(19) SUOMI - FINLAND
(FI)

PATENTTI- JA REKISTERIHALLITUS
PATENT- OCH REGISTERSTYRELSEN
FINNISH PATENT AND REGISTRATION OFFICE

(10) **FI 943354 A7**

(12) **JULKISEKSI TULLUT PATENTTIHAKEMUS**
PATENTANSÖKAN SOM BLIVIT OFFENTLIG
PATENT APPLICATION MADE AVAILABLE TO THE PUBLIC

(21) Patenttihakemus - Patentansökan - Patent application 943354

(51) Kansainvälinen patenttiluokitus - Internationell patentklassifikation - International patent classification
C12N 9/72
C12N 15/70
A61K 38/49

(22) Tekemispäivä - Ingivningsdag - Filing date **14.07.1994**

(23) Saapumispäivä - Ankomstdag - Reception date **14.07.1994**

(41) Tullut julkiseksi - Blivit offentlig - Available to the public **16.01.1995**

(43) Julkaisupäivä - Publiceringsdag - Publication date **13.06.2019**

(32) (33) (31) Etuoikeus - Prioritet - Priority

15.07.1993 DE 4323754

(71) Hakija - Sökande - Applicant

1 • **Grünenthal GmbH,** Zieglerstrasse 6, 52078 Aachen, SAKSA, (DE)

(72) Keksijä - Uppfinnare - Inventor

1 • **Steffens, Gerd J.,** Germany, SAKSA, (DE)
2 • **Wnendt, Stephan,** Germany, SAKSA, (DE)
3 • **Schneider, Johannes,** Germany, SAKSA, (DE)
4 • **Heinzel-Wieland, Regina,** Germany, SAKSA, (DE)
5 • **Saunders, Derek John,** Germany, SAKSA, (DE)

(74) Asiamies - Ombud - Agent

Berggren Oy Ab, Antinkatu 3 C, 00100 Helsinki

(54) Keksinnön nimitys - Uppfinningens benämning - Title of the invention

Bifunktionaaliset urokinaasimuunnokset, joilla on parantuneita fibrinolyyttisiä ominaisuuksia ja trombiinia inhiboivaa vaikutusta

Bifunktionella urokinasvarianter med förbättrade fibrinolytiska egenskaper och trombininhiberande verkan

Bifunktionaaliset urokinaasimuunnokset, joilla on parantuneita fibrinolyyttisiä ominaisuuksia ja trombiinia inhiboivaa vaikutusta

Oheisen keksinnön kohteena ovat bifunktionaaliset urokinaasimuunnokset, joilla on parantuneita fibrinolyyttisiä ominaisuuksia ja trombiinia inhiboivaa vaikutusta, plasmidit, joita voidaan käyttää näiden polypeptidien valmistamiseksi, sekä verisuonitukoksia hajottavat lääkkeet (trombolyytit), jotka sisältävät vaikuttavana aineena bifunktionaalista urokinaasimuunnosta.

Ihmisveren eräs tärkeä ominaisuus on kyky sulkea loukkaantunut verisuoni muodostamalla tukos. Veren hyytymisen saa aikaan joukko veressä läsnäolevia entsyymejä, jotka johtavat niinkutsutussa hyytymiskaskadissa lopulta siihen, että trombiinientsyymi muuttaa fibrinogeeni-esiasteproteiinin proteolyyttisesti fibriiniksi. Fibriini polymeroituu sulkien sisäänsä trombosyyttejä (verihiutaleita), erytrosyyttejä (punasoluja) ja veren muita aineosia vaurioituneessa kohdassa, jolloin muodostuu tukos.

Tämän lisäksi veri sisältää myös joukon sellaisia entsyymejä, jotka vastustavat hyytymistä ja jotka huolehtivat veren virtaamisesta sen jälkeen, kun verisuonen seinämä on jälleen ehyt. Trombolyysin kannalta tärkein entsyymi on plasmiini, joka vaikuttaa proteolyyttisesti fibriinimuodostelmaan ja pyrkii tällä tavalla liuottamaan tukoksen. Plasmiinia muodostuu siten, että inaktiivinen esiasteproteiini plasminogeeni pilkkoutuu proteolyyttisesti. Aktivoitumisen saavat aikaan plasminogeeniaktivaattorit plasminogeenin proteolyyttisen pilkkoutumisen seurauksena. Alalla tunnetaan kaksi ihmisen endogeenista plasminogeeniaktivaattoria, eli urokinaasi, joka on virtsassa esiintyvä plasminogeeniaktivaattori, sekä kudosplasminogeeniaktivaattori.

