets entydige aminosyresekvens på at det virker forskjellig fra B.t.k. HD-1- og HD-73-typen av proteiner. Fremstilling av to insekttoleranseproteiner med forskjellige virkningsmåter i den samme plante ville minimalisere muligheten for utvikling av insektresistens mot B.t.-proteiner i planter. Mangelen på vesentlig DNA-homologi mellom P2-gener og HD-1- og HD-73-genene minimaliserer muligheten for rekombinasjon mellom multiple insekttoleransegener i plantekromosomet.

Genene som koder for P2-proteinet, har, selv om de er forskjellige i sekvens fra B.t.k. HD-1- og HD-73-genene, til felles mange trekk med disse genene. Særlig har P2-proteingenene et høyt A+T-innhold (65%), flere potensielle polyadenyleringssignalsekvenser (26) og et stort antall ATTTA-sekvenser (10). På grunn av den totale likhet med de dårlig uttrykte villtype B.t.k. HD-1- og HD-73-genene, forventes de samme problemene med ekspresjon av villtype-P2-genet som man støtte på i de tidligere eksemplene. Basert på den ovenfor beskrevne fremgangsmåte for utforming av de syntetiske B.t.-gener, er det blitt utformet et syntetisk P2-gen som burde bli uttrykt ved adekvate nivåer for beskyttelse i planter. En sammenligning av villtype- og de syntetiske P2-genene er vist i figur 13.

Eksempel 7 - Syntetisk B.t. entomocidus-gen

B.t. entomocidus- ("Btent") proteinet er et entydig insekticid protein produsert av noen stammer av B.t.-bakterier. Det er kjennetegnet ved sitt høye aktivitetsnivå mot noen sommerfuglarter som er forholdsvis ufølsomme for B.t.k. HD-1 og HD-73, slik som Spodoptera-arter, inkludert sukkerroe-hærlarver (Visser et al., 1988). Gener som koder for Btent-proteinet, er blitt isolert og karakterisert (Honee et al., 1988). Btent-proteinene som kodes for av disse genene, har omtrent den samme lengde som B.t.k. HD-1 og HD-73. Disse proteinene har til felles bare 68% aminosyrehomologi med B.t.k. HD-1- og HD-73-proteinene. Det er sannsynlig at bare N-ende-halvdelen av Btent-proteinet er påkrevet for insekticid aktivitet, slik tilfellet er for HD-1 og HD-73. Blant de første 625 aminosyrene har Btent bare til felles 38% aminosyrehomologi med HD-1 og HD-73.

På grunn av deres sterkere aktivitet mot Spodoptera-arter som er forholdsvis ufølsomme for HD-1 og HD-73, er Btent-proteinene en ønskelig kandidat for fremstillingen av insekttolerante, transgene planter, enten alene eller i kombinasjon med de øvrige B.t.-toksiner beskrevet i eksemplene ovenfor. I noen planter kan fremstilling av Btent alene

være tilstrekkelig for å kontrollere de landbruksmessig viktige skadeinsekter. I andre planter ville produksjonen av to entydige insekttoleranseproteiner gi beskyttelse mot en større rekke insekter. Mot disse insektene hvor begge
5 proteinene er aktive, burde kombinasjonen av B.t.k. HD-1- eller HD-73-type proteinet pluss Btent-proteinet gi minst additiv insekticid virkningsfullhet og kan til og med gi en synergistisk virkning. På grunn av den entydige aminosyre-sekvens kan Btent-proteinet i tillegg ha en annen virknings-
10 måte enn HD-1 eller HD-73. Produksjon av to insekticide proteiner i den samme plante med forskjellige virkningsmåter ville minimalisere muligheten for utvikling av insektresistens mot B.t.-proteiner i planter. Den relative
15 mangel på DNA-sekvenshomologi med B.t.k.-type genene minimaliserer muligheten for rekombinasjon mellom multiple insekttoleransegener i plantekromosomet.

Genene som koder for Btent-proteinet, har, selv om de er forskjellige i sekvens fra B.t.k. HD-1- og HD-73-genene, mange fellestrekk med disse genene. Særlig har
20 Btent-proteinet et høyt A+T-innhold (62%), flere mulige polyadenyleringssignalsekvenser (39 i den fullengde-kodende sekvens og 27 i de første 1875 nukleotidene som sannsynligvis koder for det aktivt toksiske fragment) og et stort antall ATTTA-sekvenser (16 i den fullengde-
25 kodende sekvens og 12 i de første 1875 nukleotidene). På grunn av den totale likhet med de dårlig uttrykte villtype B.t.k. HD-1- og HD-73-genene, forventes villtype Btent-genene å oppvise lignende problemer med ekspresjon som man
30 støtte på med villtype HD-1- og HD-73-genene. Basert på den ovenfor beskrevne fremgangsmåte som ble brukt til å utforme de øvrige syntetiske B.t.-gener, er et syntetisk Btent-gen blitt utformet som burde bli uttrykt ved adekvate nivåer for beskyttelse i planter. En sammenligning av villtype- og
35 det syntetiske Btent-gen er vist i figur 14.

• Eksempel 8 -- Syntetiske B.t.k.-gener for ekspresjon i mais

Høynivåekspresjon av heterologe gener i maisceller er blitt vist å bli forsterket ved tilstedeværelsen av et
5 maisgenintron (Callis et al., 1987). Vanligvis har disse intronene blitt lokalisert i det 5'-utranslaterte område av kimærgenet. Det er blitt vist at CaMV35S-promoteren og NOS 3'-enden virker effektivt ved ekspresjonen av heterologe gener i maisceller (Fromm et al., 1986).

10 Ifølge figur 15 ble en planteekspresjonskasset-vektor (pMON744) konstruert som inneholder disse sekvensene. Nærmere bestemt inneholder ekspresjonskassetten den forsterkede CaMV 35S-promoter etterfulgt av intron 1 i mais-Adhl-genet (Callis et al., 1987). Dette etterfølges av et
15 multilinkerkloningssete for innføring av kodende sekvenser; denne multilinkeren inneholder bl.a. et BglII-sete. Etter multilinkeren er NOS-3'-enden. pMON744 inneholder også det selekterbare markørgen 35S/NPTII/NOS-3' for kanamycin-seleksjon av transgene maisceller. I tillegg har pMON744
20 et E. coli opprinnelsessted for replikasjon og et ampicillin-resistensgen for seleksjon av plasmidet i E. coli.

Fem B.t.k. kodende sekvenser beskrevet i de tidligere eksemplene, ble innført i BglII-setet til pMON744 for maiscelleekspresjon av B.t.k. De kodende sekvenser inn-
25 føyet og de resulterende vektorer var:

1. Villtype B.t.k. HD-1 fra pMON9921 for å lage pMON8652.

30 2. Modifisert B.t.k. HD-1 fra pMON5370 for å lage pMON8642.

3. Syntetisk B.t.k. HD-1 fra pMON5377 for å lage pMON8643.

4. Syntetisk B.t.k. HD-73 fra pMON5390 for å lage pMON8644.

35 5. Syntetisk fullengde B.t.k. HD-73 fra pMON10518 for fremstilling av pMON10902.

pMON8652 (villtype B.t.k. HD-1) ble brukt til å transformere maiscelleprotoplastere, og stabilt transformert kanamycinresistent kallus ble isolert. B.t.k.-mRNA i maiscellene ble analysert ved hjelp av nuklease S1-beskyttelse og funnet å være til stede ved et nivå som er sammenlignbart med det som ses med den samme villtype-kodende sekvens (pMON9921) i transgene tomatplanter.

pMON8652 og pMON8642 (modifisert HD-1) ble brukt til å transformere maiscelleprotoplastere i et ubestendig ekspresjonssystem. Nivået av B.t.k.-mRNA ble analysert ved hjelp av nuklease S1-beskyttelse. Den modifiserte HD-1 ga opphav til flere ganger økning i B.t.k.-mRNA sammenlignet med den villtype-kodende sekvens i de forbigående transformerte maisceller. Dette indikerte at modifikasjonene innført i B.t.k. HD-1-genet, er i stand til å fremme B.t.k.-ekspresjon i énfrøbladede celler, slik det ble demonstrert for tofrøbladede planter og celler.

pMON8642 (modifisert HD-1) og pMON8643 (syntetisk HD-1) ble brukt til å transformere "Black Mexican Sweet" (BMS) maiscelleprotoplastere ved hjelp av PEG-mediert DNA-opptak, og stabilt transformert maiskallus ble valgt ut ved dyrking på kanamycinholdig plantedyrkningsmedium. Individuelle kalluskolonier som var avledet fra enkelttransformerte celler, ble isolert og formert separat på kanamycinholdig medium.

For å fastslå ekspresjonen av B.t.k.-genene i disse cellene ble kallusprøver testet med hensyn på insekttoksisitet ved biologisk analyse mot tobakkhornlarver. For hver vektor ble 96 kalluslinjer testet ved biologisk analyse. Porsjoner av hver kallus ble plassert på sterile vannagarplater, og fem nyfødte tobakkhornlarver ble tilsatt og fikk ta til seg næring i 4 dager. For pMON8643 døde 100% av larvene etter å ha tatt til seg næring på 15 av de 96 kalli, og disse kalliene utviste liten næringsskade. For pMON8642 var bare 1 av de 96 kalliene toksisk for larvene. Dette viste at B.t.k.-genet ble uttrykt i disse prøvene ved insekticide nivåer. Observasjonen at betydelig mer kalli

• inneholdende pMON8643 var toksiske enn for pMON8642, viste at betydelig høyere nivåer av ekspresjon ble oppnådd når den syntetiske HD-1-kodende sekvens var inneholdt i mais-celler, enn når den modifiserte HD-1-kodende sekvens ble
5 brukt, på samme måte som ved de tidligere eksempler med tofrøbladete planter. En semikvantitativ, immunologisk analyse viste at de toksiske pMON8643-prøvene hadde betydelig høyere B.t.k.-proteinnivåer enn den toksiske pMON8642-prøve.

10 De 16 kallusprøvene som var toksiske for tobakk-hornlarver, ble også testet med hensyn på aktivitet mot europeisk maisborer. Europeisk maisborer er omtrent 40 ganger mindre følsom for HD-1-genproduktet enn det tobakk-hornlarver er. Larver av europeisk kornborer ble tilført
15 kallusprøvene og fikk innta næring i 4 dager. To av de 16 kalliene som ble testet, som begge inneholdt pMON8643 (syntetisk HD-1), var toksiske overfor europeiske maisborer-larver.

20 For å fastslå ekspresjonen av B.t.k.-genene i differensiert maisvev, ble det brukt en annen fremgangsmåte for DNA-avlevering. Unge blader ble skåret bort fra maisplanter, og DNA-prøver ble ført inn i bladvevet ved mikroprosjektil-bombardement. I dette systemet uttrykkes DNA-en på mikroprosjektilene forbigående i bladcellene etter bombardement.
25 Tre DNA-prøver ble brukt, og hver DNA ble testet in triplo.

1. pMON744, maisekspresjonsvektoren uten noe B.t.k.-gen.
2. pMON8643 (syntetisk HD-1).
- 30 3. pMON752, en maisekspresjonsvektor for GUS-genet, ikke noe B.t.k.-gen.

35 Bladene ble inkubert ved værelsetemperatur i 24 timer. pMON752-prøvene ble farget med et substrat som muliggjorde visuell påvisning av GUS-genproduktet. Denne analysen viste at over 100 flekker i hver prøve uttrykte GUS-produktet, og prøvene in triplo oppviste svært like nivåer av GUS-eks-

- presjon. For pMON744 og pMON8643-prøvene ble 5 tobakk-hornlarver tilsatt til hvert blad og fikk innta næring i 48 timer. Alle tre prøvene bombardert med pMON744, oppviste utstrakt næringsskade og ingen larvedødelighet. Alle tre prøver bombardert med pMON744, oppviste ingen tegn på næringsskade og 100% larvedødelighet. Prøvene ble også analysert med hensyn på tilstedeværelsen av B.t.k.-protein ved hjelp av en kvalitativ immunologisk analyse. Alle pMON8643-prøvene hadde påvisbart B.t.k.-protein. Disse resultatene viste at det syntetiske B.t.k.-gen ble uttrykt i differensiert maisplantevev ved insekticide nivåer.

Eksempel 9 -- Syntetisk potetblad-rullevirus-kappeproteingen

Ekspresjon i planter av kappeproteingener fra flere forskjellige plantevirus har vist seg å være en effektiv metode for å innarbeide resistens mot disse virus. For å oppnå virusresistens er det viktig å uttrykke virus-kappeproteinet ved et effektivt nivå. For mange plantevirus-kappeproteingener har dette vist seg å ikke være noe problem. For kappeproteingenet fra potetblad-rullevirus (PLRV) har imidlertid ekspresjon av kappeproteinet blitt observert å være lavt i forhold til andre kappeproteingener, og dette lavere proteinnivå har ikke ført til optimal resistens mot PLRV.

Genet for PLRV-kappeprotein er vist i figur 16. I figur 16 viser den øvre sekvenslinje genet slik det opprinnelig ble konstruert for planteekspresjon i vektor pMON893. Genet var inneholdt i et 749 nukleotider langt BglII-EcoRI-fragment med den kodende sekvens inneholdt mellom nukleotidene 20 og 643. Dette fragmentet inneholdt også 19 nukleotider med 5'-ikke-kodende sekvens og 104 nukleotider med 3'-ikke-kodende sekvens. Dette PLRV-kappeproteingen ble forholdsvis dårlig uttrykt i planter sammenlignet med andre viruskappeproteingener.

Et syntetisk gen ble utformet for å forbedre planteekspresjon av PLRV-kappeproteinet. I figur 16 er videre endringene gjort i det syntetiske PLRV-gen vist i den nedre

linje. Dette genet ble utformet for å kode nøyaktig for det samme protein som det naturlig forekommende gen. Legg merke til at begynnelsen av det syntetiske gen er ved nukleotid 14, og slutten på sekvensen er ved nukleotid 654. Den kodende sekvens for det syntetiske gen er fra nukleotid 20 til 643 i figuren. Endringene angitt like oppstrøms og nedstrøms for disse endepunktene, tjener bare til å innføre passende restriksjonssteder like utenfor den kodende sekvens. Størrelsen på det syntetiske gen er således 641 nukleotider som er mindre enn det naturlig forekommende gen. Det syntetiske gen er mindre fordi praktisk talt alt av den ikke-kodende sekvens både i 5'- og 3'-enden, bortsett fra segmenter som koder for BglII og EcoRI-restriksjonssetene, er blitt fjernet.

Det syntetiske gen adskiller seg fra det naturlig forekommende gen i to hovedhenseender. For det første er 41 individuelle kodoner innenfor den kodende sekvens blitt endret for å fjerne nesten alle kodoner for en bestemt aminosyre som utgjør mindre enn ca. 15% av kodonene for den aminosyren i en oversikt over tofrøbladete plantegener. For det andre er de 5'- og 3'-ikke-kodende sekvenser for det opprinnelige gen blitt fjernet. Selv om det ikke stemmer nøyaktig overens med algoritmen beskrevet i figur 1, er noen få av kodonendringene og spesielt fjerningen av det lange, 3'-ikke-kodende område i overensstemmelse med denne algoritme.

Den opprinnelige PLRV-sekvens inneholder to mulige plante-polyadenyleringssignaler (AACCAA og AAGCAT), og begge disse opptrer i den 3'-ikke-kodende sekvens som er blitt fjernet i det syntetiske gen. Det opprinnelige PLRV-gen inneholder også en ATTTA-sekvens. Denne er også inneholdt i den 3'-ikke-kodende sekvens, og er i midten av den lengste strekningen av uavbrutt A+T i genet (en strekning på 7 A+T-nukleotider). Denne sekvensen ble fjernet i det syntetiske gen. Sekvenser som algoritmen ifølge figur 1 har som mål for endring, er således blitt endret i det syntetiske PLRV-kappeproteingen ved fjerning av det 3'-ikke-

- kodende segment. Innenfor den kodende sekvens ble kodon-
endringer også gjort for å fjerne tre andre sekvensområder
beskrevet ovenfor. Spesielt har to områder med 5 fortløp-
ende A+T og ett område med 5 fortløpende G+C innenfor den
5 kodende sekvens blitt fjernet i det syntetiske gen.

Det syntetiske PLRV-kappeproteingen klones inn i en
plantetransformasjonsvektor slik som pMON893, og anvendes
til å transformere potetplanter som beskrevet ovenfor.
Disse plantene uttrykker PLRV-kappeproteinet ved høyere
10 nivåer enn det som ble oppnådd med det naturlig forekommende
gen, og disse plantene utviser økt resistens mot infeksjon
med PLRV.

15 Eksempel 10 -- Ekspresjon av syntetiske B.t.-gener med
promoterer for den lille underenhet i RUBISCO, og kloro-
plastovergangspeptider

Genene i planter som koder for den lille underenhet
i RUBISCO (SSU), uttrykkes ofte sterkt, er ofte lette å
regulere og har noen ganger vevspesifisitet. Disse eks-
20 presjonsegenskapene skyldes stort sett promotersekvensene i
disse genene. Det er blitt mulig å bruke SSU-promoterer for
å uttrykke heterologe gener i transformerte planter. Van-
ligvis vil en plante inneholde flere SSU-gener, og eks-
presjonsnivåene og vevspesifisiteten til forskjellige SSU-
25 gener vil være forskjellige. SSU-proteinene kodes for i
kjernen og syntetiseres i cytoplasmaet som forløpere som
inneholder en N-endeforlengelse kjent som kloroplast-
overgangspeptidet (CTP). CTP styrer forløperen til kloro-
plasten og fremmer opptaket av SSU-proteinet i kloroplasten.
30 I denne prosessen spaltes CTP fra SSU-proteinet. Disse CTP-
sekvensene er blitt brukt til å styre heterologe proteiner
inn i kloroplastere hos transformerte planter.

SSU-promoterne kan ha flere fordeler for ekspresjon
av B.t.k.-gener i planter. Noen SSU-promoterer uttrykkes
35 svært sterkt og kunne gi opphav til like høye ekspresjons-
nivåer som, eller høyere enn, de som er observert med
CaMV35S-promoteren. Vevfordelingen av ekspresjon fra SSU-

• promoterer er forskjellig fra vevfordelingen for CaMV35S-promoterer, så for kontroll av noen skadeinsekter kan det derfor være fordelaktig å styre ekspresjonen av B.t.k. til de cellene hvor SSU uttrykkes sterkest. Selv om det er for-
5 holdsvi grunnleggende, så uttrykkes f.eks. i løvet CaMV35S-promoterer sterkere i vaskulært vev enn i noen andre deler av løvet, mens de fleste SSU-promoterer uttrykkes sterkest i mesofyllcellene i løvet. Noen SSU-promoterer er også mer vevspesifikke slik at det ville kunne være mulig å
10 utnytte en bestemt SSU-promoter til å uttrykke B.t.k. i bare et undersett av plantevev dersom f.eks. B.t.-ekspresjon i visse celler ble funnet å være skadelig for disse cellene. F.eks. kan det for kontroll av Coloradobille i potet være fordelaktig å bruke SSU-promoterer for å styre B.t.t.-eks-
15 presjon til bladene, men ikke til de spiselige rotknoller.

Benyttelse av SSU-CTP-sekvenser for å lokalisere B.t.-proteiner til kloroplasten, kan også være fordelaktig. Lokalisering av B.t. til kloroplasten kunne beskytte pro-
teinet mot proteaser som finnes i cytoplasmaet. Dette kunne
20 stabilisere B.t.-proteinet og føre til høyere nivåer av akkumulering av aktivt protein. B.t.-gener inneholdende CTP, kunne anvendes i kombinasjon med SSU-promoterer eller sammen med andre promoterer, slik som CaMV35S.

Flere forskjellige plantetransformasjonsvektorer ble
25 konstruert for ekspresjonen av B.t.k.-gener under benyttelse av SSU-promoterer og SSU-CTP-er. Promoterne og CTP-ene som ble benyttet, var fra petunia-SSUlla-genet beskrevet av Tumer et al. (1986), og fra Arabidopsis atslA-genet (et SSU-
gen) beskrevet av Krebbers et al. (1988) og av Elionor et
30 al. (1989). Petunia-SSUlla-promoterer var inneholdt på et DNA-fragment som strakte seg omtrent 800 bp oppstrøms fra den SSU-kodende sekvens. Arabidopsis atslA-promoterer var inneholdt på et DNA-fragment som strakte seg omtrent 1,8 kb
oppstrøms fra den SSU-kodende sekvens. I den oppstrøms ende
35 ble passende steder fra multilinkeren til pUC18 brukt til å forflytte disse promoterene inn i plantetransformasjonsvektorer, slik som pMON893. Disse promoterfragmenter strakte

- seg til starten for den SSU-kodende sekvens hvor et NcoI-restriksjonssete ble konstruert for å muliggjøre innføyelse av B.t.-kodesekvensen som erstattet SSU-kodesekvensen.

Når SSU-promoterer ble brukt i kombinasjon med deres CTP, strakte DNA-fragmentene seg gjennom den kodende sekvens av CTP, og en liten del av den kodende sekvens i fullt ferdig SSU, hvor et NcoI-restriksjonssete ble konstruert ved hjelp av standardteknikker for å muliggjøre fusjonen i ramme av B.t.-kodesekvenser med CTP. For petunia-SSUlla-CTP ble spesielt B.t.-kodesekvenser fusjonert til SSU-sekvensen etter aminosyre 8 i den fullt ferdige SSU-sekvens hvor NcoI-setet var plassert. De 8 aminosyrene i fullt ferdig SSU-sekvens var inkludert ettersom preliminært in vitro kloroplastopptaksforsøk indikerte at opptak av B.t.k. ble iaktatt bare dersom dette segmentet av fullt ferdig SSU var inkludert. For Arabidopsis atslA-CTP var komplett CTP inkludert pluss 24 aminosyrer av fullt ferdig SSU-sekvens pluss sekvensen gly-gly-arg-val-asn-cys-met-gln-ala-met, med avslutning i et NcoI-sete for B.t.-fusjon. Denne korte sekvensen gjentar det naturlig forekommende SSU-CTP-spaltningssete (mellom cys og met) pluss et kort segment som omgir spaltingssetet. Denne sekvensen ble inkludert for å sikre korrekt opptak i kloroplastere. B.t.-kodesekvenser ble fusjonert til denne atslA-CTP etter met-kodonet. In vitro opptaksforsøk med denne CTP-konstruksjon og andre (ikke-B.t.) kodende sekvenser viste at denne CTP målrettet proteiner mot kloroplasten.

Når CTP-er ble brukt i kombinasjon med CaMV35S-promoterer, ble de samme CTP-segmenter brukt. De ble skåret ut like oppstrøms for ATG-startsetene i CTP ved konstruksjon av BglII-seter og plassert nedstrøms for CaMV35S-promoterer i pMON893, som BglII-NcoI-fragmenter. B.t.-kodesekvenser ble fusjonert som beskrevet ovenfor.

Villtype-B.t.k. HD-1-kodesekvens i pMON9921 (se figur 1) ble fusjonert til atslA-promoterer for å lage pMON1925, eller atslA-promoterer pluss CTP for å lage pMON1921. Disse vektorene ble brukt til å transformere

tobakksplantene, og plantene ble undersøkt med hensyn på aktivitet mot tobakkhornlarver. Ingen toksiske planter ble utvunnet. Dette er overraskende i lys av det faktum at toksiske planter kunne bli utvunnet, riktignok ved en lav frekvens, etter transformasjon med pMON9921 hvor B.t.k.-kodesekvensen ble uttrykt fra den forsterkede CaMV35S, promoter i pMON893, og i lys av det faktum at Elionor et al. (1989) rapporterer at atslA-promoterer selv er sammenlignbar i styrke med CaMV35S-promoterer og omtrent 10 ganger sterkere når CTP-sekvensen er inkludert. Dette synes ikke å være tilfelle i det minste for villtype B.t.k. HD-1-kodesekvensen.

Det ble konstruert flere forskjellige plantetransformasjonsvektorer under benyttelse av enten den avkortede, syntetiske HD-73-kodesekvens ifølge figur 4, eller B.t.k. HD-73-kodesekvensen med full lengde ifølge figur 11. Disse er angitt i tabellen nedenunder.

Tabell XV

Genkonstruksjoner med CTP-er

<u>Vektor</u>	<u>Promoter</u>	<u>CTP</u>	<u>B.t.k. HD-73-kodesekvens</u>
pMON10806	En 35S	atslA	avkortet
pMON10814	En 35S	SSU11a	full lengde
pMON10811	SSU11a	SSU11a	avkortet
pMON10819	SSU11a	ingen	avkortet
pMON10815	atslA	ingen	avkortet
pMON10817	atslA	atslA	avkortet
pMON10821	En 35S	atslA	avkortet
pMON10822	En 35S	atslA	full lengde
pMON10838	SSU11a	SSU11a	full lengde
pMON10839	atslA	atslA	full lengde

• Alle de ovenfor nevnte vektorer ble brukt til å transformere tobakksplanter. For alle vektorene inneholdende avkortede B.t.k.-gener, er bladvev fra disse plantene blitt analysert med hensyn på toksisitet overfor insekter og B.t.k.-proteinnivåer ved immunologisk analyse. pMON10806, 10811, 10819 og 10821 produserer nivåer av B.t.k.-protein som er sammenlignbare med pMON5383 og pMON5390 som inneholder syntetiske B.t.k. HD-73-kodesekvenser drevet at selve En 35S-promoter uten noe CTP. Disse plantene har også den insekticide aktivitet som forventes for B.t.k.-proteinnivåene som er påvist. For pMON10815 og pMON10817 (inneholdende atslA-promoter) er nivået av B.t.k.-protein ca. 5 ganger høyere enn det som finnes i planter inneholdende pMON5383 eller 5390. Disse plantene har også høyere insekticid aktivitet. Planter inneholdende 10815 og 10817, inneholder opptil 1% av sitt totale oppløselige bladprotein som B.t.k. HD-73. Dette er det høyeste nivå av B.t.k.-protein som hittil er blitt oppnådd med noen av de syntetiske genene.

20 Dette resultatet er overraskende i to henseender. For det første, som angitt ovenfor, utviste ikke villtype-kodesekvensene fusjonert til atslA-promoter og CTP noe bevis på høyere nivåer av ekspresjon enn for En 35S, og hadde i virkeligheten lavere ekspresjon basert på fraværet av insekticide planter. For det andre viser Elionor et al. (1989) at for to andre gener kan atslA-CTP øke ekspresjon fra atslA-promoter med ca. 10 ganger. For det syntetiske B.t.k. HD-73-gen ses det ingen overensstemmende økning ved inkludering av CTP sammenlignet med og over det som ses for atslA-promoter alene.

30 Tobakksplanter som inneholder det syntetiske, fullengde HD-73 fusjonert til SSU11A-CTP og drevet av En 35S-promoter, ga nivåer av B.t.k.-protein og insekticid aktivitet som er sammenlignbare med pMON1518 hvis innhold ikke omfatter CTP. For pMON10518 ble i tillegg B.t.k.-proteinet ekstrahert fra planter observert ved gelelektroforese å inneholde flere former med mindre enn full lengde,

- åpenbart på grunn av spaltingen av C-endedelen (ikke påkrevet for toksisitet) i cytoplasmaet. For pMON10814 syntes mesteparten av proteinet å være intakt, fullengde, noe som indikerer at proteinet er blitt stabilisert mot proteolyse ved målretting mot kloroplasten.

Eksempel 11 -- Målretting av B.t.-proteiner til det ekstracellulære rom eller vakuoler gjennom bruken av signalpeptider

B.t.-proteinene fremstilt fra de syntetiske genene som her er beskrevet, er lokalisert til cytoplasmaet i plantecellen, og denne cytoplasmiske lokalisering resulterer i planter som er insekticid effektive. Det kan være fordelaktig for noen formål å styre B.t.-proteinene til andre deler av plantecellen. Lokalisering av B.t.-proteiner i andre deler enn cytoplasmaet kan resultere i mindre eksponering av B.t.-proteinene mot cytoplasmaproteaser, noe som fører til større akkumulering av det proteinet som gir forsterket insekticid aktivitet. Ekstracellulær lokalisering kunne føre til mer effektiv eksponering av disse insekter mot B.t.-proteinene, noe som igjen ville føre til større virkningsfullhet. Dersom et B.t.-protein ble funnet å være skadelig for plantecellefunksjon, da kunne lokalisering til en ikke-cytoplasmisk del beskytte disse cellene mot proteinet.

Både i planter og andre eukaryoter syntetiseres proteiner som er bestemt for å bli lokalisert enten ekstracellulært eller i flere spesifikke områder, vanligvis med en N-ende-aminosyreforlengelse som er kjent som signalpeptidet. Dette signalpeptidet styrer proteinet slik at det kommer inn i sporet for områdeoppdeling, og det avspaltes vanligvis fra det fullt ferdige protein som et tidlig trinn i oppdelingen i områder. For et ekstracellulært protein omfatter det sekretoriske spor vanligvis samtranslasjonell innføyelse i det endoplasmiske retikulum, idet avspalting av signalpeptidet inntreffer på dette stadium. Det fullt ferdige protein passerer så gjennom Golgi-legemet inn i

• vesikler som fusjonerer med plasmamembranen og derved frigjør proteinet til det ekstracellulære rom. Proteiner bestemt for andre områder, følger et lignende spor. F.eks. følger proteiner som er bestemt for det endoplasmiske retikulum eller Golgi-legemet, dette skjemaet, men de holdes spesifikt tilbake i det korrekte område. I planter er noen proteiner også målrettet mot vakuolen, et annet membranbundet område i cytoplasmaet til mange planteceller. Vakuolmålrettede proteiner fjerner seg fra det ovenfor nevnte spor ved Golgi-legemet hvor de kommer inn i vesikler som fusjonerer med vakuolen.

Et fellestrekk ved denne proteinmålretting er signalpeptidet som initierer områdeoppdelingsprosessen. Fusjonering av et signalpeptid til et protein vil i mange tilfeller føre til målrettingen av proteinet til det endoplasmiske retikulum. Effekten av dette trinnet kan også avhenge av sekvensen til selve det fullt ferdige protein. Signalene som styrer et protein mot et bestemt område istedenfor til det ekstracellulære rom, er ikke klart definert. Det synes som om mange av signalene som styrer proteinet til bestemte områder, finnes i aminosyresekvensen til det fullt ferdige protein. Dette er blitt påvist for noen vakuolmålrettede proteiner, men det er ennå ikke mulig å definere disse sekvensene nøyaktig. Det synes som om sekresjon til det ekstracellulære rom er "normal"-sporet for et protein som inneholder en signalsekvens, men ingen andre områdeoppdelingssignaler. En strategi for å styre B.t.-proteiner ut av cytoplasmaet er å fusjonere genene for syntetiske B.t.-gener til DNA-sekvenser som koder for kjente plantesignalpeptider. Disse fusjonsgenene vil gi opphav til B.t.-proteiner som kommer inn i det sekretoriske spor og fører til ekstracellulær sekresjon eller målretting til vakuolen eller andre områder.

Signalsekvenser for flere plantegener er blitt beskrevet. En slik sekvens er for det tobakkpatogeneserelaterte protein PR1b beskrevet av Cornelissen et al.: (1986) PR1b-proteinet er normalt lokalisert til det ekstracellulære rom.

• En annen type signalpeptid er inneholdt i frølagringsproteiner hos belgplanter. Disse proteinene er lokalisert til proteinmassen i frø som er et vakuollignende område som finnes i frø. En signalpeptid-DNA-sekvens for beta-under-

5 enheten i 7S-lagringsproteinet til vanlig bønne (*Phaseolus vulgaris*), PvuB, er blitt beskrevet av Doyle et al. Basert på disse publiserte sekvenser, ble gener syntetisert ved kjemisk syntese av oligonukleotider som kodet for signal-

10 peptidene for PRLb og PvuB. De syntetiske genene for disse signalpeptidene svarer nøyaktig til de rapporterte DNA-sekvenser. Like oppstrøms for translasjonsinitierings-

15 kodonet for hvert signalpeptid ble et BamHI- og BglII-sete innføyet med BamHI-setet i 5'-enden. Dette muliggjorde innføyelsen av de signalpeptid-kodende segmenter i BglII-

20 setet til pMON893 for ekspresjon fra En 35S-promoter. For å oppnå sekresjon av eller områdeoppdeling for heterologe proteiner, har det i noen tilfeller vist seg nødvendig å inkludere en eller annen aminosyresekvens etter det normale spaltingssete for signalpeptidet. Dette kan være nød-

25 vendig for å sikre korrekt spalting av signalpeptidet. For PRLb omfattet den syntetiske DNA-sekvens også de første 10 aminosyrene i fullt ferdig PRLb. For PvuB omfattet den syntetiske DNA-sekvens de første 13 aminosyrene i fullt ferdig PvuB. Begge de syntetiske signalpeptidkodende seg-

30 menter endte med NcoI-seter for å muliggjøre fusjon i ramme til det methionin-initierende kodon i de syntetiske B.t.-gener.

Det ble konstruert fire vektorer som koder for syntetiske B.t.k. HD-73-gener inneholdende disse signalpeptidene. Det syntetiske, avkortede HD-73-gen fra pMON5383

30 ble fusjonert med signalpeptidsekvensen til PvuB og inkorporert i pMON893 for å danne pMON10827. Det syntetiske, avkortede HD-73-gen fra pMON5383 ble også fusjonert til signal-

35 peptidsekvensen i PRLb for å danne pMON10824. Det syntetiske, fullengde HD-73-gen fra pMON10518 ble fusjonert til signalpeptidsekvensen i PvuB og inkorporert i pMON893 for å danne pMON10828. Det syntetiske, fullengde HD-73-gen fra

- pMON10518 ble også fusjonert til signalpeptidsekvensen i PR1b og inkorporert i pMON893 for å danne pMON10825.

Disse vektorene ble brukt til å transformere tobakksplanter, og plantene ble analysert med hensyn på ekspresjon av B.t.k.-proteinet ved "Western blot"-analyse, og med hensyn på insekticid virkningsfullhet. pMON10824 og pMON10827 produserte mengder av B.t.k.-protein i løv som er sammenlignbare med de avkortede HD-73-vektorene, pMON5383 og pMON5390. pMON10825 og pMON10828 ga fullengde B.t.k.-protein i sammenlignbare mengder med pMON10518. I alle tilfeller var plantene insekticid aktive mot tobakkhornlarver.

15

20

25

30

35

BIBLIOGRAFI

5 Adami, G. and Nevins, J. (1988) RNA Processing, Cold Spring Harbor Laboratory, p. 26.

Adang, et al., Molecular Strategies for Crop Protection (1987), pp. 345-353, Alan R. Liss, Inc.

10 Barton, K. A. et al., Plant Physiol. (1987), 85,1103-1109.

Bevan, M. et al., Nature (1983) 304:184.

Brady, H. and Wold, W. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 224.

15 Brown, John W., Nucleic Acids Research (1986) Vol. 14, No. 24, p. 9549.

Callis, J. Fromm, M. and Walbot, V., Genes and Develop. (1987), 1:1183-1200.

Conway, L. and Wickens, M. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 40.

20 Cornelissen, B.J.C., et al., EMBO J. (1986), Vol. 5, No. 1, 37-40.

Daar, I. O. et al. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 45.

25 Dean, C. et al., Nucleic Acids Research (1986), Vol. 14, No. 5, p. 2229.

Dedrick, R., et al., The Journal of Biological Chemistry (1987), Vol. 262, No. 19, pp. 9098-1106.

Donovan, W. P. et al., The J. of Biol. Chem. (1988), Vol. 263, No. 1, pp. 561-567.

30 Doyle, J.J., et al., J. Biol. Chem. (1986), Vol. 261, No. 20, 9228-9236.

Elionor, R.P., et al., Mol. Gen. Genet. (1989), 218:78-86.

- Fischhoff, D. A. et al., Bio/Technology (1987), Vol. 5, p. 807.
- 5 Fraley, R. T. et al., Bio/Technology (1985) 3:629-635.
- Fromm, M., Taylor, L. P. and Walbot, V., Nature (1986), 319:791-793.
- Gallego, M. E. and Nadal-Ginard B. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 61.
- 10 Genovese, C. and Milcarek, C. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 62.
- Gil, A. and Proudfoot, N. J., Nature (1984), Vol. 312, p. 473.
- Goodall, G. et al. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 63.
- 15 Gross, et al. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 128.
- Hampson, R. K. and Rottman, F. M. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 68.
- 20 Hanley, Brian A and Schuler, Mary A., Nucleic Acids Research (1988), Vol. 16, No. 14, p. 7159.
- Helfman, D. M. and Ricci, W. M. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 219.
- Herrera-Estrella, L. et al., Nature (1983), 303:209.
- 25 Hoekema, A. et al., Molecular and Cellular Biology (1987), Vol. 7, pp. 2914-2924.
- Honee, G. et al., Nucleic Acids Research (1988), Vol. 16, No. 13.
- Horsch, R. B. et al., Science (1985), 227:1229.
- 30 Jarret, R. L. et al., Physiol. Plant (1980), 49:177.

- Jarret, R. L. et al., In Vitro (1981), 17:825.
- Kay, R. et al., Science (1987), 236:1299-1302.
- 5 Kessler, M. et al. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 85.
- Klee, H. J. et al., Bio/Technology (1985), 3:637-642.
- Kozak, M., Nature (1984), 308:241-246.
- 10 Krebbers, E., et al., Plant Molecular Biology (1988), 11:745-759.
- Kunkel, T. A., Proc. Natl. Acad. Sci. USA (1985), Vol. 82, pp. 488-492.
- Marrone et al., J. Econ. Entomol. (1985), 78:290-293.
- 15 Marzluff, W. and Pandey, N. (1988), RNA Processing, Cold Spring Harbor Laboratory, p. 244.
- McCormick, S. et al., Plant Cell Reports (1986), 5:81-84.
- McDevitt, M. A. et al., Cell (1984), Vol. 37, pp. 993-999.
- 20 Murashige, T. and Skoog, F., Physiol. Plant (1962), 15:473.
- Odell, J. et al., Nature (1985), 313:810.
- Pandey, N. B. and Marzluff, W. F. (1987), RNA Processing, Cold Spring Harbor Laboratory, p. 133.
- 25 Proudfoot, N. J. et al. (1987), RNA Processing, Cold Spring Harbor Laboratory, p. 17.
- Reines, D., et al., J. Mol. Biol. (1987) 196:299-312.
- 30 Sadofsky, M. and Alwine, J. C., Molecular and Cellular Biology (1984), Vol. 4, No. 8, pp. 1460-1468.

- Sanders, P. R. et al., Nucleic Acids Research (1987), Vol. 15, No. 4, p. 1543.
- 5 Schuler, M. A. et al., Nucleic Acids Research (1982), Vol. 10, No. 24, pp. 8225-8244.
- Shaw, G. & Kamen, R., Cell (1986), 46:659-667.
- Shaw, G. and Kamen, R. (1987), RNA Processing, Cold Spring Harbor Laboratory, p. 220.
- 10 Trolinder, N. L. and Goodin, J. R., Plant Cell Reports (1987), 6:231-234.
- Tsurushita, N. and Korn, L. J. (1987), RNA Processing, Cold Spring Harbor Laboratory, p. 215.
- Tumer, N.E., et al., Nucleic Acids Res. (1986), Vol. 14:8, 3325.
- 15 Vaeck, M. et al., Nature (1987), Vol. 328, p. 33.
- Velten et al., EMBO J. (1984), 3:2723-2730.
- Velten & Schell, Nucleic Acids Research (1985), 13:6981-6998.
- 20 Visser, B. et al., Mol. Gen. Genet. (1988), 212:219-224.
- Webb, K. J. et al., Plant Sci. Letters (1983), 30:1.
- Wickens, M. and Stephenson, P., Science (1984), Vol. 226, p. 1045.
- 25 Wickens, M. et al. (1987), RNA Processing, Cold Spring Harbor Laboratory, p. 9.
- Wiebauer, K. et al., Molecular and Cellular Biology (1988), Vol. 8, No. 5, pp. 2042-2051.
- 30 Yamamoto, T. and Iizuka, T., Archives of Biochemistry and Biophysics (1983), Vol. 227, No. 1, pp. 233-241.
- 35

P a t e n t k r a v

1. Modifisert, kimært gen som inneholder en promoter som
5 virker i planteceller, operativt bundet til en strukturkodende
sekvens og et 3'-ikke-translatert område som inneholder et
polyadenyleringssignal som virker i planter slik at addisjonen
av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,
k a r a k t e r i s e r t v e d at den strukturkodende
10 sekvens koder for et toksinprotein avledet fra et *Bacillus*
thuringiensis-protein, hvor den strukturkodende sekvens har en
DNA-sekvens som atskiller seg fra den naturlig forekommende
DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
har de følgende kjennetegn:

- 15 a) hvilke som helst 15 nukleotider i den struktur-
kodende sekvens som inneholder mer enn fire på
hverandre følgende nukleotider bestående av en-
ten adenin- eller tyminrester, inneholder ikke
innenfor de 15 nukleotidene to polyadenylerings-
20 sekvenser,
- b) den strukturkodende sekvens inneholder ikke en
AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i den struktur-
kodende sekvens inneholder ikke innenfor de
25 15 nukleotidene mer enn én ATTTA-sekvens.

2. Modifisert, kimært gen som inneholder en promoter som
virker i planteceller, operativt bundet til en strukturkodende
sekvens og et 3'-ikke-translatert område som inneholder et
30 polyadenyleringssignal som virker i planter slik at addisjonen
av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,
k a r a k t e r i s e r t v e d at den strukturkodende
sekvens koder for et toksinprotein avledet fra et *Bacillus*
thuringiensis-protein, hvor den strukturkodende sekvens har en
35 DNA-sekvens som atskiller seg fra den naturlig forekommende
DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
har de følgende kjennetegn:

- a) den strukturkodende sekvens inneholder ikke noen

polyadenyleringssekvens innenfor 10 nukleotider i hvilken som helst annen polyadenyleringssekvens,

- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens.

3. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes, k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og har de følgende kjennetegn:

- a) den strukturkodende sekvens som inneholder ikke noen polyadenyleringssekvens innenfor 10 nukleotider i hvilken som helst annen polyadenyleringssekvens,
- b) hvilke som helst 30 nukleotider i den strukturkodende sekvens som inneholder: (1) mer enn ca. 80% adenin- og tyminrester, og (2) mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 30 nukleotidene noen polyadenyleringssekvens,
- c) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens; og
- d) hvilke som helst 30 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 30 nukleotidene mer enn én ATTTA-sekvens.

4. Modifisert, kimært gen som inneholder en promoter som

virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,

5 k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
10 har de følgende kjennetegn:

- a) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester,
15 inneholder ikke innenfor de 15 nukleotidene to polyadenyleringssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens innenfor 5'-halvdelen av den strukturkodende sekvens, og
- 20 c) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens.

25 5. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,
30 k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
35 har de følgende kjennetegn:

- a) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke noen polyadenyleringssekvens innenfor

10 nukleotider i hvilken som helst annen polyadenyleringssekvens,

- b) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens.

10

6. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylnukleotider til 3'-enden av RNA-en bevirkes, k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og har de følgende kjennetegn i 5'-halvdelen av den strukturkodende sekvens:

25

- a) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke noen polyadenyleringssekvens innenfor 10 nukleotider i noen annen polyadenyleringssekvens,
- b) hvilke som helst 30 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder:
- (1) mer enn ca. 80% adenin- og tyminrester; og
- (2) mer enn 4 på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 30 nukleotidene noen polyadenyleringssekvens,
- c) innenfor 5'-halvdelen av den strukturkodende sekvens inneholder den strukturkodende sekvens ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- d) hvilke som helst 30 nukleotider i 5'-halvdelen

30

35

av den strukturkodende sekvens inneholder ikke innenfor de 30 nukleotidene mer enn én ATTTA-sekvens.

- 5 7. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,
 10 k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
 15 har de følgende kjennetegn:

1) den strukturkodende sekvens har et område som er i det vesentlige homologt med den følgende sekvens:

TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC

20 1 5 10 15 20 25 30 35

hvor sekvensen ikke inneholder minst én av gruppen bestående av et plantepolyadenyleringssignal og en ATTTA-sekvens, eller

- 2) hvor den strukturkodende sekvens har et område som
 25 er i det vesentlige homologt med den følgende sekvens

TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC

1 5 10 15 20 25 30 35

- 30 med minst én modifikasjon valgt fra gruppen bestående av:

nukleotid 1 er cytosin (C),

nukleotid 3 er cytosin (C),

nukleotid 6 er cytosin (C),

nukleotid 12 er guanin (G),

35 nukleotid 18 er cytosin (C),

nukleotid 24 er guanin (G), og

nukleotid 36 er tymin (T), eller

3) hvor den naturlig forekommende DNA-sekvens omfat-

ter et område med den følgende sekvens

TTAATTAACCAAAGAATAGAAGAATTCGCTAGGAAC

1 5 10 15 20 25 30 35

5

og hvor den strukturkodende sekvens omfatter modifikasjoner slik at minst nevnte område inneholder minst én færre sekvens valgt fra gruppen bestående av en AACCAA- og en AATTAA-sekvens.

10

8. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylnukleotider til 3'-enden av RNA-en bevirkes, k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og har minst ett kjennetegn valgt fra gruppen bestående av:

- a) hvilke som helst 15 nukleotider i 5'-halvdelen av den strukturkodende sekvens som inneholder:
 - (1) mer enn ca. 80% adenin- og tyminrester, og
 - (2) mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke innenfor de 15 nukleotidene to plantepolyadenyleringssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de 15 nukleotidene mer enn én ATTTA-sekvens.

35

9. Modifisert, kimært gen som inneholder en promoter som virker i planteceller, operativt bundet til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen
5 av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes, k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende
10 DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og den strukturkodende sekvens inneholder ikke mer enn 5 på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester.

15 10. Modifisert, kimært gen ifølge krav 1, k a r a k t e r i s e r t v e d at polyadenyleringssignalene er valgt fra gruppen bestående av: AATAAA, AATAAT, AAC-CAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA og CATAAA.

20

11. Modifisert, kimært gen ifølge krav 1, k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et insektdrepende protein avledet fra *B.t.k.* HD-1, idet den strukturkodende sekvens har den følgende sek-
25 vens:

30

35

	1	ATGGCTATAGAAACTGGTTACACCCCAATCGATATTCCT	40
			
5	41	TGTCGCTAACGCAATTTCTTTTGAGTGAATTGTTCCCGG	80
			
	81	TGCTGGATTTGTGTTAGGACTAGTTGATATTATCTGGGGA	120
			
	121	ATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAA	160
10			
	161	TTGAACAGCTCATCAACCAGAGAATCGAAGAGTTCGCTAG	200
			
	201	GAATCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTT	240
			
15	241	TATCAAATTTACGCAGAACTCTTTTAGAGAGTGGGAAGCAG	280
			
	281	ATCCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCA	320
			
	321	ATTCAATGACATGAACAGTGCCCTTACAACCGCTATTCCT	360
20			
	361	CTTTTTTCAGTTCAAAATTATCAAGTTCCTCTCCTCTCCG	400
			
	401	TGTACGTTCAAGCTGCCAACCTCCACCTCTCAGTTTTGAG	440
			
25	441	AGATGTTTCAGTGTTTGGACAAAGGTGGGGATTTGATGCC	480
			
	481	GCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTA	520
			
	521	TTGGCAACTATACAGATCATGCTGTACGCTGGTACAATAC	560
30			
	561	GGGATTAGAGCGTGTATGGGGACCGGATTCTAGAGATTGG	600
			
	601	ATCAGGTACAACCAAGTTCAGAAGAGAGCTTACACTAACTG	640
35			

5
641 TATTAGATATCGTTTCTCTATTTCCGAACATATGATAGTAG 680
681 AACGTATCCAATTCGAACAGTTTCCCAATTAACAAGAGAA 720
721 ATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTT 760
761 TTCGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAG 800
10
801 TCCACATTTGATGGATATACTTAATAGTATAACCATCTAT 840
841 ACGGATGCTCATAGAGGAGAATACTACTGGTCCGGTCACC 880
15
881 AGATCATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATT 920
921 CACTTTTCCGCTATATGGAACATATGGGAAATGCAGCTCCA 960
961 CAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGATATA 1000
20
1001 GAACATTATCGTCCACCTTATATAGAAGACCTTTTAACAT 1040
1041 CGGGATCAACAACCAACAACATCTGTTCTTGACGGGACA 1080
25
1081 GAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTG 1120
1121 TATACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAAT 1160
1161 ACCGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTT 1200
30
1201 AGTCATCGATTAAGCCATGTTTCAATGTTTCGTTTCAGGCT 1240
1241 TTAGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTT 1280
35
1281 CTCTTGATACATCGTAGTGCTGAGTTCAACAACATCATC 1320

1321 CCTTCATCACAAATCACCCAAATCCCCTCACCAGTCTA 1360
 1361 CTAATCTTGGCTCTGGAAGTTCTGTCGTTAAAGGACCAGG 1400
 5 1401 ATTTACAGGAGGAGATATTCTTCGAAGAACTTCACCTGGC 1440
 1441 CAGATTTCAACCTTAAGAGTAAATATTACTGCACCATTAT 1480
 10 1481 CACAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCAC 1520
 1521 AAACCTTCAGTTCCACACATCAATTGACGGAAGACCTATT 1560
 1561 AATCAGGGGAATTTTTCAGCAACTATGAGTAGTGGGAGTA 1600
 15 1601 ATTTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTAC 1640
 1641 TCCGTTTAACTTTTCAAATGGATCAAGTGTATTTACGTTA 1680
 20 1681 AGTGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAG 1720
 1721 ATCGAATTGAATTTGTTCCGGCA 1743.

25 12. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdrepende protein avledet fra B.t.k.
 HD-73, idet den strukturkodende sekvens har den følgende sek-
 vens:

30

35

1 ATGGCCATTGAAACCGGTTACACTCCCATCGACATCTCCT 40
 41 TGTCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGG 80
 81 TGCTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGT 120
 121 ATCTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAA 160

161 TTGAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAG 200
201 GAACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTC 240
5 241 TACCAAATCTATGCAGAGAGCTTCAGAGAGTGGGAAGCCG 280
281 ATCCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCA 320
10 321 ATTCAACGACATGAACAGCGCCTTGACCACAGCTATCCCA 360
361 TTGTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCG 400
401 TGTACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCG 440
15 441 AGACGTTAGCGTGTTTGGGCAAAGGTGGGGATTTCGATGCT 480
481 GCAACCATCAATAGCCGTTACAACGACCTTACTAGGCTGA 520
20 521 TTGGAAACTACACCGACCACGCTGTTTCGTTGGTACAACAC 560
561 TGGCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGG 600
601 ATTAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAG 640
25 641 TTTTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAG 680
681 AACCTACCCTATCCGTACAGTGTCCTCAACTTACCAGAGAA 720
30 721 ATCTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCT 760
761 TCCGTGGTTCTGCCCAAGGTATCGAAGGCTCCATCAGGAG 800
801 CCCACACTTGATGGACATCTTGAACAGCATAACTATCTAC 840
35 841 ACCGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACC 880
881 AGATCATGGCCTCTCCAGTTGGATTTCAGCGGGCCCGAGTT 920

	921	TACCTTTCCTCTCTATGGA	960
	961	CAACAACGTATCGTTGCTCA	1000
5	1001	GAACCTTGTCTTCCACCTTG	1040
	1041	CGGTATCAACAACCAGCAAC	1080
10	1081	GAGTTCGCCTATGGAACCTC	1120
	1121	TTTACAGAAAGAGCGGAACC	1160
	1161	CCCACCACAGAACAACAATGT	1200
15	1201	TCCCACAGGTTGAGCCACGTG	1240
	1241	TCAGCAACAGTTCCGTGAGCAT	1280
20	1281	CTCTTGGATACACCGTAGTG	1320
	1321	GCATCCGATAGTATTACTCAA	1360
	1361	ACTTTCTCTTCAACGGTTCTG	1400
25	1401	CACTGGTGGAGACCTCGTTAG	1440
	1441	AACATTCAGAATAGAGGGTATA	1480
30	1481	TCCCATCCACATCTACCAGATA	1520
	1521	TGCTTCTGTGACCCCTATTCA	1560
	1561	AATTCATCCATCTTCTCCAATA	1600
35	1601	CCTCCTTGGATAATCTCCAATC	1640

1641 TGAAAGTGCCAATGCTTTTACATCTTCACTCGGTAACATC 1680
 1681 GTGGGTGTTAGAAACTTTAGTGGGACTGCAGGAGTGATTA 1720
 1721 TCGACAGATTCGAGTTCATTCCAGTTACTGCAACACTCGA 1760
 1761 GGCTGAG 1767.

13. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdepnde protein avledet fra *B.t.k.*
 HD-1, idet den strukturkodende sekvens har den følgende sek-
 vens:

1 ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA 40
 41 ACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGA 80
 81 ACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG 120
 121 TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG 160
 161 CTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT 200
 201 CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT 240
 241 GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA 280
 281 ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA 320
 321 CCAAATCTATGCAGAGAGCTTCAGAGAGTGGAAGCCGAT 360
 361 CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT 400
 401 TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT 440

	441	GTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCGTG	480
	481	TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG	520
5	521	ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTTCGATGCTGC	560
	561	AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT	600
10	601	GGAAACTACACCGACCACGCTGTTCGTTGGTACAACACTG	640
	641	GCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGGAT	680
	681	TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT	720
15	721	TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAA	760
	761	CCTACCCATCCGTACAGTGTCCCAACTTACCAGAGAAAT	800
20	801	CTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC	840
	841	CGTGGTTCTGCCCCAAGGTATCGAAGGCTCCATCAGGAGCC	880
	881	CACACTTGATGGACATCTTGAACAGCATAACTATCTACAC	920
25	921	CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG	960
	961	ATCATGGCCTCTCCAGTTGGATTTCAGCGGGCCCGAGTTTA	1000
30	1001	CCTTTCCTCTCTATGGAACCTATGGGAAACGCCGCTCCACA	1040
	1041	ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA	1080
	1081	ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCG	1120
35	1121	GTATCAACAACCAGCAACTTTCCTTCTTGACGGAACAGA	1160

1161 GTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT 1200
1201 TACAGAAAGAGCGGAACCGTTGATTCCTTGGACGAAATCC 1240
1241 CACCACAGAACAACAATGTGCCACCCAGGCAAGGATTCTC 1280
1281 CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTCC 1320
1321 AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTCT 1360
1361 CATGGATTCATCGTAGTGCTGAGTTCAACAATATCATTCC 1400
1401 TTCCTCTCAAATCACCCAAATCCCATTGACCAAGTCTACT 1440
1441 AACCTTGGATCTGGAACCTTCTGTCGTGAAAGGACCAGGCT 1480
1481 TCACAGGAGGTGATATTCTTAGAAGAACTTCTCCTGGCCA 1520
1521 GATTAGCACCCCTCAGAGTTAACATCACTGCACCACCTTTCT 1560
1561 CAAAGATATCGTGTGTCAGGATTCGTTACGCATCTACCACTA 1600
1601 ACTTGCAATTCCACACCTCCATCGACGGAAGGCCTATCAA 1640
1641 TCAGGGTAACTTCTCCGCAACCATGTCAAGCGGCAGCAAC 1680
1681 TTGCAATCCGGCAGCTTCAGAACCGTCGGTTTCACTACTC 1720
1721 CTTTCAACTTCTCTAACGGATCAAGCGTTTTACCCCTTAG 1760
1761 CGCTCATGTGTTCAATTCTGGCAATGAAGTGTACATTGAC 1800
1801 CGTATTGAGTTTGTGCCTGCCGAAGTTACCTTCGAGGCTG 1840
1341 AGTAC 1845.

14. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdrepende protein avledet fra *B.t.k.*
 HD-73, idet den strukturkodende sekvens har den følgende sek-
 5 vens:

	1	ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA	40
10	41	ACTGCTTGAGTAACCCAGAAGTTGAGTACTTGGTGGAGA	80
	81	ACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG	120
	121	TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG	160
15	161	CTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT	200
	201	CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT	240
20	241	GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA	280
	281	ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA	320
	321	CCAAATCTATGCAGAGAGCTTCAGAGAGTGGGAAGCCGAT	360
25	361	CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT	400
	401	TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT	440
30	441	GTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCGTG	480
	481	TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG	520
	521	ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTTCGATGCTGC	560
35	561	AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT	600

601 GGAAACTACACCGACCACGCTGTTTCGTTGGTACAACACTG 640
641 GCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGGAT 680
5 681 TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT 720
721 TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAA 760
10 761 CCTACCCTATCCGTACAGTGTCCCAACTTACCAGAGAAAT 800
801 CTATACTAACCCAGTTCCTTGAGAACTTCGACGGTAGCTTC 840
841 CGTGGTTCTGCCCAAGGTATCGAAGGCTCCATCAGGAGCC 880
15 881 CACACTTGATGGACATCTTGAACAGCATAACTATCTACAC 920
921 CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG 960
20 961 ATCATGGCCTCTCCAGTTGGATTTCAGCGGGCCCGAGTTTA 1000
1001 CCTTTCCTCTCTATGGAACCTATGGGAAACGCCGCTCCACA 1040
1041 ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA 1080
25 1081 ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCG 1120
1121 GTATCAACAACCAGCAACTTTCCGTTCTTGACGGAACAGA 1160
30 1161 GTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT 1200
1201 TACAGAAAGAGCGGAACCGTTGATTCCCTTGGACGAAATCC 1240
1241 CACCACAGAACAACAATGTGCCACCCAGGCAAGGATTCTC 1280
35 1281 CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTC 1320

1321 AGCAACAGTTC CGTGAGCATCATCAGAGCTCCTATGTTCT 1360
 5 1361 CTTGGATACACCGTAGTGCTGAGTTCAACAACATCATCGC 1400
 1401 ATCCGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC 1440
 10 1441 TTTCTCTTCAACGGTTCTGTCATTTCAAGGACCAGGATTCA 1480
 1481 CTGGTGGAGACCTCGTTAGACTCAACAGCAGTGGAAATAA 1520
 1521 CATTCAGAATAGAGGGTATATTGAAGTTCCAATTCAC TTC 1560
 15 1561 CCATCCACATCTACCAGATATAGAGTTCGTGTGAGGTATG 1600
 1601 CTTCTGTGACCCCTATTACCTCAACGTTAATTGGGGTAA 1640
 20 1641 TTCATCCATCTTCTCCAATACAGTTCCAGCTACAGCTACC 1680
 1681 TCCTTGGATAATCTCCAATCCAGCGATTTCCGGTTACTTTG 1720
 1721 AAAGTGCCAATGCTTTTACATCTTCACTCGGTAACATCGT 1760
 25 1761 GGGTGTTAGAAACTTTAGTGGGACTGCAGGAGTGATTATC 1800
 1801 GACAGATTCGAGTTCATTCCAGTTACTGCAACACTCGAGG 1840
 30 1841 CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTAATGCG 1880
 1881 CTGTTTACGTCTACAAACCAGCTTGGACTCAAGACAAATG 1920.

35

15. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdepnde protein avledet fra *B.t.k.*

HD-73, idet den strukturkodende sekvens har den følgende sekvens:

5	1	ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA	40
	41	ACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGA	80
	81	ACGCATTGAAACCGGTTAACTCCCATCGACATCTCCTTG	120
10	121	TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG	160
	161	CTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT	200
15	201	CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT	240
	241	GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA	280
	281	ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA	320
20	321	CCAAATCTATGCAGAGAGCTTCAGAGAGTGGAAGCCGAT	360
	361	CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT	400
25	401	TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT	440
	441	GTTTCGCAGTCCAGAACTACCAAGTTCCCTCTCTTGTCGGTG	480
	481	TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG	520
30	521	ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTTCGATGCTGC	560
	561	AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT	600
35	601	GGAAACTACACCGACCACGCTGTTTCGTTGGTACAACACTG	640
	641	GCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGGAT	680

581 TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT 720
721 TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAA 760
5 761 CCTACCCTATCCGTACAGTGTCCCAACTTACCAGAGAAAT 800
801 CTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC 840
10 841 CGTGGTTCTGCCCCAAGGTATCGAAGGCTCCATCAGGAGCC 880
881 CACACTTGATGGACA'TCTTGAACAGCATAACTATCTACAC 920
921 CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG 960
15 961 ATCATGGCCTCTCCAGTTGGATTCAGCGGGCCCGAGTTTA 1000
1001 CCTTTCCTCTCTATGGA ACTATGGGAAACGCCGCTCCACA 1040
20 1041 ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA 1080
1081 ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCG 1120
1121 GTATCAACAACCAGCAACTTTCGGTTCTTGACGGAACAGA 1160
25 1161 GTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT 1200
1201 TACAGAAAGAGCGGAACCGTTGATTCCTTGGACGAAATCC 1240
30 1241 CACCACAGAACAACAATGTGCCACCCAGGCAAGGATTCTC 1280
1281 CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTTC 1320
1321 AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTTCT 1360
35 1361 CTTGGATACACCGTAGTGCTGAGTTCAACAACATCATCGC 1400

1401 ATCCGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC 1440
1441 TTTCTCTTCAACGGTTCTGTCAATTCAGGACCAGGATTCA 1480
5 1481 CTGGTGGAGACCTCGTTAGACTCAACAGCAGTGGAAATAA 1520
1521 CATTTCAGAATAGAGGGTATATTGAAGTTCCAATTCAC TTC 1560
10 1561 CCATCCACATCTACCAGATATAGAGTTCGTGTGAGGTATG 1600
1601 CTTCTGTGACCCCTATTACCTCAACGTTAATTGGGGTAA 1640
1641 TTCATCCATCTTCTCCAATACAGTTCAGCTACAGCTACC 1680
15 1681 TCCTTGGATAATCTCCAATCCAGCGATTTTCGGTTACTTTG 1720
1721 AAAGTGCCAATGCTTTTACATCTTCACTCGGTAACATCGT 1760
20 1761 GGGTGT TAGAACTTTAGTGGGACTGCAGGAGTGATTATC 1800
1801 GACAGATTCGAGTTCATTCCAGTTACTGCAACACTCGAGG 1840
1841 CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC 1880
25 1881 GCTGTTTACGTCTACAAACCAGCTCGGCCTCAAGACCAAT 1920
1921 GTGACGGATTATCATATTGATCAAGTGTCCAACCTGGTGA 1960
30 1961 CCTACCTCAGCGATGAGTTCTGTCTGGATGAAAAGCGAGA 2000
2001 ATTGTCCGAGAAAGTCAAACATGCGAAGCGACTCAGTGAT 2040
2041 GAACGCAATTTACTCCAAGATTCAAATTTCAAAGACATTA 2080
35 2081 ATAGGCAACCAGAACGTGGGTGGGGCGGAAGTACAGGGAT 2120

	2121	TACCATCCAGGGAGGTTGACGACGTGTTCAAGGAGAACTAC	2160
	2161	GTCACACTATCAGGTACCTTTGATGAGTGCTATCCAACAT	2200
5	2201	ACCTCTACCAGAAGATCGACGAGTCCAAGTTGAAAGCCTT	2240
	2241	TACCCGTTATCAATTAAGAGGGTATATCGAAGATAGTCAA	2280
10	2281	GACCTCGAGATCTACCTCATCCGCTACAATGCAAAACATG	2320
	2321	AAACAGTAAATGTGCCAGGTACGGGTTCTTATGGCCGCT	2360
	2361	TTCAGCCCCAAAGTCCAATCGGAAAGTGTGGAGAGCCGAAT	2400
15	2401	CGATGCGCGCCACACCTTGAATGGAATCCTGACTTAGATT	2440
	2441	GTTCGTGTAGGGATGGAGAAAAGTGTGCCCATCATTCGCA	2480
20	2481	TCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTA	2520
	2521	AATGAGGACCTAGGTGTATGGGTGATCTTTAAGATTAAGA	2560
	2561	CGCAAGATGGGCACGCAAGACTAGGGAATCTAGAGTTTCT	2600
25	2601	CGAAGAGAAACCATTAGTAGGAGAAGCGCTAGCTCGTGTG	2640
	2641	AAAAGAGCGGAGAAAAAATGGAGAGACAAACGTGAGAAGT	2680
30	2681	TGGAATGGGAGACCAACATCGTCTACAAAGAGGCAAAGA	2720
	2721	ATCTGTAGATGCTTTATTTGTAAACTCTCAATATGATCAA	2760
	2761	TTACAAGCGGATACGAATATTGCCATGATTCATGCGGCAG	2800
35	2801	ATAAACGTGTTCATAGCATTCGAGAAGCTTATCTGCCTGA	2840

2841 GCTGTCTGTGATTCCGGGTCTCAATGCGGCTATTTTGTAA 2880
2881 GAATTAGAAGGGCGTATTTTCACTGCATTCTCCCTCTACG 2920
5 2921 ATGCCAGAAACGTCATCAAGAACGGTGAAGTCAACAATGG 2960
2961 CTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAA 3000
10 3001 GAACAAAACAACCAACGTTTCGGTCCTTGTTGTTCCGGAAT 3040
3041 GGAAGCAGAAGTGTACAAGAAGTTCGTGTCTGTCCGGG 3080
3081 TCGTGGCTATATCCTTCGTGTACAGCGTACAAGGAGGGA 3120
15 3121 TATGGAGAAGGTTGCGTAACCATTGATGAGATCGAGAACA 3160
3161 ATACAGACGAAGTGAAGTTTAGCAACTGCGTAGAAGAGGA 3200
20 3201 AATCTATCCAAATAACACGGTAACGTGTAATGATTATACT 3240
3241 GTAAATCAAGAAGAATACGGAGGTGCGTACACTTCTCGTA 3280
3281 ATCGAGGATATAACGAAGCTCCTTCCGTACCAGCTGATTA 3320
25 3321 TCGGTCAGTCTATGAAGAAAATCGTATACAGATGGACGA 3360
3361 AGAGAGAATCCTTGTAATTTAACAGAGGGTATAGGGATT 3400
30 3401 ACACGCCACTACCAGTTGGTTATGTGACAAAAGAATTAGA 3440
3441 ATACTTCCCAGAAACCGATAAGGTATGGATTGAGATTGGA 3480
3481 GAAACGGAAGGAACATTTATCGTGGACAGCGTGGAATTAC 3520
35 3521 TCCTTATGGAGGAA 3534.

16. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdrepende protein avledet fra *B.t.k.*
 HD-73, idet den strukturkodende sekvens har den følgende sek-
 5 vens:

	1	ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA	40
10	41	ACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGA	80
	81	ACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG	120
	121	TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG	160
15	161	CTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT	200
	201	CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT	240
20	241	GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA	280
	281	ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA	320
	321	CCAAATCTATGCAGAGAGCTTCAGAGAGTGGGAAGCCGAT	360
25	361	CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT	400
	401	TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT	440
30	441	GTTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCTGTG	480
	481	TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG	520
	521	ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTCGATGCTGC	560
35	561	AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT	600
	601	GGAAACTACACCGACCACGCTGTTCGTTGGTACAACACTG	640

641 GCTTGGAGCGTGTCTGGGGTCTGATTCTAGAGATTGGAT 680
681 TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT 720
721 TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAA 760
761 CCTACCCTATCCGTACAGTGTCCCAACTTACCAGAGAAAT 800
801 CTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC 840
841 CGTGGTTCTGCCCCAAGGTATCGAAGGCTCCATCAGGAGCC 880
881 CACACTTGATGGACATCTTGAACAGCATAACTATCTACAC 920
921 CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG 960
961 ATCATGGCCTCTCCAGTTGGATTTCAGCGGGCCCGAGTTTA 1000
1001 CCTTTCCTCTCTATGGAACATATGGGAAACGCCGCTCCACA 1040
1041 ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA 1080
1081 ACCTTGTCCTCCACCTTGTACAGAAGACCCTTCAATATCG 1120
1121 GTATCAACAACCAGCAACTTTCGGTTCTTGACGGAACAGA 1160
1161 GTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT 1200
1201 TACAGAAAGAGCGGAACCGTTGATTCCCTTGGACGAAATCC 1240
1241 CACCACAGAACAACAATGTGCCACCCAGGCAAGGATTCTC 1280
1281 CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTC 1320
1321 AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTCT 1360
1361 CTTGGATACACCGTAGTGCTGAGTTCAACAACATCATCGC 1400

1401 ATCCGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC 1440
1441 TTTCTCTTCAACGGTTCTGTCAATTCAGGACCAGGATTCA 1480
1481 CTGGTGGAGACCTCGTTAGACTCAACAGCAGTGGAAATAA 1520
1521 CATTCAGAATAGAGGGTATATTGAAGTTCCAATTCACTTC 1560
1561 CCATCCACATCTACCAGATATAGAGTTCGTGTGAGGTATG 1600
1601 CTTCTGTGACCCCTATTACCTCAACGTTAATTGGGGTAA 1640
1641 TTCATCCATCTTCTCCAATACAGTTCCAGCTACAGCTACC 1680
1681 TCCTTGGATAATCTCCAATCCAGCGATTCGGTTACTTTG 1720
1721 AAAGTGCCAATGCTTTTACATCTTCACTCGGTAACATCGT 1760
1761 GGGTGTTAGAAACTTTAGTGGGACTGCAGGAGTGATTATC 1800
1801 GACAGATTTCGAGTTCATTCCAGTTACTGCAACACTCGAGG 1840
1841 CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC 1880
1881 GCTGTTTACGTCTACAAACCAACTAGGGCTAAAAACAAAT 1920
1921 GTAACGGATTATCATATTGATCAAGTGTCCAATTTAGTTA 1960
1961 CGTATTTATCGGATGAATTTTGTCTGGATGAAAAGCGAGA 2000
2001 ATTGTC CGAGAAAGTCAAACATGCGAAGCGACTCAGTGAT 2040
2041 GAACGCAATTTACTCCAAGATTCAAATTTCAAAGACATTA 2080
2081 ATAGGCAACCAGAACGTGGGTGGGGCGGAAGTACAGGGAT 2120

2121 TACCATCCAAGGAGGGGATGACGTATTTAAAGAAAATTAC 2160
2161 GTCACACTATCAGGTACCTTTGATGAGTGCTATCCAACAT 2200
5 2201 ATTTGTATCAAAAAATCGATGAATCAAAATTAAAAGCCTT 2240
2241 TACCCGTTATCAATTAAGAGGGTATATCGAAGATAGTCAA 2280
10 2281 GACTTAGAAATCTATTTAATTGCTACAATGCAAAACATG 2320
2321 AAACAGTAAATGTGCCAGGTACGGGTTCTTATGGCCGCT 2360
2361 TTCAGCCCAAAGTCCAATCGGAAAGTGTGGAGAGCCGAAT 2400
15 2401 CGATGCGCGCCACACCTTGAATGGAATCCTGACTTAGATT 2440
2441 GTTCGTGTAGGGATGGAGAAAAGTGTGCCCATCATTCGCA 2480
20 2481 TCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTA 2520
2521 AATGAGGACCTAGGTGTATGGGTGATCTTTAAGATTAAGA 2560
2561 CGCAAGATGGGCACGCAAGACTAGGGAATCTAGAGTTTCT 2600
25 2601 CGAAGAGAAACCATTAGTAGGAGAAGCGCTAGCTCGTGTG 2640
2641 AAAAGAGCGGAGAAAAAATGGAGAGACAAACGTGAAAAAT 2680
30 2681 TGGAATGGGAAACAAATATCGTTTATAAGAGGCAAAGA 2720
2721 ATCTGTAGATGCTTTATTTGTAACTCTCAATATGATCAA 2760
2761 TTACAAGCGGATACGAATATTGCCATGATTCATGCGGCAG 2800
35 2801 ATAAACGTGTTTCATAGCATTCGAGAAGCTTATCTGCCTGA 2840

120

2841	GCTGTCTGTGATTCCGGGTGTCAATGCGGCTATTTTGTAA	2880
2881	GAATTAGAAGGGCGTATTTTCACTGCATTCTCCCTATATG	2920
2921	ATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATGG	2960
2961	CTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAA	3000
3001	GAACAAAACAACCAACGTTCCGGTCCTTGTTGTTCCGGAAT	3040
3041	GGGAAGCAGAAGTGTACACAAGAAGTTCGTGTCTGTCCGGG	3080
3081	TCGTGGCTATATCCTTCGTGTCACAGCGTACAAGGAGGGA	3120
3121	TATGGAGAAGGTTGCGTAACCATTTCATGAGATCGAGAACA	3160
3161	ATACAGACGAACTGAAGTTTAGCAACTGCGTAGAAGAGGA	3200
3201	AATCTATCCAAATAACACGGTAACGTGTAATGATTATACT	3240
3241	GTAAATCAAGAAGAATACGGAGGTGCGTACACTTCTCGTA	3280
3281	ATCGAGGATATAACGAAGCTCCTTCCGTACCAGCTGATTA	3320
3321	TGCGTCAGTCTATGAAGAAAAATCGTATACAGATGGACGA	3360
3361	AGAGAGAATCCTTGTGAATTTAACAGAGGGTATAGGGATT	3400
3401	ACACGCCACTACCAGTTGGTTATGTGACAAAAGAATTAGA	3440
3441	ATACTTCCCAGAAACCGATAAGGTATGGATTGAGATTGGA	3480
3481	GAAACGGAAGGAACATTTATCGTGACAGCGTGGAATTAC	3520
3521	TCCTTATGGAGGAA	3534

17. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et insektdepnde protein avledet fra *B.t.k.*
 HD-73, idet den strukturkodende sekvens har den følgende sek-
 5 vens:

```

      .       .       .       .
1   ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA   40

      .       .       .       .
10  41  ACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGA   80

      .       .       .       .
      81  ACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG   120

      .       .       .       .
      121 TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG   160

      .       .       .       .
15  161  CTGGGTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT   200

      .       .       .       .
      201 CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT   240

      .       .       .       .
20  241  GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA   280

      .       .       .       .
      281 ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA   320

      .       .       .       .
      321 CCAAATCTATGCAGAGAGCTTCAGAGAGTGGAAGCCGAT   360

      .       .       .       .
25  361  CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT   400

      .       .       .       .
      401 TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT   440

      .       .       .       .
30  441  GTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCCTG   480

      .       .       .       .
      481 TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG   520

      .       .       .       .
      521 ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTTCGATGCTGC   560

      .       .       .       .
35  561  AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT   600

      .       .       .       .
      601 GGAAACTACACCGACCACGCTGTTCGTTGGTACAACACTG   640

```

	641	GCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGGAT	680
5	681	TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT	720
	721	TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAA	760
	761	CCTACCCTATCCGTACAGTGTCCCAACTTACCAGAGAAAT	800
10	801	CTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC	840
	841	CGTGGTTCTGCCCCAAGGTATCGAAGGCTCCATCAGGAGCC	880
15	881	CACACTTGATGGACATCTTGAACAGCAATAACTATCTACAC	920
	921	CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAAG	960
	961	ATCATGGCCTCTCCAGTTGGATTACGCGGGCCCGAGTTTA	1000
20	1001	CCTTTCTCTCTATGGAACCTATGGGAAACGCCGCTCCACA	1040
	1041	ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA	1080
25	1081	ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCG	1120
	1121	GTATCAACAACCAGCAACTTTCCGTTCTTGACGGAACAGA	1160
	1161	GTTGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT	1200
30	1201	TACAGAAAGAGCGGAACCGTTGATTCTTGACGAAATCC	1240
	1241	CACCACAGAACAAATGTGCCACCCAGGCAAGGATTCTC	1280
35	1281	CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTC	1320

	1321	AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTCT	1360
5	1361	CTTGGATACACCGTAGTGCTGAGTTCAACAACATCATCGC	1400
	1401	ATCCGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC	1440
	1441	TTTCTCTTCAACGGTTCTGTCAATTCAGGACCAGGATTCA	1480
10	1481	CTGGTGGAGACCTCGTTAGACTCAACAGCAGTGGAAATAA	1520
	1521	CATTCAGAATAGAGGGTATATTGAAGTTCCAATTCACTTC	1560
15	1561	CCATCCACATCTACCAGATATAGAGTTCGTGTGAGGTATG	1600
	1601	CTTCTGTGACCCCTATTACCTCAACGTTAATTGGGGTAA	1640
	1641	TTTCATCCATCTTCTCCAATACAGTTCCAGCTACAGCTACC	1680
20	1681	TCCTTGCATAATCTCCAATCCAGCGATTTTCGGTTACTTTG	1720
	1721	AAAGTGCCAATGCTTTTACATCTTCACTCGGTAACATCGT	1760
25	1761	GGGTGTTAGAACTTTAGTGGGACTGCAGGAGTGATTATC	1800
	1801	GACAGATTCGAGTTCATTCCAGTTACTGCAACACTCGAGG	1840
	1841	CTGAGTACAACCTTGAGAGAGCCCAGAAGGCTGTGAACGC	1880
30	1881	CCTCTTTACCTCCACCAATCAGCTTGGCTTGAAAATAAC	1920
	1921	GTTACTGACTATCACATTGACCAAGTGTCCAACCTTGGTCA	1960
35	1961	CCTACCTTAGCGATGAGTTCTGCCTCGACGAGAAGCGTGA	2000

	2001	ACTCTCCGAGAAAGTTAAACACGCCAAGCGTCTCAGCGAC	2040
5	2041	GAGAGGAATCTCTTGCAAGACTCCAACTTCAAAGACATCA	2080
	2081	ACAGGCAGCCAGAACGTGGTTGGGGTGAAGCACCGGGAT	2120
	2121	CACCATCCAAGGAGGCGACGATGTGTTCAAGGAGAACTAC	2160
10	2161	GTCACCCTCTCCGGAACCTTCGACGAGTGCTACCCTACCT	2200
	2201	ACTTGTACCAGAAGATCGATGAGTCCAAACTCAAAGCCTT	2240
15	2241	CACCAGGTATCAACTTAGAGGCTACATCGAAGACAGCCAA	2280
	2281	GACCTTGAAATCTACTCGATCAGGTACAATGCCAAGCACG	2320
	2321	AGACCGTGAATGTCCCAAGGTACTGGTTCCCTCTGGCCACT	2360
20	2361	TTCTGCCCAATCTCCCATTTGGGAAGTGTGGAGAGCCTAAC	2400
	2401	AGATGCGCTCCACACCTTGAGTGGAATCCTGACTTGGACT	2440
25	2441	GCTCCTGCAGGGATGGCGAGAAGTGTGCCCAACCATTCTCA	2480
	2481	TCACTTCTCCTTGGACATCGATGTGGGATGTACTGACCTG	2520
	2521	AATGAGGACCTCGGAGTCTGGGTCATCTTCAAGATCAAGA	2560
30	2561	CCCAAGACGGACACGCAAGACTTGGCAACCTTGAGTTTCT	2600
	2601	CGAAGAGAAACCATTGGTCGGTGAAGCTCTCGCTCGTGTG	2640
35	2641	AAGAGAGCAGAGAAGAAGTGGAGGGACAAACGTGAGAAAC	2680

	2681	TCGAATGGGAAACTAACATCGTTTACAAGGAGGCCAAAGA	2720
5	2721	GTCCGTGGATGCTTTGTTCGTGAACTCCCAATATGATCAG	2760
	2761	TTGCAAGCCGACACCAACATCGCCATGATCCACGCCGCAG	2800
	2801	ACAAACGTGTGCACAGCATTCGTGAGGCTTACTTGCCTGA	2840
10	2841	GTTGTCCGTGATCCCTGGTGTGAACGCTGCCATCTTCGAG	2880
	2881	GAAGTTGAGGGACGTATCTTTACCGCATTCTCCTTGTACG	2920
15	2921	ATGCCAGAAACGTCATCAAGAACGGTGACTTCAACAATGG	2960
	2961	CCTCAGCTGCTGGAATGTGAAAGGTCATGTGGACGTGGAG	3000
	3001	GAACAGACAATCAGCGTTCGGTCCTGGTTGTGCCTGAGT	3040
20	3041	GGGAAGCTGAAGTGTCCCAAGAGGTTAGAGTCTGTCCAGG	3080
	3081	TAGAGGCTACATTCTCCGTGTGACCGCTTACAAGGAGGGA	3120
25	3121	TACGGTGAGGGTTGCGTGACCATCCACGAGATCGAGAACA	3160
	3161	ACACCGACGAGCTTAAGTTCTCCAAGTGCCTCGAGGAAGA	3200
	3201	AATCTATCCCAACAACACCGTTACTTGCAACGACTACACT	3240
30	3241	GTGAATCAGGAAGAGTACGGAGGTGCCTACACTAGCCGTA	3280
	3281	ACAGAGGTTACAACGAAGCTCCTTCCGTTCTGCTGACTA	3320
35	3321	TGCCTCCGTGTACGAGGAGAAATCCTACACAGATGGCAGA	3360

3361 CGTGAGAACCCTTGCGAGTTCAACAGAGGTTACAGGGACT 3400
 3401 ACACACCACTTCCAGTTGGCTATGTTACCAAGGAGCTTGA 3440
 5 3441 GTACTTTCCTGAGACCGACAAAGTGTGGATCGAGATCGGT 3480
 3481 GAAACCGAGGGAAACCTTCATCGTCGACAGCGTGGAGCTTC 3520
 10 3521 TCTTGATGGAGGAA 3534

18. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d a t den strukturkodende
 15 sekvens koder for et insektdrepende protein avledet fra *B.t.*
tenebrionis, idet den strukturkodende sekvens har den følgende
 sekvens:

20 1 ATGACTGCACACAAACAACACCGAAGCCCTCGACAGTTCTA 40
 41 CCACTAAGGATGTTATCCAGAAGGGTATCTCCGTTGTGGG 80
 81 AGACCTCTTGGGCGTGGTTGGATTTCCCTTCGGTGGAGCC 120
 25 121 CTCGTGAGCTTCTATACAAACTTTCTCAACACCATTGTC 160
 161 CAAGCGAGGACCCCTTGGAAGCATTATGGAGCAAGTTGA 200
 30 201 AGCTCTTATGGATCAGAAGATTGCAGATTATGCCAAGAAC 240
 241 AAGGCTTTGGCAGAACTCCAGGGCCTTCAGAACAAATGTGG 280
 281 AGGACTACGTGAGTGCATTGTCCAGCTGGCAGAAAGAACCC 320
 35 321 TGTTAGCTCCAGAAATCCTCACAGCCAAGGTAGGATCAGA 360

361 GAGTTGTTCTCTCAAGCCGAATCCCACCTTCAGAAATTCCA 400
.
401 TGCCTAGCTTTGCTATCTCCGGTTACGAGGTTCTTTTCCT 440
.
441 CACTACCTATGCTCAAGCTGCCAACACCCACTTGTTTCTC 480
.
481 CTTAAGGACGCTCAAATCTATGGAGAAGAGTGGGGATACG 520
.
521 AGAAAGAGGACATTGCTGAGTTCTACAAGCGTCAACTTAA 560
.
561 GCTCACCCAAGAGTACACTGACCATTGCGTGAAATGGTAT 600
.
601 AACGTTGGTCTCGATAAGCTCAGAGGCTCTTCCTACGAGT 640
.
641 CTTGGGTGAACTTCAACAGATACAGGAGAGAGATGACCTT 680
.
681 GACTGTGCTCGATCTTATCGCACTCTTCCCTTGTTACGAT 720
.
721 GTGAGACTCTACCCAAAGGAAGTGAAAAGTCTGAGCTTACCA 760
.
761 GAGACGTGCTCACTGACCCTATTGTCTGGAGTCAACAACCT 800
.
801 TAGGGGTATGGAAGTACCTTCAGCAATATCGAAAAGTAC 840
.
841 ATTAGGAAACCACATCTCTTCGACTATCTTCACAGAATTC 880
.
881 AATTCCACACAAGGTTTCAACCAGGATACTATGGTAACGA 920
.
921 CTCCTTCAACTATTGGTCCGGTAACTATGTTTCCACCAGA 960
.
961 CCAAGCATTGGATCTAATGACATCATCACATCTCCCTTCT 1000
.
1001 ATGGTAACAAGTCCAGTGAACCTGTGCAGAACCTTGAGTT 1040
.
.
.
.

	1041	CAACGGCGAGAAAGTCTATAGAGCCGTCGCAAACACCAAT	1080
5	1081	CTCGCTGTGTGGCCATCCGCAGTTTACTCAGGCGTCACAA	1120
	1121	AGGTGGAGTTTAGTCAGTATAACGATCAGACCGATGAGGC	1160
	1161	CAGCACCCAGACTTACGACTCCAAACGTAACGTTGGCGCA	1200
10	1201	GTCTCTTGGGATTCTATCGACCAATTGCCTCCAGAAACCA	1240
	1241	CAGACGAACCATTGGAGAAGGGCTACAGCCACCAACTTAA	1280
15	1281	CTATGTGATGTGCTTCTTGATGCAAGGTTCCAGAGGGACC	1320
	1321	ATTCCAGTGTTGACCTGGACACACAAGTCCGTGGACTTCT	1360
	1361	TCAACATGATCGATAGCAAGAAGATCACTCAACTTCCCTT	1400
20	1401	GGTGAAAGCCTACAAGCTGCAATCTGGTGCTCCGTTGTCT	1440
	1441	GCAGGTCCCAGATTCACTGGAGGTGACATCATCCAGTGCA	1480
25	1481	CAGAGAACGGCAGCGCAGCTACTATCTACGTGACACCTGA	1520
	1521	TGTGTCTTACTCTCAGAAGTACAGGGCACGTATTCATTAC	1560
	1561	GCATCTACCAGCCAGATCACCTTCACACTCAGCTTGGATG	1600
30	1601	GAGCACCCCTTCAACCAGTATTACTTTGACAAGACCATCAA	1640
	1641	CAAAGGTGACACTCTCACATACAATAGCTTCAACTTGGCA	1680
35	1681	AGTTTCAGCACACCATTTGAACTCTCAGGCAACAATCTTC	1720

1721 AGATCGGCGTCACCGGTCTCAGCGCCGGAGACAAAGTCTA 1760

1761 CATCGACAAGATTGAGTTCATCCCAGTGAAC 1791.

5

19. Modifisert, kimært gen ifølge krav 1,
k a r a k t e r i s e r t v e d at den strukturkodende
sekvens koder for et insektdrepende protein avledet fra *B.t.*
entomocidus, idet den strukturkodende sekvens har den følgende
10 sekvens:

15

20

25

30

35

1 ATGGAGGAGAACAACCAAAACCAATGCATTCCATACAAC 40
41 GCTTGAGTAACCCAGAAGAGGTATTGCTTGATGGAGAACG 80
81 CA'FTTCAACCGGTAACCTCTCCATCGACATCTCCTTGTCC 120
121 TTGGTCCAGTTTCTGGTCAGCAACTTCGTGCCAGGTGGTG 160
161 GGTTCCCTTGTCGGACTAATTGACTTCGTCTGGGGTATCGT 200
201 TGGTCCATCTCAATGGGATGCATTCCCTGGTGCAAATTGAG 240
241 CAGTTGATCAACGAGAGGATCGCTGAGTTCGCCAGGAACG 280
281 CTGCCATCGCTAACTTGAAGGATTGGGCAATAACTTCAA 320
321 CATCTATGTGGAGGCCTTCAAAGAGTGGGAAGAGGACCCT 360
361 AACAACCCAGAGACCCGCACTAGGGTGATCGACAGATTCA 400
401 GAATCTTGGACGGCCTCTTGGAGAGAGATATCCCATCCTT 440
441 CAGAATCTCTGGCTTCGAAGTTCCTCTCTTGTCCGTGTAC 480
481 GCTCAAGCAGCTAATCTTCACCTCGCTATCCTTCGAGACA 520

521 GTGTCATCTTTGGGGAAAGGTGGGGATTGACCACTATCAA 560
561 CGTCAATGAGAATTACAACAGACTTATCAGGCACATTGAC 600
601 GAGTACGCCGACCACTGTGCTAACACCTACAACCGTGGCT 640
641 TGAACAATCTCCCTAAGTCTACTTATCAAGATTGGATTAC 680
681 CTACAACAGGTTGAGGAGAGACTTGACCCTCACAGTTTTG 720
721 GACATTGCAGCTTTCTTCCCGAACTATGACAACAGGAGAT 760
761 ACCCTATCCAACCAGTGGGTCAACTTACCAGAGAAGTCTA 800
801 TACTGACCCACTTATCAACTTCAACCCTCAGTTGCAAAGT 840
841 GTCGCCCAACTTCCCAATTCAACGTCATGGAGTCCAGCC 880
881 GTATCAGGAACCCACACTTGTTTGACATCTTGAACAACCT 920
921 TACTATCTTCACCGATTGGTTCAGCGTTGGGCGTAACTTC 960
961 TATTGGGGTGGACACAGGGTCATCTCCTCTCTTATTGGAG 1000
1001 GTGGGAACATTACCTCTCCTATCTATGGACGTGAGGCAAA 1040
1041 CCAGGAGCCACCACGTAGTTTCACCTTCAACGGTCCAGTC 1080
1081 TTCAGAACCTTGTCTAACCCCTACCTTGAGATTGCTCCAGC 1120
1121 AACCTTGGCCAGCTCCACCTTTCAACCTTAGAGGTGTTGA 1160
1161 GGGCGTTGAGTTCTCTACTCCTACCAACTCCTTCACTTAC 1200

	1201	AGAGGTAGAGGAACCGTTGATTCCTTGACCGAACTCCCAC	1240
5	1241	CAGAGGACAATAGCGTGCCACCCAGGGAAGGCTACTCCCA	1280
	1281	CAGGTTGTGCCACGCAACCTTCGTGCAGCGTCCGGA	1320
	1321	CCATTCCTCACTACAGGAGTTGTGTTCTCATGGACTGATC	1360
10	1361	GTAGTGCTACTCTCACTAATACCATTGATCCCGAGAGGAT	1400
	1401	CAATCAAATCCCATTGGTCAAGGGTTCCGTGTGTGGGGA	1440
15	1441	GGAAGTTCTGTCATCACAGGACCAGGCTTCACAGGAGGTG	1480
	1481	ATATCTCTTAGAAGAAACACTTTTGGCGACTTTGTGAGCCT	1520
	1521	CCAAGTTAACATCAACTCTCCAATTACTCAAAGATATCGT	1560
20	1561	CTCAGGTTTCGTTACGCATCTTCCCGTGACGCTAGAGTCA	1600
	1601	TCGTGCTCACCGGAGCAGCTTCTACCGGTGTCGGTGGACA	1640
25	1641	AGTCTCCGTGAACATGCCACTCCAGAAGACTATGGAGATC	1680
	1681	GGCGAGAACTTGACATCCAGGACCTTCAGATACACCGACT	1720
	1721	TCTCTAACCCTTTCAGTTTCGCGTCCAACCCTGACATCAT	1760
30	1761	TGGCATTAGCGAACAACCTCTCTTTGGAGCTGGTAGCATC	1800
	1801	TCATCTGGCGAATTGTACATTGACAAGATTGAGATCATT	1840
35	1841	TTGCCGACGCTACCTTCGAGGCTGAGTCTGACCTTGAGAG	1880

	1881	AGCCCAGAAGGCTGTGAACGCCCTCTTTACCTCCTCTAAT	1920
5	1921	CAGATTGGCTTGAAAAGTACGTTACTGACTATCACATTG	1960
	1961	ACCAAGTGTCCAAGTTGGTCGACTGCCTTAGCGATGAGTT	2000
	2001	CTGCCTCGACGAGAAGCGTGAACTCTCCGAGAAAGTTAAA	2040
10	2041	CACGCCAAGCGTCTCAGCGACGAGAGGAATCTCTTGCAAG	2080
	2081	ACCCCAACTTCAGAGGCATCAACAGGCAGCCAGACCGTGG	2120
15	2121	TTGGAGAGGAAGCACCGACATCACCATCCAAGGAGGCGAC	2160
	2161	GATGTGTTCAAGGAGAACTACGTCACCCTCCCAGGAACTG	2200
	2201	TGGACGAGTGCTACCCTACCTACTTGTACCAGAAGATCGA	2240
20	2241	TGAGTCCAAACTCAAAGCCTACACCAGGTATGAACTTAGA	2280
	2281	GGCTACATCGAAGACAGCCAAGACCTTGAAATCTACCTCA	2320
25	2321	TCAGGTACAATGCCAAGCACGAGATCGTGAATGTCCCAGG	2360
	2361	TACTGGTTCCCTCTGGCCACTTTCTGCCCAAATGCCCATT	2400
	2401	GGGAAGTGTGGAGAGCCTAACAGATGCGCTCCACACCTTG	2440
30	2441	AGTGGAATCCTGACTTGGACTGCTCCTGCAGGGATGGCGA	2480
	2481	GAAGTGTGCCCACCATTTCTCATCACTTCACCTTGGACATC	2520
35	2521	GATGTGGGATGTACTGACCTGAATGAGGACCTCGGAGTCT	2560

2561 GGGTCATCTTCAAGATCAAGACCCAAGACGGACACGCAAG 2600
5 2601 ACTTGGCAACCTTGAGTTTCTCGAAGAGAAACCATTGCTC 2640
2641 GGTGAAGCTCTCGCTCGTGTGAAGAGAGCAGAGAAGAAGT 2680
2681 GGAGGGACAAACGTGAGAACTCCAACTCGAGACTAACAT 2720
10 2721 CGTTTACAAGGAGGCCAAAGAGTCCGTGGATGCTTTGTTC 2760
2761 GTGAACTCCCAATATGATAGGTTGCAAGTGGACACCAACA 2800
15 2801 TCGCCATGATCCACGCTGCAGACAAACGTGTGCACAGGAT 2840
2841 TCGTGAGGCTTACTTGCCTGAGTTGTCCGTGATCCTGGT 2880
2881 GTGAACGC1GCCATCTTCGAG3AACTTGAGGGACGTATCT 2920
20 2921 TTACCGCATACTCCTTGACGATGCCAGAAACGTCATCAA 2960
2961 GAACGGTGACTTCAACAATGGCCTCTTGTGCTGGAATGTG 3000
25 3001 AAAGGTCATGTGGACGTGGAGGAACAGAACAATCACCGTT 3040
3041 CCGTCCTGGTTATCCCTGAGTGGGAAGCTGAAGTGTCCCA 3080
3081 AGAGGTTAGAGTCTGTCCAGGTAGAGGCTACATTCTCCGT 3120
30 3121 GTGACCGCTTACAAGGAGGGATACGGTGAGGGTTGCGTGA 3160
3161 CCATCCACGAGATCGAGGACAACACCGACGAGCTTAAGTT 3200
35 3201 CTCCAAC1GCGTCGAGGAAGAAGTCTATCCCAACAACACC 3240

3241 GTTACTTGCAACAACACTACACTGGGACCCAGGAAGAGTACG 3280
 3281 AAGGTACCTACACTAGCCGTAACCAAGGTTACGACGAAGC 3320
 5 3321 TTACGGAAACAATCCTTCCGTTCCCTGCTGACTATGCCTCC 3360
 3361 GTGTACGAGGAGAAATCCTACACAGATGGCAGACGTGAGA 3400
 10 3401 ACCCTTGCGAGTCCAACAGAGGTTACGGTGACTACACACC 3440
 3441 ACTTCCAGCAGGCTATGTTACCAAGGACCTTGAGTACTTT 3480
 3481 CCTGAGACCGACAAAGTGTGGATCGAGATCGGTGAAACCG 3520
 15 3521 AGGGAACCTTCATCGTGGACAGCGTGGAGCTTCTCTTGAT 3560
 3561 GGAGGAA 3567.

20

20. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et B.t. P2 insektdrepende protein, idet den
 strukturkodende sekvens har den følgende sekvens:

25

1 ATGGACAACAACGTCTTGAACCTCTGGTAGAACAACCATCT 40
 41 GCGACGCATACAAACGTCGTGGCTCACGATCCATTTCAGCTT 80
 30 81 CGAACACAAGAGCCTCGACACTATTTCAGAAGGAGTGGATG 120
 121 GAATGGAAACGTACTGACCACTCTCTCTACGTCGCACCTG 160
 35 161 TGGTTGGAACAGTGTCCAGCTTCCTTCTCAAGAAGGTCGG 200
 201 CTCTCTCATCGGAAAACGTATCTTGTCCGAACCTCTGGGGT 240

241 ATCATCTTTCCATCTGGGTCCACTAATCTCATGCAAGACA 280
5
281 TCTTGAGGGAGACCGAACAGTTTCTCAACCAGCGTCTCAA 320
321 CACTGATACCTTGGCTAGAGTCAACGCTGAGTTGATCGGT 360
361 CTCCAAGCAAACATTTCGTGAGTTCAACCAGCAAGTGGACA 400
10
401 ACTTCTTGAATCCAACTCAGAATCCTGTGCCTCTTTCCAT 440
441 CACTTCTTCCGTGAACACTATGCAGCAACTCTTCCTCAAC 480
15
481 AGATTGCCTCAGTTTCAGATTCAAGGCTACCAGTTGCTCC 520
521 TTCTTCCACTCTTTGCTCAGGCTGCCAACATGCACTTGTC 560
561 CTTCCATACGTGACGTGATCCTCAACGCTGACGAATGGGGA 600
20
601 ATCTCTGCAGCCACTCTTAGGACATACAGAGACTACTTGA 640
641 GGAACTACACTCGTGATTACTCCAACATATTGCATCAACAC 680
25
681 TTATCAGACTGCCTTTTCGTGGACTCAATACTAGGCTTCAC 720
721 GACATGCTTGAGTTCAGGACCTACAATGTTTCCTTAACGTGT 760
761 TTGAGTACGTCAGCATTTGGAGTCTCTTCAAGTACCAGAG 800
30
801 CTTGATGGTGTCTCTGAGCCAATCTCTACGCCTCTGGC 840
841 AGTGGACCACAGCAAACCTCAGAGCTTCACAGCTCAGAACT 880
35
881 GGCCATTCTTGTATAGCTTGTTCCAAGTCAACTCCAACATA 920

	921	CATTCTCAGTGGTATCTCTGGGACCAGACTCTCCATAACC	960
5	961	TTTCCCAACATTGGTGGACTTCCAGGCTCCACTACAACCC	1000
	1001	ATAGCCTTAACCTCTGCCAGAGTGAACCTACAGTGGAGGTGT	1040
	1041	CAGCTCTGGATTGATTGGTGCAACTAACTTGAACCACAAC	1080
10	1081	TTCAATTGCTCCACCGTCTTGCCACCTCTGAGCACACCGT	1120
	1121	TTGTGAGGTCCTGGCTTGACAGCGGTACTGATCGCGAAGG	1160
15	1161	AGTTGCTACCTCTACAAACTGGCAAACCGAGTCCTTCCAA	1200
	1201	ACCACTCTTAGCCCTCGGTGTGGAGCTTCTCTGCACGTG	1240
	1241	GGAATTCAAACCTACTTTCCAGACTACTTCATTAGGAACAT	1280
20	1281	CTCTGGTGTTCCTCTCGTCATCAGGAATGAAGACCTCACC	1320
	1321	CGTCCACTTCATTACAACCAGATTAGGAACATCGAGTCTC	1360
25	1361	CATCCGGTACTCCAGGAGGTGCAAGAGCTTACCTCGTGTC	1400
	1401	TGTCCATAACAGGAAGAACAACATCTACGCTGCCAACGAG	1440
	1441	AATGGCACCATGATTACCTTGCACCAGAAGATTACACTG	1480
30	1481	GATTCACCATCTCTCCAATCCATGCTACCCAAGTGAACAA	1520
	1521	TCAGACACGCACCTTCATCTCCGAAAAGTTCGGAAATCAA	1560
35	1561	GGTGACTCCTTGAGGTTCGAGCAATCCAACTACTACCGCTA	1600

1601 GGTACACTTTGAGAGGCAATGGAAACAGCTACAACCTTTA 1640
 1641 CTTGAGAGTTAGCTCCATTGGTAACTCCACCATCCGTGTT 1680
 1681 ACCATCAACGGACGTGTTTACACAGTCTCTAATGTGAACA 1720
 1721 CTACAACGAACAATGATGGCGTTAACGACAACGGAGCCAG 1760
 1761 ATTCAGCGACATCAACATTGGCAACATCGTGGCCTCTGAC 1800
 1801 AACACTAACGTTACTTTGGACATCAATGTGACCCTCAATT 1840
 1841 CTGGA ACTCCATTGATCTCATGAACATCATGTTTGTGCC 1880
 1881 AACTAACCTCCCTCCATTGTACTAA 1905.

21. Modifisert, kimær gensekvens ifølge krav 1,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for et fusjonsprotein som omfatter de N-termin-
 ale 610 aminosyrene i et toksinprotein avledet fra B.t.k. HD-1
 og de C-terminale 567 aminosyrene i et toksinprotein avledet
 fra B.t.k. HD-73, hvor den strukturkodende sekvens har den
 følgende sekvens:

1 ATGGACAACAACCCAAACATCAACGAATGCATTCCATACA 40
 41 ACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGA 80
 81 ACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG 120
 121 TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTG 160
 161 CTGGGTTTCGTTCTCGGACTAGTTGACATCATCTGGGGTAT 200
 201 CTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT 240

241 GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGA 280
281 ACCAGGCCATCTCTAGGTTGGAAGGATTGAGCAATCTCTA 320
5 321 CCAAATCTATGCAGAGAGCTTCAGAGAGTGGGAAGCCGAT 360
361 CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAAT 400
10 401 TCAACGACATGAACAGCGCCTTGACCACAGCTATCCCATT 440
441 GTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCTGTG 480
481 TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAG 520
15 521 ACGTTAGCGTGTTTGGGCAAAGGTGGGGATTCGATGCTGC 560
561 AACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT 600
20 601 GGAAACTACACCGACCACGCTGTTCGTTGGTACAACACTG 640
641 GCTTGGAGCGTGTCTGGGGTCCTGATTCTAGAGATTGGAT 680
681 TAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT 720
25 721 TTGGACATTGTGTCTCTCTTCCGAACTATGACTCCAGAA 760
761 CCTACCCTATCCGTACAGTGTCCCAACTTACCAGAGAAAT 800
30 801 CTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC 840
841 CGTGGTTCTGCCCCAAGGTATCGAAGGCTCCATCAGGAGCC 880
881 CACACTTGATGGACATCTTGAACAGCATAACTATCTACAC 920
35 921 CGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG 960

5
10
15
20
25
30
35

961 ATCATGGCCTCTCCAGTTGGATTTCAGCGGGCCCGAGTTTA 1000
1001 CCTTTCCCTCTCTATGGAACATATGGGAAACGCCGCTCCACA 1040
1041 ACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA 1080
1081 ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCG 1120
1121 GTATCAACAACCAGCAACTTTCCGTTCTTGACGGAACAGA 1160
1161 GTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT 1200
1201 TACAGAAAGAGCGGAACCGTTGATTTCCTTGGACGAAATCC 1240
1241 CACCACAGAACAACAATGTGCCACCCAGGCAAGGATTCTC 1280
1281 CCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTTC 1320
1321 AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTCT 1360
1361 CATGGATTTCATCGTAGTGCTGAGTTCAACAATATCATTCC 1400
1401 TTCCTCTCAAATCACCCAAATCCCATTGACCAAGTCTACT 1440
1441 AACCTTGGATCTGGAACCTTCTGTCGTGAAAGGACCAGGCT 1480
1481 TCACAGGAGGTGATATTCTTAGAAGAACTTCTCCTGGCCA 1520
1521 GATTAGCACCCCTCAGAGTTAACATCACTGCACCACTTTCT 1560
1561 CAAAGATATCGTGTCAGGATTCGTTACGCATCTACCACTA 1600
1601 ACTTGCAATTCCACACCTCCATCGACGGAAGGCCTATCAA 1640
1641 TCAGGGTAACTTCTCCGCAACCATGTCAAGCGGCAGCAAC 1680

	1681	TTGCAATCCGGCAGCTTCAGAACCGTCGGTTTCACTACTC	1720
	1721	CTTTCAACTTCTCTAACGGATCAAGCGTTTTTCACCCTTAG	1760
5	1761	CGCTCATGTGTTC AATTCTGGCAATGAAGTGTACATTGAC	1800
	1801	CGTATTGAGTTTGTGCCTGCCGAAGTTACCCTCGAGGCTG	1840
10	1841	AGTACAACCTTGAGAGAGCCCAGAAGGCTGTGAACGCCCT	1880
	1881	CTTTACCTCCACCAATCAGCTTGGCTTGAAAAC TAACGTT	1920
	1921	ACTGACTATCACATTGACCAAGTGTCCA ACTTGGTCACCT	1960
15	1961	ACCTTAGCGATGAGTTCTGCCTCGACGAGAAGCGTGA ACT	2000
	2001	CTCCGAGAAAGTTAAACACGCCAAGCGTCTCAGCGACGAG	2040
20	2041	AGGAATCTCTTGCAAGACTCCA ACTTCAAAGACATCAACA	2080
	2081	GGCAGCCAGAACGTGGTTGGGGTGAAGCACCGGGATCAC	2120
	2121	CATCCAAGGAGGCGACGATGTGTTCAAGGAGAACTACGTC	2160
25	2161	ACCCTCTCCGGA ACTTTTCGACGAGTGCTACCCTACCTACT	2200
	2201	TGTACCAGAAGATCGATGAGTCCAA ACTCAAAGCCTTCAC	2240
30	2241	CAGGTATCA ACTTAGAGGCTACATCGAAGACAGCCAAGAC	2280
	2281	CTTGAAATCTACTCGATCAGGTACAATGCCAAGCACGAGA	2320
	2321	CCGTGAATGTCCCAGGTACTGGTTCCCTCTGGCCACTTTC	2360
35	2361	TGCCCAATCTCCCAT TGGGAAGTGTGGAGAGCCTAACAGA	2400

	2401	TGCGCTCCACACCTTGAGTGGAATCCTGACTTGGACTGCT	2440
	2441	CCTGCAGGGATGGCGAGAAGTGTGCCCACCATTCTCATCA	2480
5	2481	CTTCTCCTTGGACATCGATGTGGGATGTACTGACCTGAAT	2520
	2521	GAGGACCTCGGAGTCTGGGTCATCTTCAAGATCAAGACCC	2560
10	2561	AAGACGGACACGCAAGACTTGGCAACCTTGAGTTTCTCGA	2600
	2601	AGAGAAACCATTTGGTCGGTGAAGCTCTCGCTCGTGTGAAG	2640
	2641	AGAGCAGAGAAGAAGTGGAGGGACAAACGTGAGAACTCG	2680
15	2681	AATGGGAAACTAACATCGTTTACAAGGAGGCCAAAGAGTC	2720
	2721	CGTGGATGCTTTGTTCGTGAACTCCCAATATGATCAGTTG	2760
20	2761	CAAGCCGACACCAACATCGCCATGATCCACGCCGCAGACA	2800
	2801	AACGTGTGCACAGCATTCGTGAGGCTTACTTGCCTGAGTT	2840
	2841	GTCCGTGATCCCTGGTGTGAACGCTGCCATCTTCGAGGAA	2880
25	2881	CTTGAGGGACGTATCTTTACCGCATTCTCCTTGACGATG	2920
	2921	CCAGAAACGTCATCAAGAACGGTGACTTCAACAATGSCCT	2960
30	2961	CAGCTGCTGGAATGTGAAAGGTCATGTGGACGTGGAGGAA	3000
	3001	CAGAACAATCAGCGTTCCGTCCTGGTTGTGCCTGAGTGGG	3040
	3041	AAGCTGAAGTGTCCCAAGAGGTTAGAGTCTGTCCAGGTAG	3080
35	3081	AGGCTACATTCTCCGTGTGACCGCTTACAAGGAGGGATAC	3120
	3121	GGTGAGGGTTGCGTGACCATCCACGAGATCGAGAACAACA	3160

3161 CCGACGAGCTTAAGTTCTCCAACCTGCGTCGAGGAAGAAAT 3200
 3201 CTATCCCAACAACACCGTTACTTGCAACGACTACACTGTG 3240
 5 3241 AATCAGGAAGAGTACGGAGGTGCCTACACTAGCCGTAACA 3280
 3281 GAGGTTACAACGAAGCTCCTTCCGTTCTGCTGACTATGC 3320
 10 3321 CTCCGTGTACGAGGAGAAATCCTACACAGATGGCAGACGT 3360
 3361 GAGAACCCTTGCGAGTTCAACAGAGGTTACAGGGACTACA 3400
 3401 CACCACTTCCAGTTGGCTATGTTACCAAGGAGCTTGAGTA 3440
 15 3441 CTTTCCTGAGACCGACAAAGTGTGGATCGAGATCGGTGAA 3480
 3481 ACCGAGGGAACCTTCATCGTGGACAGCGTGGAGCTTCTCT 3520
 20 3521 TGATGGAGGAA 3531.

22. Modifisert, kimært gen ifølge krav 1,
 k a r a k t e r i s e r t v e d at nevnte *Bacillus thurin-*
 25 *giensis* er valgt fra gruppen bestående av *B.t. kurstaki* HD-1,
B.t. kurstaki HD-73, *B.t. sotto*, *B.t. berliner*, *B.t. thurin-*
giensis, *B.t. tolworthi*, *B.t. dendrolimus*, *B.t. alesti*, *B.t.*
galleriae, *B.t. aizawai*, *B.t. subtoxicus*, *B.t. entomocidus* og
B.t. tenebrionis.

30

23. Modifisert, kimært gen ifølge krav 7,
 k a r a k t e r i s e r t v e d at den strukturkodende
 sekvens koder for minst 5'-endehalvdelen av det insektdrepende
 protein fra *B.t. kurstaki* HD-1.

35

24. Plantetransformasjonsvektor som inneholder et
 plantegen som inneholder et modifisert, kimært gen innehold-
 ende en promoter som virker i planteceller, operativt bundet

til en strukturkodende sekvens og et 3'-ikke-translatert område som inneholder et polyadenyleringssignal som virker i planter slik at addisjonen av polyadenylatnukleotider til 3'-enden av RNA-en bevirkes,

5 k a r a k t e r i s e r t v e d at den strukturkodende sekvens koder for et toksinprotein avledet fra et *Bacillus thuringiensis*-protein, hvor den strukturkodende sekvens har en DNA-sekvens som atskiller seg fra den naturlig forekommende DNA-sekvens som koder for *Bacillus thuringiensis*-proteinet, og
10 har de følgende kjennetegn:

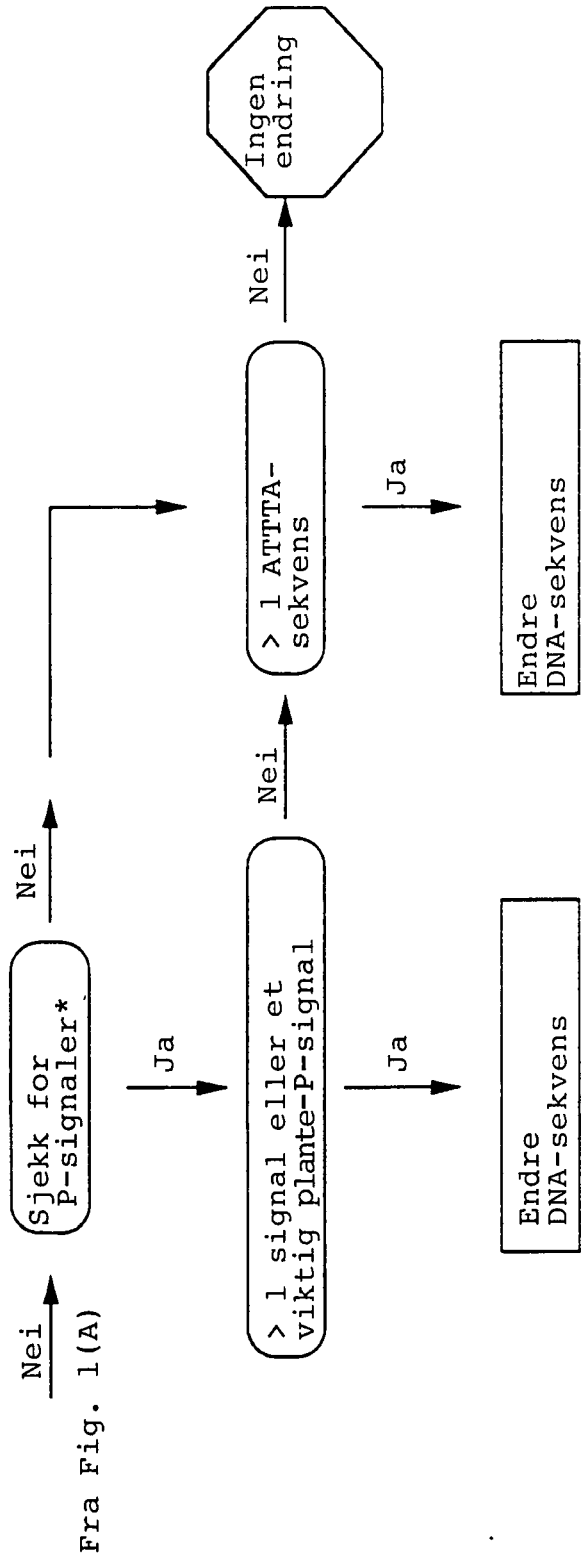
- a) hvilke som helst 15 nukleotider i den strukturkodende sekvens som inneholder mer enn fire på hverandre følgende nukleotider bestående av enten adenin- eller tyminrester, inneholder ikke
15 innenfor de 15 nukleotidene to polyadenyleringsssekvenser,
- b) den strukturkodende sekvens inneholder ikke en AATAAA- eller AATAAT-DNA-sekvens, og
- c) hvilke som helst 15 nukleotider i den strukturkodende sekvens inneholder ikke innenfor de
20 15 nukleotidene mer enn én ATTTA-sekvens.

25

30

35

Bestemmelse av DNA-områder i gener som skal modifiseres ved
stedrettet mutagenese for økt ekspresjon i planter



* Polyadenyleringssignal-sekvens

FIGUR 1 (B)

1 ATGGCTATAGAACTGGTTACACCCCAATCGATATTTCTT 40
 41 TGTCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGG 80
 81 TGCTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGA 120
 T C
 121 ATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAA 160
 161 TTGAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAG 200
 C C C G C G
 201 GAACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTT 240
 T
 241 TATCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAG 280
 281 ATCCTACTAATCCAGCATTAAAGAGAAGAGATGCGTATTCA 320
 321 ATTCAATGACATGAACAGTGCCCTTACAACCGCTATTCCT 360
 361 CTTTTTGCAGTTCAAATTATCAAGTTCCTCTTTTATCAG 400
 CC C C
 401 TATATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAG 440
 G C C CC C CC C
 441 AGATGTTTCAGTGTGGACAAAGGTGGGGATTTGATGCC 480
 481 GCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTA 520
 521 TTGGCAACTATACAGATCATGCTGTACGCTGGTACAATAC 560
 561 GGGATTAGAGCGTGTATGGGGACCGGATTCTAGAGATTGG 600
 601 ATAAGATATAATCAATTTAGAAGAGAATTAACACTAACTG 640
 C G C C G C GC T
 641 TATTAGATATCGTTTCTCTATTTCCGAACATGATAGTAG 680
 681 AACGTATCCAATTCGAACAGTTTCCCAATTAACAAGAGAA 720

FIGUR 2A

721	ATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTT	760
761	TTCGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAG	800
801	TCCACATTTGATGGATATACTTAATAGTATAACCATCTAT	840
841	ACGGATGCTCATAGAGGAGAAATATTATTGGTCAGGGCATC C C C T C	880
881	AAATAATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATT G C	920
921	CACTTTTCCGCTATATGGAACCTATGGGAAATGCAGCTCCA	960
961	CAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGATAT	1000
1001	GAACATTATCGTCCACCTTATATAGAAGACCTTTTAATAT C	1040
1041	AGGGATAAATAATCAACAACCTATCTGTTCTTGACGGGACA C C C C	1080
1081	GAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTG	1120
1121	TATACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAAT	1160
1161	ACCGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTT	1200
1201	AGTCATCGATTAAGCCATGTTTCAATGTTTCGTTTCAGGCT	1240
1241	TTAGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTT	1280
1281	CTCTTGGATACATCGTAGTGCTGAATTTAATAATATAATT G C C C C C	1320
1321	CCTTCATCACAAATTACACAAATACCTTTAACAAAATCTA C C C AC C C G	1360
1361	CTAATCTTGGCTCTGGAACCTCTGTCGTTAAAGGACCAGG	1400

FIGUR 2B

1401 ATTTACAGGAGGAGATATTCTTCGAAGAACTTCACCTGGC 1440
1441 CAGATTTCAACCTTAAGAGTAAATATTACTGCACCATTAT 1480
1481 CACAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCAC 1520
1521 AAATTTACAATTCCATACATCAATTGACGGAAGACCTATT 1560
CC T G C
1561 AATCAGGGGAATTTTTCAGCAACTATGAGTAGTGGGAGTA 1600
1601 ATTTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTAC 1640
1641 TCCGTTTAACTTTTCAAATGGATCAAGTGTATTACGTTA 1680
1681 AGTGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAG 1720
1721 ATCGAATTGAATTTGTTCCGGCA 1743

FIGUR 2C

1	ATGGATAACAATCCGAACATCAATGAATGCATTCCTTATA	40
	C C A C A C	
41	ATTGTTTAAGTAACCCCTGAAGTAGAAGTATTAGGTGGAGA	80
	C C G A T C T	
81	AAGAATAGAACTGGTTACACCCCAATCGATATTCCTTG	120
	C C T C T C C C	
121	TCGCTAACGCAATTTCTTTTGAAGTGAATTTGTTCCCGGTG	160
	CT G A G GC C C G C G A	
161	CTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGAAT	200
	G C TC C C C C T	
201	TTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT	240
	C A T C G G	
241	GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGA	280
	G G C G G C G C	
281	ACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTA	320
	G C G G T G C	
321	TCAAATTTACGCAGAATCTTTTAGAGAGTGGAAGCAGAT	360
	C C T GAGC C C	
361	CCTACTAATCCAGCATTAAAGAGAAGAGATGCGTATTCAAT	400
	C TC CC C G A	
401	TCAATGACATGAACAGTGCCCTTACAACCGCTATTCCTCT	440
	C C T G C A C AT	
441	TTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA	480
	G C C G C C C G C G	
481	TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAG	520
	C A T C T CC CAGC GC TC	
521	ATGTTTCAGTGTTTGGACAAAGGTGGGGATTTGATGCCGC	560
	C AGC G C T	
561	GACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT	600
	A C C C C CC T G	
601	GGCAACTATACAGATcATGCTGTaCGCTGGTACAATACGG	640
	A C C CC C T T C T	
641	GATTAGAGCGTGTATGGGGACCGGATTCTAGAGATTGGAT	680
	C G C T T	

FIGUR 3A

681	AAGATATAATCAATTTAGAAGAGAATTAACACTAACTGTÄ	720
	T C C G C G G C C A T	
721	TTAGATATCGTTTCTCTATTTCCGAACATGATAGTAGAA	760
	G C T G C C CTCC	
761	CGTATCCAATTCGAACAGTTTCCCAATTAACAAGAGAAAT	800
	C C T C T G C T C	
801	TTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT	840
	C T TC T G C C C C C	
841	CGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAGTC	880
	T T T C A T C CTCC C C	
881	CACATTTGATGGATATACTTAATAGTATAACCATCTATAC	920
	C C CT G C C T C	
921	GGATGCTCATAGAGGAGAATATTATTGGTCAGGGCATCAA	960
	C C G C T A C G	
961	ATAATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATTCA	1000
	C C A T A CAGC C G T	
1001	CTTTTCCGCTATATGGAACATATGGGAAATGCAGCTCCACA	1040
	C T C C C	
1041	ACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA	1080
	C T C C	
1081	ACATTATCGTCCACCTTATATAGAAGACCTTTTAATATAG	1120
	C G T G C C C C	
1121	GGATAAATAATCAACAACCTATCTGTTCTTGACGGGACAGA	1160
	T C C C G T C A	
1161	ATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTÄ	1200
	G C C T T C T	
1201	TACAGAAAAGCGGAACGGTAGATTCGCTGGATGAAATAC	1240
	G C T CT C C	
1241	CGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTTAG	1280
	A C T C CTC	
1281	TCATCGATTAAGCCATGTTTCAATGTTTCGTTTCAGGCTTT	1320
	C CA G G C G C C C A C	
1321	AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCT	1360
	C C TCC G C C C	
1361	CTTGGATACATCGTAGTGCTGAATTTAATAATATAATTCC	1400
	A T G C C C	

FIGUR 3B

1401	TTCATCACAAATTACACAAATACCTTTAACAAAATCTACT	1440
	C T C C C A G C G	
1441	AATCTTGGCTCTGGAACCTTCTGTCGTTAAAGGACCAGGAT	1480
	C A G C	
1481	TTACAGGAGGAGATATTCTTCGAAGAACTTCACCTGGCCA	1520
	C T A T	
1521	GATTTCAACCTTAAGAGTAAATATTACTGCACCATTATCA	1560
	AGC C C T C C C T T	
1561	CAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCACAA	1600
	T C G T A A	
1601	ATTTACAATTCCATACATCAATTGACGGAAGACCTATTAA	1640
	C G C C C C G C	
1641	TCAGGGGAATTTTTCAGCAACTATGAGTAGTGGGAGTAAT	1680
	T C C C C TCA C C C C	
1681	TTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTACTC	1720
	G A C C A C C C	
1721	CGTTTAACTTTTCAAATGGATCAAGTGTATTTACGTTAAG	1760
	T C C T C C T C CC T	
1761	TGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAGAT	1800
	C G T G C T C	
1801	CGAATTGAATTTGTTCCGGCAGAAGTAACCTTTGAGGCAG	1840
	T G G T C T C T	
1841	AATAT	1845
	G C	

FIGUR 3C

1	ATGGATAACAATCCGAACATCAATGAATGCATTTCCTTATA	40
	C C A C A C	
41	ATTGTTTAAGTAACCCTGAAGTAGAAGTATTAGGTGGAGA	80
	C C G A T C T	
81	AAGAATAGAACTGGTTACACCCCAATCGATATTTCCCTTG	120
	C C T C T C C C	
121	TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTG	160
	CT G A G GC C C G C G A	
161	CTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGAAT	200
	G C TC C C C C T	
201	TTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT	240
	C A T C G G	
241	GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGA	280
	G G C G G C G C	
281	ACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTA	320
	G C G G T G C	
321	TCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT	360
	C C T GAGC C C	
361	CCTACTAATCCAGCATTAAAGAGAAGAGATGCGTATTCAAT	400
	C TC CC C G A	
401	TCAATGACATGAACAGTGCCCTTACAACCGCTATTCCCTCT	440
	C C T G C A C AT	
441	TTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA	480
	G C C G C C C G C G	
481	TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAG	520
	C A T C T CC CAGC GC TC	
521	ATGTTTCAGTGTGTTGGACAAAGGTGGGGATTTGATGCCGC	560
	C AGC G C T	
561	GACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT	600
	A C C C C CC T G	
601	GGCAACTATACAGATTATGCTGTACGCTGGTACAATACGG	640
	A C C CC C T T C T	
641	GATTAGAACGTGTATGGGGACCGGATTCTAGAGATTGGGT	680
	C G G C T T A	

FIGUR 4A

681	AAGGTATAATCAATTTAGAAGAGAATTAACACTAACTGTA	720
	T A C C G C G G C C A T	
721	TTAGATATCGTTGCTCTGTTCCCGAATTATGATAGTAGAA	760
	G C T GT C C CTCC	
761	GATATCCAATTCGAACAGTTTCCCAATTAACAAGAGAAAT	800
	CC C T C T G C T C	
801	TTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT	840
	C T TC T G C C C C C	
841	CGAGGCTCGGCTCAGGGCATAGAAAGAAGTATTAGGAGTC	880
	T T T C A T C G CTCC C C	
881	CACATTTGATGGATATACTTAACAGTATAACCATCTATAC	920
	C C CT G C T C	
921	GGATGCTCATAGGGGTTATTATTATTGGTCAGGGCATCAA	960
	C C A AG G C T A C G	
961	ATAATGGCTTCTCCTGTAGGGTTTTCGGGGCCAGAATTCA	1000
	C C A T A CAGC C G T	
1001	CTTTTCCGCTATATGGAACCTATGGGAAATGCAGCTCCACA	1040
	C T C C C	
1041	ACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA	1080
	C T C C	
1081	ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAG	1120
	C G T C G C C C C	
1121	GGATAAATAATCAACAACCTATCTGTTCTTGACGGGACAGA	1160
	T C C C G T C A	
1161	ATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA	1200
	G C C T T C T	
1201	TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATAC	1240
	G C T CT C C	
1241	CGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTTAG	1280
	A C T C CTC	
1281	TCATCGATTAAAGCCATGTTTCAATGTTTCGTTTCAGGCTTT	1320
	C CA G G C G C C C A C	
1321	AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCT	1360
	C C TCC G C C C	
1361	CTTGGATACATCGTAGTGCTGAATTTAATAATATAATTGC	1400
	C G C C C C	

FIGUR 4B

```

1401  ATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC  1440
      C
1441  TTTCTTTTAAATGGTTCTGTAATTTTCAGGACCAGGATTTA  1480
      C C C C C
1481  CTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAAATAA  1520
      A C C C C C C
1521  CATTCAGAATAGAGGGTATATTGAAGTTCCAATTCACTTC  1560
1561  CCATCGACATCTACCAGATATCGAGTTCGTGTACGGTATG  1600
      C A GA
1601  CTTCTGTAACCCCGATTCACTCAACGTTAATTGGGGGTAA  1640
      G T
1641  TTCATCCATTTTTTCCAATACAGTACCAGCTACAGCTACG  1680
      C C T C
1681  TCATTAGATAATCTACAATCAAGTGATTTTGGTTATTTTG  1720
      C G C C C C C
1721  AAAGTGCCAATGCTTTTACATCTTCATTAGGTAATATAGT  1760
      C C C C
1761  AGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATA  1800
      G C T C
1801  GACAGATTGGAATTTATTCCAGTTACTGCAACACTCGAGG  1840
      C G C
1841  CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC  1880
      A TGCG
1881  GCTGTTTACGTCTACAAACCAACTAGGGCTAAAAACAAAT  1920
      CTGT ACGTCTACA C AGCT G ACTC G CA TG
1921  G 1921

```

FIGUR 4C

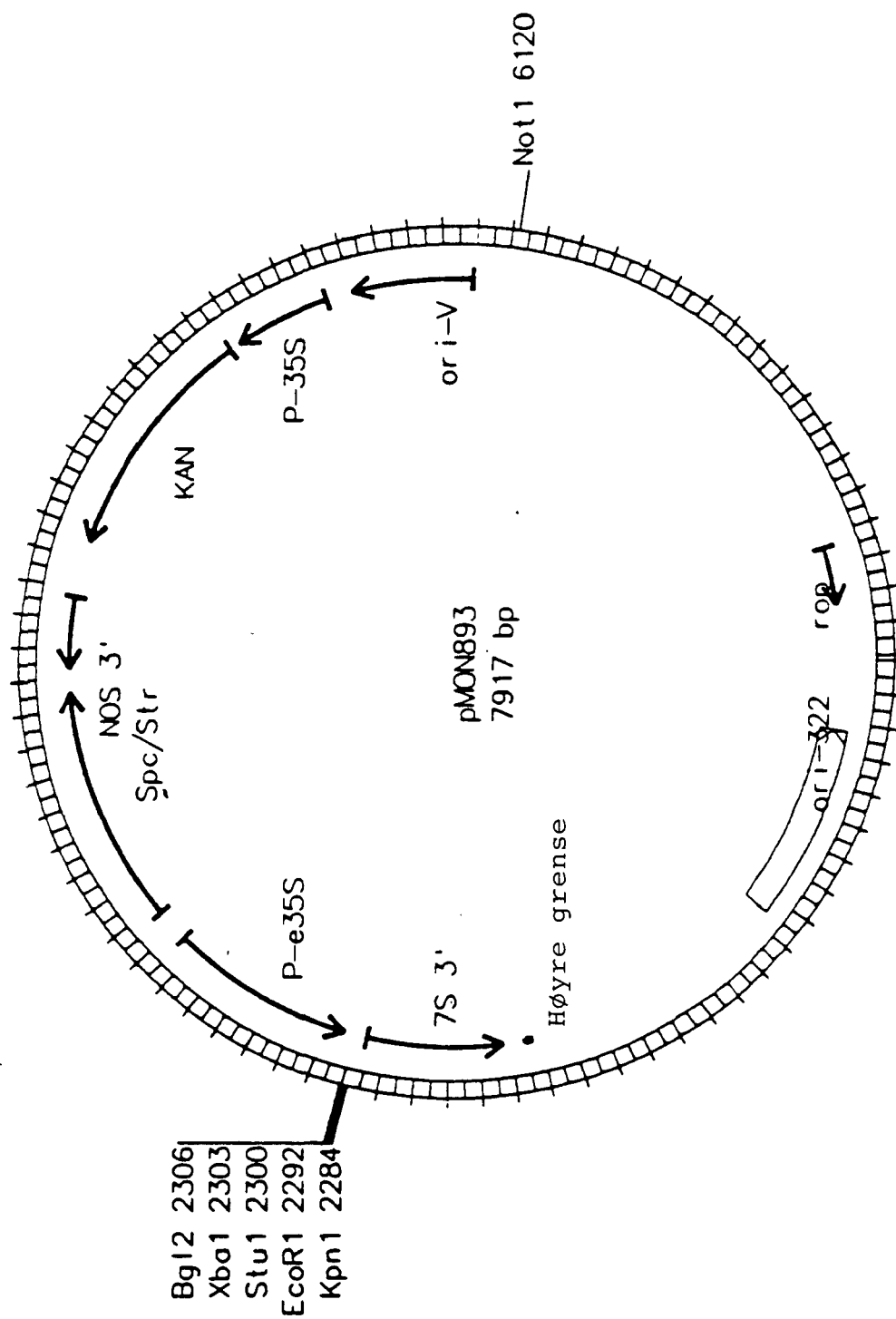
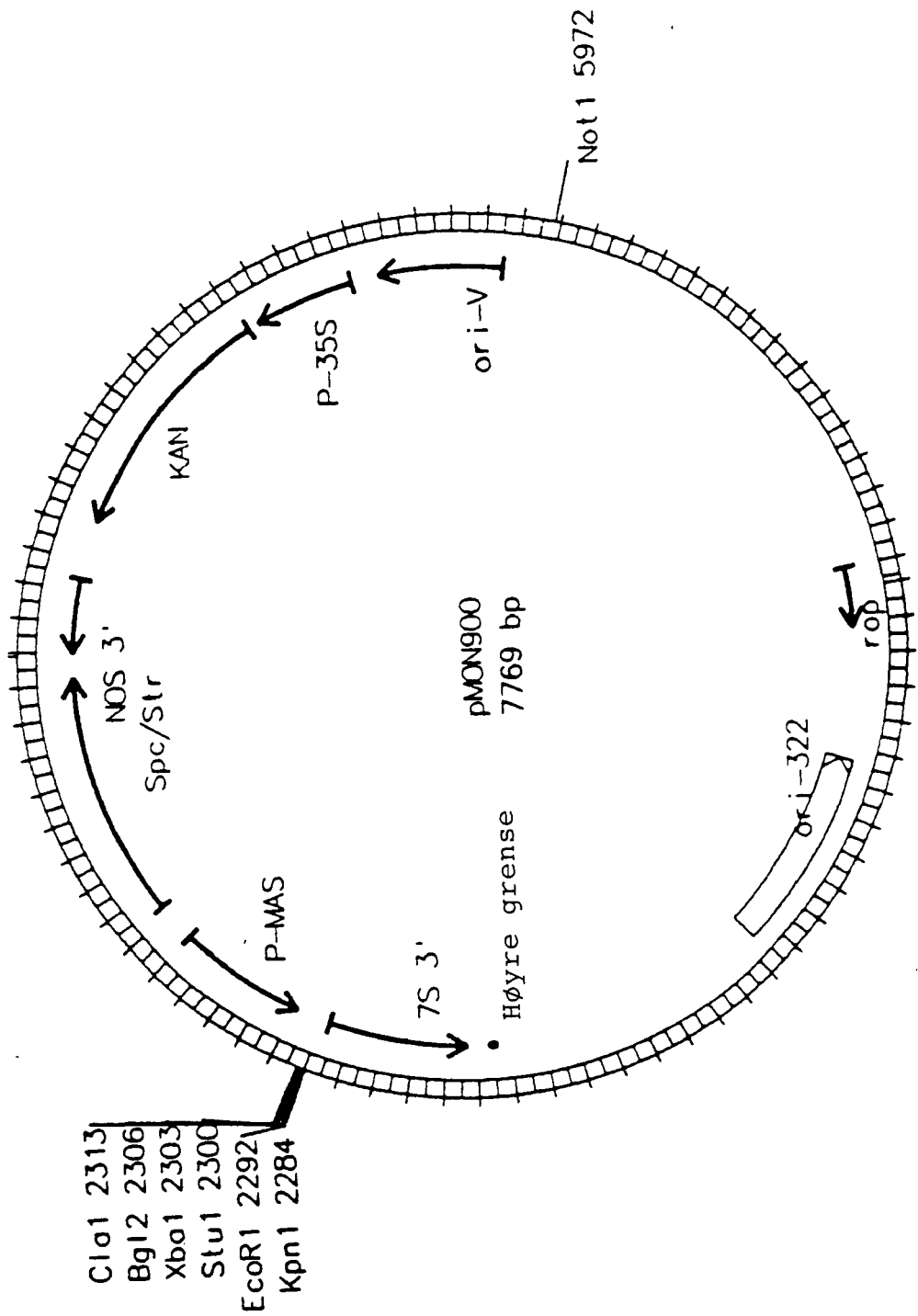
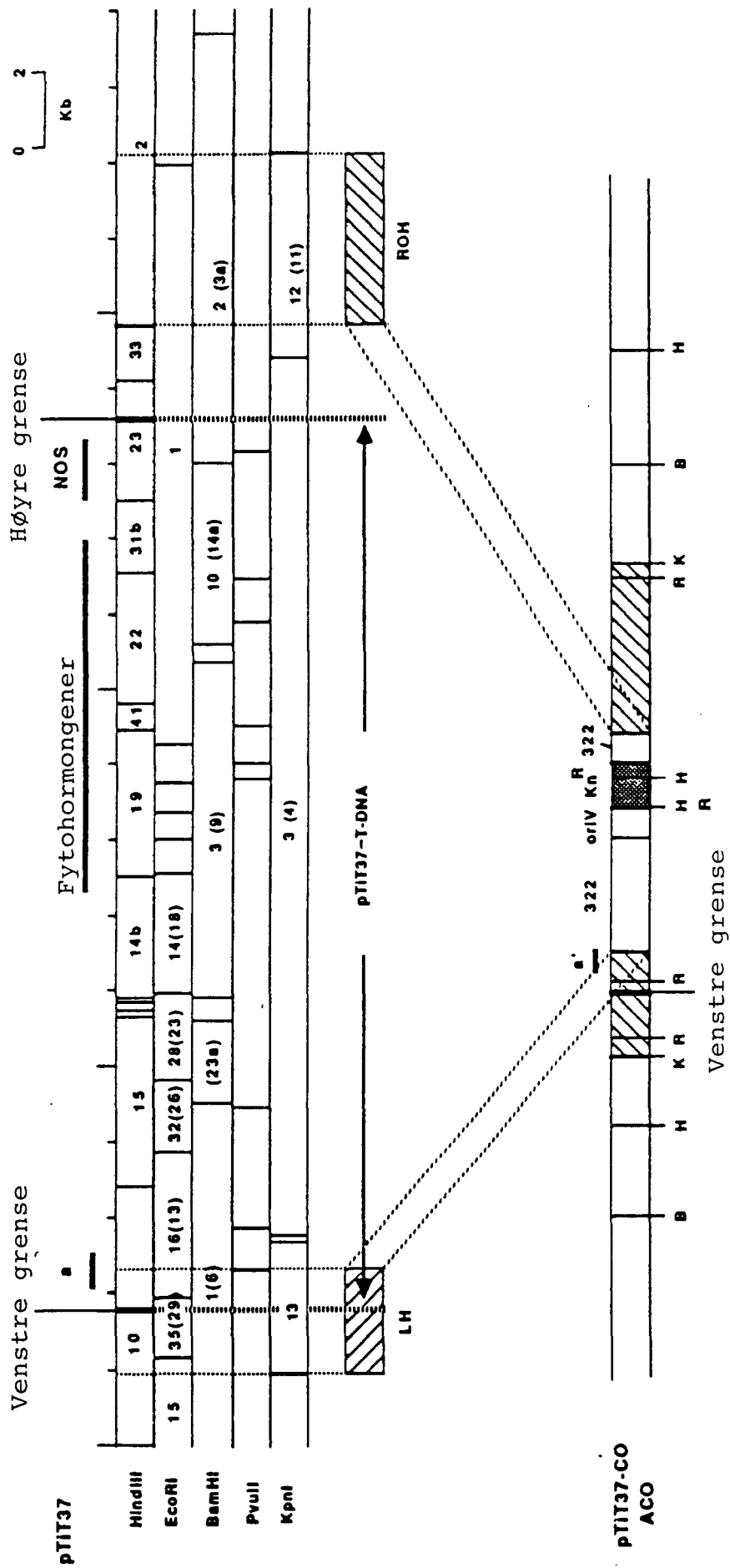


FIGURE 5



FIGUR 6



FIGUR 7

1	GAAAGAATAGAACTGGTTACACCCCAATCGATATTTTCCT	40
	ATGGCC T C T C C C	
41	TGTCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGG	80
	CT G A G GC C C G C G A	
81	TGCTGGATTGTGTAGGACTAGTTGATATAATATGGGGA	120
	G C TC C C C C T	
121	ATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTAACAA	160
	C A T C G G	
161	TTGAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAG	200
	G G C G G C G C	
201	GAACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTT	240
	G C G G T G C	
241	TATCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAG	280
	C C T GAGC C C	
281	ATCCTACTAATCCAGCATTAAAGAGAAGAGATGCGTATTCA	320
	C TC CC C G A	
321	ATTCAATGACATGAACAGTGCCCTTACAACCGCTATTCCT	360
	C C T G C A C A	
361	CTTTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAG	400
	T G C C G C C C G C	
401	TATATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAG	440
	G C A T C T CC CAGC GC TC	
441	AGATGTTTCAGTGTTTGGACAAAGGTGGGGATTGATGCC	480
	C AGC G C T	
481	GCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTA	520
	A C C C C CC T G	
521	TTGGCAACTATACAGATTATGCTGTACGCTGGTACAATAC	560
	A C C CC C T T C	
561	GGGATTAGAACGTGTATGGGGACCGGATTCTAGAGATTGG	600
	T C G G C T T	
601	GTAAGGTATAATCAATTTAGAAGAGAATTAACACTAACTG	640
	A T A C C G C G G C C A	
641	TATTAGATATCGTTGCTCTGTTCCCGAATTATGATAGTAG	680
	T G C T GT C C CTCC	

FIGURE 8A

681	AAGATATCCAATTTCGAACAGTTTCCCAATTAACAAGAGAA	720
	CC C T C T G C T C	
721	ATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTT	760
	C T TC T G C C C C	
761	TTCGAGGCTCGGCTCAGGGCATAGAAAGAAGTATTAGGAG	800
	C T T T C A T C G CTCC C	
801	TCCACATTTGATGGATATACTTAACAGTATAACCATCTAT	840
	C C C CT G C T C	
841	ACGGATGCTCATAGGGGTTATTATTATTGGTCAGGGCATC	880
	C C A AG G C T A C	
881	AAATAATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATT	920
	G C C A T A CAGC C G	
921	CACTTTTCCGCTATATGGAACCTATGGGAAATGCAGCTCCA	960
	T C T C C C	
961	CAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATA	1000
	C T C C	
1001	GAACATTATCGTCCACTTTTATATAGAAGACCTTTTAATAT	1040
	C G T C G C C C	
1041	AGGGATAAATAATCAACAACCTATCTGTTCTTGACGGGACA	1080
	C T C C C G T C A	
1081	GAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTG	1120
	G C C T T C	
1121	TATACAGAAAAAGCGGAACGGTAGATTTCGCTGGATGAAAT	1160
	T G C T CT C	
1161	ACCGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTT	1200
	C A C T C C	
1201	AGTCATCGATTAAGCCATGTTTCAATGTTTCGTTTCAGGCT	1240
	TCC CA G G C G C C A	
1241	TTAGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTT	1280
	C C C TCC G C C C	
1281	CTCTTGGATACATCGTAGTGCTGAATTTAATAATATAATT	1320
	C G C C C C C	
1321	GCATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAA	1360
	C	
1361	ACTTTCTTTTAAATGGTTCTGTAATTTTCAGGACCAGGATT	1400
	C C C C	

FIGUR 8B

1401	TACTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAAT	1440
	C A C C C C C C	
1441	AACATTCAGAAATAGAGGGTATATTGAAGTTCCAATTCCT	1480
1481	TCCCATCGACATCTACCAGATATCGAGTTCGTGTACGGTA	1520
	C A GA	
1521	TGCTTCTGTAAACCCCGATTACCTCAACGTTAATTGGGGT	1560
	G T	
1561	AATTCATCCATTTTTTCCAATACAGTACCAGCTACAGCTA	1600
	C C T	
1601	CGTCATTAGATAATCTACAATCAAGTGATTTTGGTTATTT	1640
	C C G C C C C C	
1641	TGAAAGTGCCAATGCTTTTACATCTTCATTAGGTAATATA	1680
	C C C C	
1681	GTAGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAA	1720
	G C T	
1721	TAGACAGATTTGAATTTATTCCAGTTACTGCAACACTCGA	1760
	C C G C	
1761	GGCTGAA 1767	
	G	

FIGUR 8C

1	ATGGATAACAATCCGAACATCAATGAATGCATTTCCTTATA	40
	C C A C A C	
41	ATTGTTTAAGTAACCCTGAAGTAGAAGTATTAGGTGGAGA	80
	C C G A T C T	
81	AAGAATAGAACTGGTTACACCCCAATCGATATTCCTTG	120
	C C T C T C C C	
121	TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTG	160
	CT G A G GC C C G C G A	
161	CTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGAAT	200
	G C TC C C C C T	
201	TTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT	240
	C A T C G G	
241	GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGA	280
	G G C G G C G C	
281	ACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTA	320
	G C G G T G C	
321	TCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT	360
	C C T GAGC C C	
361	CCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCAAT	400
	C TC CC C G A	
401	TCAATGACATGAACAGTGCCCTTACAACCGCTATTCCTCT	440
	C C T G C A C AT	
441	TTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA	480
	G C C G C C C G C G	
481	TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAG	520
	C A T C T CC CAGC GC TC	
521	ATGTTTCAGTGTTTGGACAAAGGTGGGGATTTGATGCCGC	560
	C AGC G C T	
561	GACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT	600
	A C C C C CC T G	
601	GGCAACTATACAGATTATGCTGTACGCTGGTACAATACGG	640
	A C C CC C T T C T	
641	GATTAGAACGTGTATGGGGACCGGATTCTAGAGATTGGGT	680
	C G G C T T A	

FIGUR 9A

681	AAGGTATAATCAATTTAGAAGAGAATTAACACTAACTGTA	720
	T A C C G C G G C C A T	
721	TTAGATATCGTTGCTCTGTTCCCGAATTATGATAGTAGAA	760
	G C T GT C C CTCC	
761	GATATCCAATTCGAACAGTTTCCCAATTAACAAGAGAAAT	800
	CC C T C T G C T C	
801	TTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT	840
	C T TC T G C C C C C	
841	CGAGGCTCGGCTCAGGGCATAGAAAGAAGTATTAGGAGTC	880
	T T T C A T C G CTCC C C	
881	CACATTTGATGGATATACTTAACAGTATAACCATCTATAC	920
	C C CT G C T C	
921	GGATGCTCATAGGGGTTATTATTATTGGTCAGGGCATCAA	960
	C C A AG G C T A C G	
961	ATAATGGCTTCTCCTGTAGGGTTTTCGGGGCCAGAATTCA	1000
	C C A T A CAGC C G T	
1001	CTTTTCCGCTATATGGAACCTATGGGAAATGCAGCTCCACA	1040
	C T C C C	
1041	ACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA	1080
	C T C C	
1081	ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAG	1120
	C G T C G C C C C	
1121	GGATAAATAATCAACAACCTATCTGTTCTTGACGGGACAGA	1160
	T C C C G T C A	
1161	ATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA	1200
	G C C T T C T	
1201	TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATAC	1240
	G C T CT C C	
1241	CGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTTAG	1280
	A C T C CTC	
1281	TCATCGATTAAAGCCATGTTTCAATGTTTCGTTTCAGGCTTT	1320
	C CA G G C G C C C A C	
1321	AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCT	1360
	C C TCC G C C C	
1361	CTTGGATACATCGTAGTGCTGAATTTAATAATATAATTGC	1400
	C G C C C C C	

FIGUR 9B

1401	ATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC C	1440
1441	TTTCTTTTAAATGGTTCTGTAATTCAGGACCAGGATTTA C C C C C	1480
1481	CTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAATAA A C C C C C C	1520
1521	CATTCAGAATAGAGGGTATATTGAAGTTCCTCAATTCCTTC	1560
1561	CCATCGACATCTACCAGATATCGAGTTCGTGTACGGTATG C A GA	1600
1601	CTTCTGTAAACCCCGATTCACTCAACGTTAATTGGGGTAA G T	1640
1641	TTCATCCATTTTTTCCAATACAGTACCAGCTACAGCTACG C C T C	1680
1681	TCATTAGATAATCTACAATCAAGTGATTTTGGTTATTTTG C G C C C C C	1720
1721	AAAGTGCCAATGCTTTTACATCTTCATTAGGTAATATAGT C C C C	1760
1761	AGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATA G C T C	1800
1801	GACAGATTGGAATTTATTCCAGTTACTGCAACACTCGAGG C G C	1840
1841	CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC	1880
1881	GCTGTTTACGTCTACAAACCACTAGGGCTAAAAACAAAT	1920
1921	GTAACGGATTATCATATTGATCAAGTGTCCAATTTAGTTA	1960
1961	CGTATTTATCGGATGAATTTTGTCTGGATGAAAAGCGAGA	2000
2001	ATTGTCCGAGAAAGTCAAACATGCGAAGCGACTCAGTGAT	2040
2041	GAACGCAATTTACTCCAAGATTCAAATTTCAAAGACATTÀ	2080
2081	ATAGGCAACCAGAACGTGGGTGGGGCGGAAGTACAGGGAT	2120

FIGUR 9C

2121	TACCATCCAAGGAGGGGATGACGTATTTAAAGAAAATTAC	2160
2161	GTCACACTATCAGGTACCTTTGATGAGTGCTATCCAACAT	2200
2201	ATTTGTATCAAAAAATCGATGAATCAAAATTAAAAGCCTT	2240
2241	TACCCGTTATCAATTAAGAGGGTATATCGAAGATAGTCAA	2280
2281	GACTTAGAAATCTATTTAATTCGCTACAATGCAAAACATG	2320
2321	AAACAGTAAATGTGCCAGGTACGGGTTTCCTTATGGCCGCT	2360
2361	TTCAGCCCAAAGTCCAATCGGAAAGTGTGGAGAGCCGAAT	2400
2401	CGATGCGCGCCACACCTTGAATGGAATCCTGACTTAGATT	2440
2441	GTTCTGTAGGGATGGAGAAAAGTGTGCCCATCATTCGCA	2480
2481	TCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTA	2520
2521	AATGAGGACCTAGGTGTATGGGTGATCTTTAAGATTAGA	2560
2561	CGCAAGATGGGCACGCAAGACTAGGGAATCTAGAGTTTCT	2600
2601	CGAAGAGAAACCATTAGTAGGAGAAGCGCTAGCTCGTGTG	2640
2641	AAAAGAGCGGAGAAAAAATGGAGAGACAAACGTGAAAAAT	2680
2681	TGGAATGGGAAACAAATATCGTTTATAAAGAGGCAAAGA	2720
2721	ATCTGTAGATGCTTTATTTGTAAACTCTCAATATGATCAA	2760
2761	TTACAAGCGGATACGAATATTGCCATGATTCATGCGGCAG	2800
2801	ATAAACGTGTTTCATAGCATTTCGAGAAGCTTATCTGCCTGA	2840

FIGUR 9D

2841	GCTGTCTGTGATTCCGGGTGTCAATGCGGCTATTTTGAÄ	2880
2881	GAATTAGAAGGGCGTATTTTCACTGCATTCTCCCTATATG	2920
2921	ATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATGG	2960
2961	CTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAA	3000
3001	GAACAAAACAACCAACGTTTCGGTCCTTGTTGTTCCGGAAT	3040
3041	GGGAAGCAGAAGTGTCAACAAGAAGTTCGTGTCTGTCCGGG	3080
3081	TCGTGGCTATATCCTTCGTGTCACAGCGTACAAGGAGGGA	3120
3121	TATGGAGAAGGTTGCGTAAACCATTCATGAGATCGAGAACA	3160
3161	ATACAGACGAACTGAAGTTTAGCAACTGCGTAGAAGAGGA	3200
3201	AATCTATCCAAATAACACGGTAACGTGTAATGATTATACT	3240
3241	GTAAATCAAGAAGAATACGGAGGTGCGTACACTTCTCGTA	3280
3281	ATCGAGGATATAACGAAGCTCCTTCCGTACCAGCTGATTÄ	3320
3321	TGCGTCAGTCTATGAAGAAAATCGTATACAGATGGACGA	3360
3361	AGAGAGAATCCTTGTGAATTTAACAGAGGGTATAGGGATT	3400
3401	ACACGCCACTACCAGTTGGTTATGTGACAAAAGAATTAGA	3440
3441	ATACTTCCCAGAAACCGATAAGGTATGGATTGAGATTGGA	3480
3481	GAAACGGAAGGAACATTTATCGTGGACAGCGTGGAATTAC	3520
3521	TCCTTATGGAGGAA	3534

FIGUR 9E

1	ATGGATAACAATCCGAACATCAATGAATGCATTCCTTATA	40
	C C A C A C	
41	ATTGTTTAAGTAACCCTGAAGTAGAAGTATTAGGTGGAGA	80
	C C G A T C T	
81	AAGAATAGAACTGGTTACACCCCAATCGATATTTCCCTTG	120
	C C T C T C C C	
121	TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTG	160
	CT G A G GC C C G C G A	
161	CTGGATTGTGTAGGACTAGTTGATATAATATGGGGAAT	200
	G C TC C C C C T	
201	TTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT	240
	C A T C G G	
241	GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGA	280
	G G C G G C G C	
281	ACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTA	320
	G C G G T G C	
321	TCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT	360
	C C T GAGC C C	
361	CCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCAAT	400
	C TC CC C G A	
401	TCAATGACATGAACAGTGCCCTTACAACCGCTATTCCTCT	440
	C C T G C A C AT	
441	TTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA	480
	G C C G C C C G C G	
481	TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAG	520
	C A T C T CC CAGC GC TC	
521	ATGTTTCAGTGTGTTGGACAAAGGTGGGGATTTGATGCCGC	560
	C AGC G C T	
561	GACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT	600
	A C C C C CC T G	
601	GGCAACTATACAGATTATGCTGTACGCTGGTACAATACGG	640
	A C C CC C T T C T	

FIGUR 10A

641	GATTAGAACGTGTATGGGGACCGGATTCTAGAGATTGGGT C G G C T T A	680
681	AAGGTATAATCAATTTAGAAGAGAATTAACACTAACTGT T A C C G C G G C C A T	720
721	TTAGATATCGTTGCTCTGTTCCTCGAATTATGATAGTAGAA G C T GT C C CTCC	760
761	GATATCCAATTTCGAACAGTTTCCCAATTAACAAGAGAAAT CC C T C T G C T C	800
801	TTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT C T TC T G C C C C C	840
841	CGAGGCTCGGCTCAGGGCATAGAAAGAAGTATTAGGAGTC T T T C A T C G CTCC C C	880
881	CACATTTGATGGATATACTTAACAGTATAACCATCTATAC C C CT G C T C	920
921	GGATGCTCATAGGGGTTATTATTATTGGTCAGGGCATCAA C C A AG G C T A C G	960
961	ATAATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATTCA C C A T A CAGC C G T	1000
1001	CTTTTCCGCTATATGGAACCTATGGGAAATGCAGCTCCACA C T C C C	1040
1041	ACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA C T C C	1080
1081	ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAG C G T C G C C C C	1120
1121	GGATAAATAATCAACAACCTATCTGTTCTTGACGGGACAGA T C C C G T C A	1160
1161	ATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGT G C C T T C T	1200
1201	TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATAC G C T CT C C	1240
1241	CGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTTAG A C T C CTC	1280
1281	TCATCGATTAAAGCCATGTTTCAATGTTTCGTTTCAGGCTTT C CA G G C G C C C A C	1320
1321	AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCT C C TCC G C C C	1360

FIGUR 10B

1361	CTTGGATACATCGTAGTGCTGAATTTAATAATATAATTGC	1400
	C G C C C C C	
1401	ATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC	1440
	C	
1441	TTTCTTTTTTAATGGTTCTGTAAATTTTCAGGACCAGGATTTA	1480
	C C C C C	
1481	CTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAAATAA	1520
	A C C C C C C	
1521	CATTCAGAATAGAGGGTATATTGAAGTTCCAATTCACTTC	1560
1561	CCATCGACATCTACCAGATATCGAGTTCGTGTACGGTATG	1600
	C A GA	
1601	CTTCTGTAACCCCGATTCACTCAACGTTAATTGGGGTAA	1640
	G T	
1641	TTCATCCATTTTTTTCCAATACAGTACCAGCTACAGCTACG	1680
	C C T C	
1681	TCATTAGATAATCTACAATCAAGTGATTTTGGTTATTTTG	1720
	C G C C C C C C	
1721	AAAGTGCCAATGCTTTTACATCTTCATTAGGTAATATAGT	1760
	C C C C	
1761	AGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATA	1800
	G C T C	
1801	GACAGATTTGAATTTATTCCAGTTACTGCAACACTCGAGG	1840
	C G C	
1841	CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC	1880
1881	GCTGTTTACGTCTACAAACCAACTAGGGCTAAAAACAAAT	1920
	G C C C G C	
1921	GTAACGGATTATCATATTGATCAAGTGTCCAATTTAGTTA	1960
	G C G G	
1961	CGTATTTATCGGATGAATTTTGTCTGGATGAAAAGCGAGA	2000
	C CC CAGC G C	
2001	ATTGTCCGAGAAAGTCAAACATGCGAAGCGACTCAGTGAT	2040
2041	GAACGCAATTTACTCCAAGATTCAAATTTCAAAGACATTA	2080

FIGUR 10C

2081	ATAGGCAACCAGAACGTGGGTGGGGCGGAAGTACAGGGAT	2120
2121	TACCATCCAAGGAGGGGATGACGTATTTAAAGAAAATTAC G T C G C G G C	2160
2161	GTCACACTATCAGGTACCTTTGATGAGTGCTATCCAACAT	2200
2201	ATTTGTATCAAAAAATCGATGAATCAAAATTAAAAGCCTT CC C C G G C G C G G	2240
2241	TACCCGTTATCAATTAAGAGGGTATATCGAAGATAGTCAA	2280
2281	GACTTAGAAATCTATTTAATTCGCTACAATGCAAAACATG C C G CC C C	2320
2321	AAACAGTAAATGTGCCAGGTACGGGTTCCTTATGGCCGCT	2360
2361	TTCAGCCCAAAGTCCAATCGGAAAGTGTGGAGAGCCGAAT	2400
2401	CGATGCGCGCCACACCTTGAATGGAATCCTGACTTAGATT	2440
2441	GTTCGTGTAGGGATGGAGAAAAGTGTGCCCATCATTCGCA	2480
2481	TCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTA	2520
2521	AATGAGGACCTAGGTGTATGGGTGATCTTTAAGATTAAGA	2560
2561	CGCAAGATGGGCACGCAAGACTAGGGAATCTAGAGTTTCT	2600
2601	CGAAGAGAAACCATTAGTAGGAGAAGCGCTAGCTCGTGTG	2640
2641	AAAAGAGCGGAGAAAAAATGGAGAGACAAACGTGAAAAAT G G	2680
2681	TGGAATGGGAAACAAATATCGTTTATAAAGAGGCAAAGA G C C C C	2720
2721	ATCTGTAGATGCTTTATTTGTAACTCTCAATATGATCAA	2760
2761	TTACAAGCGGATACGAATATTGCCATGATTCATGCGGCAG	2800

FIGUR 10D

2801	ATAAACGTGTTTCATAGCATTTCGAGAAGCTTATCTGCCTGA	2840
2841	GCTGTCTGTGATTCCGGGTGTCAATGCGGCTATTTTGGAA	2880
2881	GAATTAGAAGGGCGTATTTTCACTGCATTCTCCCTATATG C C	2920
2921	ATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATGG C C C G C C C C	2960
2961	CTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAA	3000
3001	GAACAAAACAACCAACGTTCCGGTCCTTGTTGTTCCGGAAT	3040
3041	GGGAAGCAGAAAGTGTCAACAAGAAGTTCGTGTCTGTCCGGG	3080
3081	TCGTGGCTATATCCTTCGTGTACAGCGTACAAGGAGGGA	3120
3121	TATGGAGAAGGTTGCGTAACCATTCATGAGATCGAGAACÀ	3160
3161	ATACAGACGAACTGAAGTTTAGCAACTGCGTAGAAGAGGA	3200
3201	AATCTATCCAAATAACACGGTAACGTGTAATGATTATACT	3240
3241	GTAAATCAAGAAGAATACGGAGGTGCGTACACTTCTCGTA	3280
3281	ATCGAGGATATAACGAAGCTCCTTCCGTACCAGCTGATTA	3320
3321	TGCGTCAGTCTATGAAGAAAAATCGTATACAGATGGACGA	3360
3361	AGAGAGAATCCTTGTGAATTTAACAGAGGGTATAGGGATT	3400
3401	ACACGCCACTACCAGTTGGTTATGTGACAAAAGAATTAGA	3440
3441	ATACTTCCCAGAAACCGATAAGGTATGGATTGAGATTGGA	3480
3481	GAAACGGAAGGAACATTTATCGTGGACAGCGTGGAATTAC	3520
3521	TCCTTATGGAGGAA	3534

1	ATGGATAACAATCCGAACATCAATGAATGCATTTCCTTATA	40
	C C A C A C	
41	ATTGTTTAAGTAACCCTGAAGTAGAAGTATTAGGTGGAGA	80
	C C G A T C T	
81	AAGAATAGAACTGGTTACACCCCAATCGATATTCCTTG	120
	C C T C T C C C	
121	TCGCTAACGCAATTTCTTTTGAAGTGAATTGTTCCTGGTG	160
	CT G A G GC C C G C G A	
161	CTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGAAT	200
	G C TC C C C C T	
201	TTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT	240
	C A T C G G	
241	GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGA	280
	G G C G G C G C	
281	ACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTA	320
	G C G G T G C	
321	TCAAATTTACGCAGAATCTTTTAGAGAGTGGAAGCAGAT	360
	C C T GAGC C C	
361	CCTACTAATCCAGCATTAAAGAGAAGAGATGCGTATTCAAT	400
	C TC CC C G A	
401	TCAATGACATGAACAGTGCCCTTACAACCGCTATTCCTCT	440
	C C T G C A C AT	
441	TTTTGCAGTTCAAATTATCAAGTTCCTCTTTTATCAGTA	480
	G C C G C C C G C G	
481	TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAG	520
	C A T C T CC CAGC GC TC	
521	ATGTTTCAGTGTTTGGACAAAGGTGGGGATTGATGCCGC	560
	C AGC G C T	
561	GACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT	600
	A C C C C CC T G	
601	GGCAACTATACAGATTATGCTGTACGCTGGTACAATACGG	640
	A C C CC C T T C T	
641	GATTAGAACGTGTATGGGGACCGGATTCTAGAGATTGGGT	680
	C G G C T T A	

FIGUR 11A

681	AAGGTATAATCAATTTAGAAGAGAATTAACACTAACTGTA	720
	T A C C G C G G C C A T	
721	TTAGATATCGTTGCTCTGTTCCCGAATTATGATAGTAGAA	760
	G C T GT C C CTCC	
761	GATATCCAATTTCGAACAGTTTCCCAATTAACAAGAGAAAT	800
	CC C T C T G C T C	
801	TTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT	840
	C T TC T G C C C C C	
841	CGAGGCTCGGCTCAGGGCATAGAAAGAAGTATTAGGAGTC	880
	T T T C A T C G CTCC C C	
881	CACATTTGATGGATATACTTAACAGTATAACCATCTATAC	920
	C C CT G C T C	
921	GGATGCTCATAGGGGTTATTATTATTGGTCAGGGCATCAA	960
	C C A AG G C T A C G	
961	ATAATGGCTTCTCCTGTAGGGTTTTTCGGGGCCAGAATTCA	1000
	C C A T A CAGC C G T	
1001	CTTTTCCGCTATATGGAACATATGGGAAATGCAGCTCCACA	1040
	C T C C C	
1041	ACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA	1080
	C T C C	
1081	ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAG	1120
	C G T C G C C C C	
1121	GGATAAATAATCAACAACCTATCTGTTCTTGACGGGACAGA	1160
	T C C C G T C A	
1161	ATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA	1200
	G C C T T C T	
1201	TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATAC	1240
	G C T CT C C	
1241	CGCCACAGAATAACAACGTGCCACCTAGGCAAGGATTTAG	1280
	A C T C CTC	
1281	TCATCGATTAAAGCCATGTTTCAATGTTTCGTTTCAGGCTTT	1320
	C CA G G C G C C C A C	
1321	AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCT	1360
	C C TCC G C C C	
1361	CTTGGATACATCGTAGTGCTGAATTTAATAATATAATTGC	1400
	C G C C C C C	

FIGUR 11B

1401	ATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAAC C	1440
1441	TTTCTTTTAAATGGTTCTGTAATTCAGGACCAGGATTTA C C C C C	1480
1481	CTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAAATAA A C C C C C C	1520
1521	CATTCAGAAATAGAGGGTATATTGAAGTTCCAATTCACTTC	1560
1561	CCATCGACATCTACCAGATATCGAGTTCGTGTACGGTATG C A GA	1600
1601	CTTCTGTAACCCCGATTCACTCAACGTTAATTGGGGTAA G T	1640
1641	TTCATCCATTTTTTCCAATACAGTACCAGCTACAGCTACG C C T C	1680
1681	TCATTAGATAATCTACAATCAAGTGATTTTGGTTATTTTG C G C C C C	1720
1721	AAAGTGCCAATGCTTTTACATCTTCATTAGGTAATATAGT C C C C	1760
1761	AGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATA G C T C	1800
1801	GACAGATTTGAATTTATTCCAGTTACTGCAACACTCGAGG C G C	1840
1841	CTGAATATAATCTGGAAAGAGCGCAGAAGGCGGTGAATGC G C C T G C T C	1880
1881	GCTGTTTACGTCTACAAACCACTAGGGCTAAAAACAAAT C C C C C T G T CT G T C	1920
1921	GTAACGGATTATCATATTGATCAAGTGTCCAATTTAGTTA T T C C C C G C	1960
1961	CGTATTTATCGGATGAATTTGTCTGGATGAAAAGCGAGA C CC TAGC G C C C C G T	2000
2001	ATTGTCCGAGAAAGTCAAACATGCGAAGCGACTCAGTGAT C C T C C T C C	2040
2041	GAACGCAATTTACTCCAAGATTCAAATTTCAAAGACATT GA G C CT G C C C C	2080
2081	ATAGGCAACCAGAACGTGGGTGGGGCGGAAGTACAGGGAT C G T T C C	2120

FIGURE 11C

2121	TACCATCCAAGGAGGGGATGACGTATTTAAAGAAAATTAC	2160
	C C C T G C G G C	
2161	GTCACACTATCAGGTACCTTTGATGAGTGCTATCCAACAT	2200
	C C C A T C C C T C	
2201	ATTTGTATCAAAAAATCGATGAATCAAAATTAAAAGCCTT	2240
	C C G G G C C C	
2241	TACCCGTTATCAATTAAGAGGGTATATCGAAGATAGTCAA	2280
	C A G C T C C C C	
2281	GACTTAGAAATCTATTTAATTCGCTACAATGCAAAACATG	2320
	C T C CG CA G C G C	
2321	AAACAGTAAATGTGCCAGGTACGGGTTTCCTTATGGCCGCT	2360
	G C G C T C C A	
2361	TTCAGCCCAAAGTCCAATCGGAAAGTGTGGAGAGCCGAAT	2400
	T TC C T G T C	
2401	CGATGCGCGCCACACCTTGAATGGAATCCTGACTTAGATT	2440
	A T G G G C	
2441	GTTCTGTAGGGATGGAGAAAAGTGTGCCCATCATTGCGCA	2480
	C C C C G C T	
2481	TCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTA	2520
	C G C G T C G	
2521	AATGAGGACCTAGGTGTATGGGTGATCTTTAAGATTAAGA	2560
	C A C C C C	
2561	CGCAAGATGGGCACGCAAGACTAGGGAATCTAGAGTTTCT	2600
	C C A T C C T	
2601	CGAAGAGAAACCATTAGTAGGAGAAGCGCTAGCTCGTGTG	2640
	G C T T C	
2641	AAAAGAGCGGAGAAAAAATGGAGAGACAAACGTGAAAAAT	2680
	G A G G G G C	
2681	TGGAATGGGAAACAAATATCGTTTATAAAGAGGCAAAAGA	2720
	C T C C G C	
2721	ATCTGTAGATGCTTTATTTGTAAACTCTCAATATGATCAA	2760
	G C G G C G C G	
2761	TTACAAGCGGATACGAATATTGCCATGATTCATGCGGCAG	2800
	G C C C C C C C C	
2801	ATAAACGTGTTTCATAGCATTTCGAGAAGCTTATCTGCCTGA	2840
	C G C T G CT	

FIGUR 11D

2841	GCTGTCTGTGATTCCGGGTGTCAATGCGGCTATTTTGGAA	2880
	T C C T G C T C C C G	
2881	GAATTAGAAGGGCGTATTTTCACTGCATTCTCCCTATATG	2920
	C T G A C T C T G C	
2921	ATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATGG	2960
	C C C G C C C C	
2961	CTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAA	3000
	C CAG T T G C G G	
3001	GAACAAAACAACCAACGTTCCGGTCCTTGTTGTTCCGGAAT	3040
	G T G C G G T G	
3041	GGGAAGCAGAAGTGTACACAAGAAGTTCGTGTCTGTCCGGG	3080
	T C G A A A	
3081	TCGTGGCTATATCCTTCGTGTACAGCGTACAAGGAGGGA	3120
	A A C T C G C T	
3121	TATGGAGAAGGTTGCGTAACCATTCATGAGATCGAGAACAA	3160
	C T G G C C	
3161	ATACAGACGAACTGAAGTTTAGCAACTGCGTAGAAGAGGA	3200
	C C G T CTC C G A	
3201	AATCTATCCAAATAACACGGTAACGTGTAATGATTATACT	3240
	C C C T T C C C C	
3241	GTAAATCAAGAAGATACGGAGGTGCGTACACTTCTCGTA	3280
	G G G C AGC	
3281	ATCGAGGATATAACGAAGCTCCTTCCGTACCAGCTGATTAA	3320
	CA T C T T C	
3321	TGCGTCAGTCTATGAAGAAAAATCGTATACAGATGGACGA	3360
	C C G C G G C C CA	
3361	AGAGAGAATCCTTGTTGAATTTAACAGAGGGTATAGGGATT	3400
	C T C C G C T C C	
3401	ACACGCCACTACCAGTTGGTTATGTGACAAAAGAATTAGA	3440
	A T C T C G GC T	
3441	ATACTTCCCAGAAACCGATAAGGTATGGATTGAGATTGGA	3480
	G T T G C A G C C T	
3481	GAAACGGAAGGAACATTTATCGTGGACAGCGTGAATTAC	3520
	C G C C GC T	
3521	TCCTTATGGAGGAA	3534
	T G	

FIGUR 11E

1	ATGACTGCAGATAATAATACGGAAGCACTAGATAGCTCTA	40
	C C C C C C C T	
41	CAACAAAAGATGTCATTCAAAAAGGCATTTCCGTAGTAGG	80
	C T G T C G G T C T G	
81	TGATCTCCTAGGCGTAGTAGGTTTCCCGTTTGGTGGAGCG	120
	A C T G G T A T C C C	
121	CTTGTTTCGTTTTTATACAAACTTTTTTAAATACTATTTGGC	160
	C GAGC C C C C C	
161	CAAGTGAAGACCCGTGGAAGGCTTTTATGGAACAAGTAGA	200
	C G T A A C G T	
201	AGCATTGATGGATCAGAAAATAGCTGATTATGCAAAAAAT	240
	TC T G T A C G C	
241	AAAGCTCTTGCAGAGTTACAGGGCCTTCAAAATAATGTGCG	280
	G T G AC C G C G	
281	AAGATTATGTGAGTGCATTGAGTTCATGGCAAAAAAATCC	320
	G C C TCCAGC G G C	
321	TGTGAGTTCACGAAATCCACATAGCCAGGGGCGGATAAGA	360
	T C CA T C A TA C	
361	GAGCTGTTTTCTCAAGCAGAAAGTCATTTTCGTAATTCAA	400
	T C C TCC C CA A C	
401	TGCCTTCGTTTGCAATTTCTGGATACGAGGTTCTATTTCT	440
	AGC T C C T T C	
441	AACAACATATGCACAAGCTGCCAACACACATTTATTTTTTA	480
	C T C T C C G C C	
481	CTAAAAGACGCTCAAATTTATGGAGAAGAATGGGGATACG	520
	T G C G	
521	AAAAAGAAGATATTGCTGAATTTTATAAAAGACAATAAAA	560
	G G C G C C GC T T	
561	ACTTACGCAAGAATATACTGACCATTGTGTCAAATGGTAT	600
	G C C G C C G	
601	AATGTTGGATTAGATAAATTAAGAGGTTCACTTATGAAT	640
	C TC C GC C C T C C G	
641	CTTGGGTAAACTTTAACCGTTATCGCAGAGAGATGACATT	680
	G C A A CA G C	

FIGUR 12A

681	AACAGTATTAGATTTAATTGCACTATTTCCATTGTATGAT	720
	G T GC C C T C C C C	
721	GTTCTGGCTATACCCAAAAGAAAGTTAAAACCGAATTAACAA	760
	GA A C G G T GC T C	
761	GAGACGTTTTTAACAGATCCAATTGTCGGAGTCAACAACCT	800
	GC C T C T	
801	TAGGGGCTATGGAACAACCTTCTCTAATATAGAAAATTAT	840
	T T AGC C C C	
841	ATTCGAAAACCAACATCTATTTGACTATCTGCATAGAATTC	880
	A G C C T C	
881	AATTTACACGCGGTTCCAACCAGGATATTATGGAAATGA	920
	C AA T C T C	
921	CTCTTTCAATTATTGGTCCGGTAATTATGTTTCAACTAGA	960
	C C C C C C	
961	CCAAGCATAGGATCAAATGATATAATCACATCTCCATTCT	1000
	T T C C C	
1001	ATGGAAATAAATCCAGTGAACCTGTACAAAATTTAGAATT	1040
	T C G G G CC T G	
1041	TAATGGAGAAAAAGTCTATAGAGCCGTAGCAAATACAAAT	1080
	C C C G C C C	
1081	CTTGCGGTCTGGCCGTCCGCTGTATATTCAGGTGTTACAA	1120
	C T G A A T C C C	
1121	AAGTGAATTTAGCCAATATAATGATCAAACAGATGAAGC	1160
	G G T G C G C G	
1161	AAGTACACAAACGTACGACTCAAAAAGAAATGTTGGCGCG	1200
	C C C G T C C T C A	
1201	GTCAGCTGGGATTCTATCGATCAATTGCCTCCAGAAACAA	1240
	TCT C C	
1241	CAGATGAACCTCTAGAAAAGGGATATAGCCATCAACTCAA	1280
	C AT G G C C C T	
1281	TTATGTAATGTGCTTTTTAATGCAGGGTAGTAGAGGAACA	1320
	C G C G A TCC G C	
1321	ATCCCAGTGTTAACTTGGACACATAAAAGTGTAGACTTTT	1360
	T G C C GTCC G C	
1361	TTAACATGATTGATTTCGAAAAAATTACACAACCTTCCGTT	1400
	C C AGC G G C T C	

FIGUR 12B

1401	AGTAAAGGCATATAAGTTACAATCTGGTGCTTCCGTTGTC	1440
	G G A C C C G	
1441	GCAGGTCCTAGGTTTACAGGAGGAGATATCATTCAATGCA	1480
	C A C T T C C G	
1481	CAGAAAATGGAAGTGCGGCAACTATTTACGTTACACCGGA	1520
	G C C C A T C G T	
1521	TGTGTCGTA CTCTCAAAAATATCGAGCTAGAATTCATTAT	1560
	T G G C A G A C T C	
1561	GCTTCTACATCTCAGATAACATTTACACTCAGTTTACAGC	1600
	A CAGC C C C C G T	
1601	GGGCACCATTTTAATCAATACTATTTTCGATAAAAACGATAAA	1640
	A C C C G T C T C G C C	
1641	TAAAGGAGACACATTAACGTATAATTCATTTAATTTAGCA	1680
	C T T C C A C A G C C C G	
1681	AGTTTCAGCACACCATTTCGAATTATCAGGGAATAACTTAC	1720
	T C C C C T C T	
1721	AAATAGGCGTCACAGGATTAAGTGCTGGAGATAAAGTTTA	1760
	G C C T C C C C C C	
1761	TATAGACAAAATTGAATTTATTCCAGTGAAT	1791
	C C G G C C C	

FIGUR 12C

```

1  ATG  AATAATGTATTGAATAGTGGAAGAACAACACTATTT  40
    GAC  C  C  C      CTC  T      C  C

41  GTGATGCGTATAATGTAGTAGCCCATGATCCATTTAGTTT  80
    C  C  A  C  C  C  G  T  C      C  C

81  TGAACATAAATCATTAGATACCATCCAAAAAGAATGGATG  120
    C      C  GAGCC C  C  T  T  G  G  G

121 GAGTGGAAAAAGAACAGATCATAGTTTATATGTAGCTCCTG  160
    A      C  T  T  C  CTC C C  C  C  A

161 TAGTCGGAACGTGTGTCTAGTTTTTTTGCTAAAGAAAGTGGG  200
    G  T      A      C  C  CC T  C      G  C

201 GAGTCTTATTGGAAAAAGGATATTGAGTGAATTATGGGGG  240
    CTC  C  C      C  T  C  TCC  C  C      T

241 ATAATATTTCTAGTGGTAGTACAAATCTAATGCAAGATA  280
    C  C      ATC  GTCC  T      C      C

281 TTTTAAGGGAGACAGAACAATTCCTAAATCAAAGACTTAA  320
    C  G      C      G  T  C  C  GC T  C

321 TACAGATACCCTTGCTCGTGTAATGCAGAATTGATAGGG  360
    C  T      T  G  A  A  C  C  T  G      C  T

361 CTCCAAGCGAATATAAGGGAGTTTAATCAACAAGTAGATA  400
    A  C  TC T      C  C  G      G  C

401 ATTTTTTAAACCCTACTCAAACCCTGTTCTTTTATCAAT  440
    C  C  G  T  A      G  T      G  C  T  C

441 AACTTCTTCGGTTAATAACAATGCAGCAATTATTTCTAAAT  480
    C      C  G  C  T      C  C  C  C  C

481 AGATTACCCCAGTTCCAGATACAAGGATAACAGTTGTTAT  520
    G  T      T      T      C      C  CC

521 TATTACCTTTATTTGCACAGGCAGCCAATATGCATCTTTC  560
    TC T  AC C      T      T      C      CT G

561 TTTTATTAGAGATGTTATTCTTAATGCAGATGAATGGGGT  600
    C  C  AC T  C  G  C  C  C  T  C      A

601 ATTTCAGCAGCAACATTACGTACGTATCGAGATTACCTGA  640
    C  T      C  TC TA G  A  CA      C  T

641 GAAATTATACAAGAGATTATTCTAATTATTGTATAAATAC  680
    G  C  C  TC T      C  C  C      C  C  C

```

FIGUR 13A

681	GTATCAAAC [•] TGCGTTT [•] AGAGGGT [•] TAAACAC [•] CCGTTTACAC [•]	720
	T G C C T AC C T TA GC T	
721	GATATGTTAGAATT [•] TTAGAACATATATGTT [•] TTTAAATGTAT [•]	760
	C C T G C G C C CC T C G	
761	TTGAATATGTATCCATT [•] TGGTCATTGTTT [•] TAAATATCAGAG [•]	800
	G C CAG AGTC C C G C	
801	TCTTATGGTATCTTCTG [•] CGCTAATTTATATGCTAGCGGT [•]	840
	CT G G C A C C C C CTCT C	
841	AGTGGACCACAGCAGACACAATCATTTACAGCACAAAAC [•] T	880
	A T GAGC C T G	
881	GGCCATTTT [•] TATATTCTCTTTTCCAAGTTAATT [•] CGAATTA [•]	920
	C G AGCT G C C C C	
921	TATATTATCTGGTATTAGTGGTACTAGGCTTTCTATTACC [•]	960
	C TC CAG CTC G C A C C A	
961	TTCCCTAATATTGGTGGTTT [•] TACGGGTTAGTACTACAAC [•] TC [•]	1000
	T C C AC T A CTCC C	
1001	ATTCATTGAATAGTGCCAGGGTTAATTATAGCGGAGGAGT [•]	1040
	AGCC T CTC A G C C T T	
1041	TTCATCTGGTCTCATAGGGGCGACTAATCTCAATCACAAC [•]	1080
	CAGC AT G T T A CT G C	
1081	TTTAATTGCAGCACGGTCCTCCCTCCTTTATCAACACCAT [•]	1120
	C TC C T G A C GAGC G	
1121	TTGTTAGAAGTTGGCTGGATT [•] CAGGTACAGATCGAGAGGG [•]	1160
	G GTCC T CAGC T C A	
1161	CGTTGCTACCTCTACGAATTGGCAGACAGAATCCTTTCAA [•]	1200
	A A C A C G C	
1201	ACAACTTTAAGTTTAAGGTGTGGTGCTTTTTCAGCCCGTG [•]	1240
	C C T CC TC A C T A	
1241	GAAATTCAAAC [•] TATTTCCCAGATTATTTTATCCGTAATAT [•]	1280
	G C T C C C TA G C	
1281	TTCTGGGGTTCCTTTTAGTTATTAGAAACGAAGATCTAAC [•] A	1320
	C T C C C C G T C C C	
1321	AGACCGTTACACTATAACCAATAAGAAATATAGAAAGTC [•]	1360
	C T AC T T C G T G C C GTC	
1361	CTTCGGGAACACCTGGTGGAGCACGGGCCTATTTGGTATC [•]	1400
	A C T T A A T A A T CC C G	

FIGUR 13B

1401	TGTGCATAACAGAAAAAATAATATCTATGCCGCTAATGAA	1440
	C G G C C C T C C G	
1441	AATGGTACTATGATCCATTGGCGCCAGAAGATTATACAG	1480
	C C T CC T A C T	
1481	GATTTACTATATCGCCAATACATGCCACTCAAGTGAATAA	1520
	C C C T C T C C	
1521	TCAAACCTCGAACATTTATTTCTGAAAAATTTGGAAATCAA	1560
	G A C C C C C G C	
1561	GGTGATTCCTTAAGATTTGAACAAAGCAACACGACAGCTC	1600
	C G G C G TC T C A	
1601	GTTATACGCTTAGAGGGAATGGAAATAGTTACAATCTTTA	1640
	G C TT G C C C C	
1641	TTTAAGAGTATCTTCAATAGGAAATTCAACTATTCGAGTT	1680
	C G TAGC C T T C C C C T	
1681	ACTATAAACGGTAGAGTTTATACTGTTTCAATGTTAATA	1720
	C C AC T C A C T G C	
1721	CCACTACAAATAACGATGGAGTTAATGATAATGGAGCTCG	1760
	T A G C T C C C C CA	
1761	TTTTTCAGATATTAATATCGGTAATATAGTAGCAAGTGAT	1800
	A CAGC C C C T C C C G CTC C	
1801	AATACTAATGTAACGCTAGATATAAATGTGACATTAAACT	1840
	C C T TT G C C CC C T	
1841	CCGGTACTCCATTTGATCTCATGAATATTATGTTTGTGCC	1880
	T A C C	
1881	AACTAATCTTCCACCACTTTAT	1902
	C C T T G C	

FIGUR 13C

1	ATGGAGGAAAATAATCAAAATCAATGCATACCTTACAATT	40
	G C C C T A C	
41	GTTTAAGTAATCCTGAAGAAGTACTTTTGGATGGAGAACG	80
	C G C A G T G C T	
81	GATATCAACTGGTAATTCATCAATTGATAATTTCTCTGTCA	120
	C T C C T C C C C C T C	
121	CTTGTTTCAGTTTCTGGTATCTAACTTTGTACCAGGGGGAG	160
	T G C C A G C C G T T	
161	GATTTTTAGTTGGATTAATAGATTTTGTATGGGGAATAGT	200
	G C C T C C T C C C T C	
201	TGGCCCTTCTCAATGGGATGCATTTCTAGTACAAATTGAA	240
	T A C G G G	
241	CAATTAATTAATGAAAGAATAGCTGAATTTGCTAGGAATG	280
	G G C C G G C G C C C	
281	CTGCTATTGCTAATTTAGAAGGATTAGGAAACAATTTCAA	320
	C C C G G C T C	
321	TATATATGTGGAAGCATTAAAGAATGGGAAGAAGATCCT	360
	C C G C C G G C	
361	AATAATCCAGAAACCAGGACCAGAGTAATTGATCGCTTTC	400
	C G C C T G G C C A A C A	
401	GTATACTTGATGGGCTACTTGAAAGGGACATTCCTTCGTT	440
	A C T G C C C T G G A T C A C	
441	TCGAATTTCTGGATTTGAAGTACCCCTTTTATCCGTTTAT	480
	C A C C C T T C G G C	
481	GCTCAAGCGGCCAATCTGCATCTAGCTATATTAAGAGATT	520
	A T T C C C C T C A	
521	CTGTAATTTTGGAGAAAGATGGGGATTGACAACGATAAA	560
	G C C G G C T C	
561	TGTCAATGAAAACATAATAGACTAATTAGGCATATTGAT	600
	C G T C C T C C C	
601	GAATATGCTGATCACTGTGCAAATACGTATAATCGGGGAT	640
	G C C C T C C C C T C	
641	TAAATAATTTACCGAAATCTACGTATCAAGATTGGATAAC	680
	G C C C T G T T	
681	ATATAATCGATTACGGAGAGACTTAACATTGACTGTATTA	720
	C C C A G G A G C C C A T G	

FIGUR 14A

721	GATATCGCCGCTTTCTTTCCAACTATGACAATAGGAGAT	760
	C T A C G C	
761	ATCCAATTCAGCCAGTTGGTCAACTAACAAGGGAAGTTTA	800
	C T C A G T C A C	
801	TACGGACCCATTAATTAATTTTAATCCACAGTTACAGTCT	840
	T C T C C C T G AAG	
841	GTAGCTCAATTACCTACTTTTAACGTTATGGAGAGCAGCC	880
	C C C T C A C C TC	
881	GAATTAGAAATCCTCATTATTTTGATATATTGAATAATCT	920
	T C G C A C G C C C C	
921	TACAATCTTTACGGATTGGTTTAGTGTTGGACGCAATTTT	960
	T C C C C G T C C	
961	TATTGGGGAGGACATCGAGTAATATCTAGCCTTATAGGAG	1000
	T CA G C C CTCT T	
1001	GTGGTAACATAACATCTCCTATATATGGAAGAGAGGCGAA	1040
	G T C C C T A	
1041	CCAGGAGCCTCCAAGATCCTTTACTTTTAATGGACCGGTAA	1080
	A C TAGT C C C C T A C	
1081	TTTAGGACTTTATCAAATCCTACTTTACGATTATTACAGC	1120
	C A C G T C C GA GC C	
1121	AACCTTGGCCAGCGCCACCATTTAATTTACGTGGTGTGTA	1160
	T T C CC TA A	
1161	AGGAGTAGAATTTTCTACACCTACAAATAGCTTTACGTAT	1200
	G C T G C T C CTC C T C	
1201	CGAGGAAGAGGTACGGTTGATTCTTTAACTGAATTACCGC	1240
	A T A C C G C C C A	
1241	CTGAGGATAATAGTGTGCCACCTCGCGAAGGATATAGTCA	1280
	A C C CA G C CTCC	
1281	TCGTTTATGTCATGCAACTTTTGTTCAAAGATCTGGAACA	1320
	CA G G C C C C G GC T C T	
1321	CCTTTTTTAACAACTGGTGTAGTATTTTCTTGGACCGATC	1360
	A CC C T A A T G C A T	
1361	GTAGTGCAACTCTTACAAATACAATTGATCCAGAGAGAAT	1400
	T C T C C G	

FIGUR 14B .

1401	TAATCAAATACCTTTAGTGAAAGGATTTAGAGTTTGGGGG	1440
	C C A G C G T C C T G A	
1441	GGCACCTCTGTCATTACAGGACCAGGATTTACAGGAGGGG	1480
	A T C C C T	
1481	ATATCCTTCGAAGAAATACCTTTGGTGATTTTGTATCTCT	1520
	T A C T C C GAGC	
1521	ACAAGTCAATATTAATTCACCAATTACCCAAAGATACCGT	1560
	C T C C C C T T T	
1561	TTAAGATTTTCGTTACGCTTCCAGTAGGGATGCACGAGTTA	1600
	C C G A T T C C C T C T A C	
1601	TAGTATTAACAGGAGCGGCATCCACAGGAGTGGGAGGCCA	1640
	C G C C C C A T T C T C T A	
1641	AGTTAGTGTAAATATGCCTCTTCAGAAAATATGGAAATA	1680
	CTCC G C A C G G C	
1681	GGGGAGAACTTAACATCTAGAACATTTAGATATACCGATT	1720
	C G C G C C C C	
1721	TTAGTAATCCTTTTTTCATTTAGAGCTAATCCAGATATAAT	1760
	CTC C CAGT C C T C C T C C	
1761	TGGGATAAGTGAACAACCTCTATTTGGTGCAGGTTCTATT	1800
	C T C C A T AGC C	
1801	AGTAGCGGTGAACTTTATATAGATAAAATTGAAATTATTC	1840
	TCATCT C T G C T C G G C	
1841	TAGCAGATGCAACATTTGAAGCAGAATCTGATTTAGAAAG	1880
	T C C T C C C G T G A C A C C T G	
1881	AGCACAAAAGGCGGTGAATGCCCTGTTTACTTCTTCCAAT	1920
	C G T C C C CA	
1921	CAAATCGGGTTAAAAACCGATGTGACGGATTATCATATTG	1960
	GC T C G T A C T T C C	
1961	ATCAAGTATCCAATTTAGTGATTGTTTATCAGATGAATT	2000
	C G C G CACC ACC TAGC G	
2001	TTGTCTGGATGAAAAGCGAGAATTGTCCGAGAAAGTCAA	2040
	C C C C G T C C T	
2041	CATGCGAAGCGACTCAGTGATGAGCGGAATTTACTTCAAG	2080
	C C T C C A C CT G	
2081	ATCCAAACTTCAGAGGGATCAATAGACAACCAGACCGTGG	2120
	CT C A AC C G G A	

FIGURE 14C

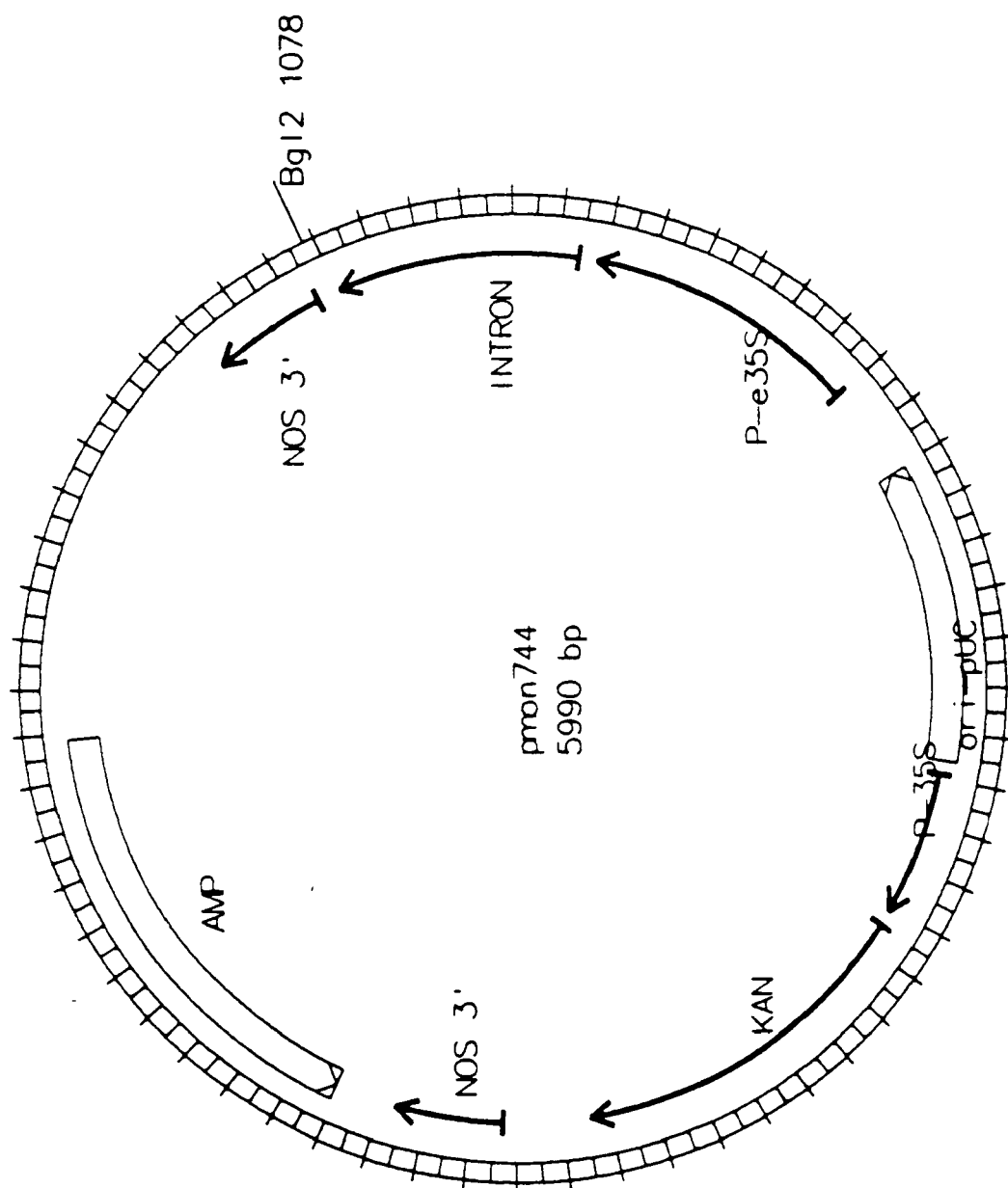
2121	CTGGAGAGGAAGTACAGATATTACCATCCAAGGAGGAGAT	2160
	T G T C C GG C C C	
2161	GACGTATTCAAAGAGAATTACGTACACTACCGGGTACCG	2200
	T G G C C CT C A TT	
2201	TTGATGAGTGCTATCCAACGTATTTATATCAGAAAATAGA	2240
	C C C T C C G C G C	
2241	TGAGTCGAAATTAAGCTTATACCCGTTATGAATTAAGA	2280
	C C C C TC A G C C T	
2281	GGGTATATCGAAGATAGTCAAGACTTAGAAATCTATTTGA	2320
	C C C C C T C C	
2321	TCCGTTACAATGCAAAACACGAAATAGTAAATGTGCCAGG	2360
	A G C G G CC G C	
2361	CACGGGTTCCTTATGGCCGCTTTCAGCCCAAATGCCAATC	2400
	T T C C A T TCT C T	
2401	GGAAAGTGTGGAGAACCGAATCGATGCGCGCCACACCTTG	2440
	G G T CA T	
2441	AATGGAATCCTGATCTAGATTGTTCTGTCAGAGACGGGGA	2480
	G CT G C C G T C	
2481	AAAATGTGCACATCATTCCCATCATTTACCTTGGATATT	2520
	G G C C T C T C C	
2521	GATGTTGGATGTACAGACTTAAATGAGGACTTAGGTGTAT	2560
	G T C G C C A C	
2561	GGGTGATATTCAAGATTAAGACGCAAGATGGCCATGCAAG	2600
	C C C C C A C	
2601	ACTAGGGAATCTAGAGTTTCTCGAAGAGAAACCATTATTA	2640
	T C C T GG C	
2641	GGGGAAGCACTAGCTCGTGTGAAAAGAGCGGAGAAGAAGT	2680
	T T C G A	
2681	GGAGAGACAAACGAGAGAAACTGCAGTTGGAAACAAATAT	2720
	G T CG A G T C	
2721	TGTTTATAAAGAGGCAAAAGAATCTGTAGATGCTTTATTT	2760
	C C G C G C G G C	
2761	GTAAACTCTCAATATGATAGATTACAAGTGGATACGAACA	2800
	G C CAG G CC C C	
2801	TCGCCATGATTCATGCGGCAGATAAACGCGTTCATAGAAT	2840
	C C C C T G C C	

FIGUR 14D

2841 CCGGGAAGCGTATCTGCCAGAGTTGTCTGTGATTCCAGGT 2880
 T T G T CT T C C T
 2881 GTCAATGCGGCCATTTTCGAAGAATTAGAGGGACGTATTT 2920
 G C T C G C T C
 2921 TTACAGCGTATTCCCTTATATGATGCGAGAAATGTCATTAA 2960
 C A TC G C C C C
 2961 AAATGGCGATTTCAATAATGGCTTATTATGCTGGAACGTG 3000
 G C T C C C CAGC T
 3001 AAAGGTCATGTAGATGTAGAAGAGCAAAACAACCACCGTT 3040
 G C G G A G T G
 3041 CGGTCCTTGTTATCCCAGAATGGGAGGCAGAAGTGTACAC 3080
 C G G G T G A T C
 3081 AGAGGTTTCGTGTCTGTCCAGGTCGTGGCTATATCCTTCGT 3120
 A A A A C T C
 3121 GTCACAGCATATAAAGAGGGATATGGAGAGGGCTGCGTAA 3160
 G C T C G C T T T G
 3161 CGATCCATGAGATCGAAGACAATACAGACGAACTGAAATT 3200
 C C GA C C G T G
 3201 CAGCAACTGTGTAGAAGAGGAAGTATATCCAAACAACACA 3240
 TC C C G A A C C C
 3241 GTAACGTGTAATAATTATACTGGGACTCAAGAAGAATATG 3280
 T T C CG C C T A G G C
 3281 AGGGTACGTACACTTCTCGTAATCAAGGATATGACGAAGC 3320
 GA G C AGC CAG T CA
 3321 CTATGGTAATAACCCTTCCGTACCAGCTGATTACGCTTCA 3360
 TCC TCXXXXXXXXXXXX T T C T C C
 3361 GTCTATGAAGAAAAATCGTATACAGATGGACGAAGAGAGA 3400
 G C G G C C CA C T
 3401 ATCCTTGTGAATCTAACAGAGGCTATGGGGATTACACACC 3440
 C C G TC T CA C
 3441 ACTACCGGCTGGTTATGTAACAAAGGATTTAGAGTACTTC 3480
 T A T C T C GC T T
 3481 CCAGAGACCGATAAGGTATGGATTGAGATCGGAGAAACAG 3520
 T C A G C T C
 3521 AAGGAACATTCATCGTGGATAGCGTGGAATTACTCCTTAT 3560
 G C C GC T T G

3561 GGAGGAA 3567

FIGUR 14E



FIGUR 15

1	AGATCTAGAGGTAATTGTTATGAGTACTGTCGTGGTTAAG GATC	40
41	GGAAACGTCAACGGTGGTGTACAACAACCTAGAAGGAGGA G T A	80
81	GAAGGCAATCCCTTCGCAGGAGGGCTAACAGAGTACAGCC T A T	120
121	AGTGGTTATGGTCACTGCTCCTGGCGAACCCAGGAGGAGG GC A A A	160
161	AGACGCAGAAGAGGAGGCAATCGCAGGTCAAGAAGAAGT A G T A	200
201	GAGTTCCCAGGGGAAGGGGCTCAAGCGAGACATTTCGTGTT A A T	240
241	TACAAAGGACAACCTCGTGGGCAACTCCCAAGGAAGTTTC	280
281	ACCTTCGGACCAAGTGTATCAGACTGTCCAGCATTCAAGG T	320
321	ATGGAATACTCAAGGCCTACCATGAGTACAAGATCACAAG T	360
361	TATCCTTCTTCAGTTCGTCAGCGAGGCCTCTTCCACCTCA T G T	400
401	CCAGGATCCATCGCTTATGAGTTGGACCCACATTGCAAAG C A T	440
441	TATCATCCCTCCAGTCCTACGTCAACAAGTTCCAAATCAC T	480
481	AAAGGGAGGAGCTAAGACCTATCAAGCTAGGATGATCAAC T T C T	520
521	GGAGTAGAATGGCACGATTCTCTGAGGATCAGTGCAGGA T T A	560
561	TACTTTGGAAAGGAAGTGGAAATCTTCAGACCCAGCAGG C A G T T	600
601	ATCTTTCAGAGTCACCATCAGAGTGGCTCTTCAAACCCCC T T A	640
641	AAGTAATAGACTCCGGATCAGAGCCTGGTCCAAGCCCACA A T	680

FIGURE 16A

681 ACCAACACCCACTCCAACCTCCCCAAAAGCATGAGCGATT 720
721 ATTGCTTACGTCGGCATACTATGCTGACCATTCAAGAAT 760
761 TC 762

FIGUR. 16B