Sydäninfarkti ja aivoinfarkti liittyvät läheisesti tukosten patologiseen muodostumiseen. Näissä molemmissa infarktimuodois-

sa verisuonten seinämiin syntyy tietyissä olosuhteissa tukoksia, useimmiten valtimoiden sairaalloisesta kovettumisesta johtuvien muutosten seurauksena. Nämä tukokset häiritsevät veren virtausta valtimoissa siten, ettei kudoksiin pääse enää riittävästi happea. Sydäninfarktin tapauksessa tämä johtaa sydänlihaksen osittaiseen tai täydelliseen kuolemiseen. Vastaavasti aivovaltimoiden tukkeutuminen johtaa aivokudoksen vakaviin vaurioihin.

Infarktipotilaiden hoitamiseen käytetään tukosten hajottajina plasminogeeniaktivaattoreita, jotka käynnistävät tukosten hajoamisen plasmiinin vaikutuksesta. Tällä hetkellä hoitamiseen on käytettävissä streptokinaasia, APSAC-kompleksi (anisoloidun plasminogeenin ja streptokinaasiaktivaattorin välinen kompleksi), kaksiketjuista urokinaasia (UK), yhdistelmäteknistä yksiketjuista urokinaasia (yhdistelmätekninen prourokinaasi) sekä kudosplasminogeeniaktivaattoria (tPA) [Collen ja Lijnen, Blood, 78, 3114-3124 (1991)]. Streptokinaasi on punasoluja hajottavista streptokokeista peräisin oleva proteiini. Streptokinaasi aktivoi plasminogeenia muodostamalla kompleksin plasminogeenin kanssa, jolloin plasminogeeni saadaan aktiiviseen konformaatioon. Tämä kompleksi itse muuttaa vapaan plasminogeenin plasmiiniksi, joka puolestaan pilkkoo streptokinaasin sitomaa plasminogeenia. Eräs streptokinaasin kehittyneempi muoto on APSAC, joka on streptokinaasista ja ihmisen plasminogeenista in vitro valmistettu yhdiste. APSAC-yhdisteen biologinen puoliintumisaika on pidentynyt streptokinaasiin verrattuna plasminogeenin aktiivisen keskuksen kemiallisen muokkautumisen seurauksena.

Urokinaasi on ihmisen proteiini, jota voidaan saada virtsasta kahdessa eri muodossa olevana proteolyyttisesti aktiivisena proteiinina, eli urokinaasina, jolla on suuri molekyylipaino (HUK) ja urokinaasina, jolla on pieni molekyylipaino (LUK) [Stump et al., J. Biol. Chem., 261, 1267-1273 (1986)]. HUK ja LUK ovat kaksiketjuisia molekyylejä. Urokinaasia muodostuu yksiketjuisena urokinaasina (prourokinaasina) eri kudoksissa, ja se voidaan osoittaa pieninä proentsyymimäärinä ihmisen verestä

[Wun et al., J. Biol. Chem. 257, 3276-3283 (1982)]. Prourokinaasin aktivoituneen HUK-muodon molekyylipaino on 54 kilodaltonia ja se koostuu kolmesta domeenista: aminopäätteen puoleisesta kasvutekijädomeenista, Kringel-domeenista ja seriiniproteaasidomeenista [Günzler et al., Hoppe-Seyler's Z. Physiol. Chem. 363, 1155-1165 (1982); Steffens et al., Hoppe-Seyler's Z. Physiol. Chem., 363,1043-1058 (1982)]. Vaikka prourokinaasi ja plasminogeeni ovatkin proentsyymeinä, niin tästä huolimatta prourokinaasi kykenee sisäisen aktiivisuutensa seurauksena muuttamaan plasminogeenin aktiiviseksi plasmiiniksi. Tämä plasminogeeniaktivaattori saavuttaa täydellisen aktiivisuutensa kuitenkin vasta sitten, kun muodostunut plasmiini on puolestaan pilkkonut prourokinaasin aminohappojen ^{158}lysiini ja ^{159}isoleusiini välistä [Lijnen et al., J. Biol. Chem., 261, 1253-1258 (1986)]. Urokinaasin geeniteknisen valmistuksen Escherichia coli-bakteerissa kuvasi ensimmäisenä Heyneken (Proceedings of the IVth International Symposium on Genetics of Industrial Micro-organisms 1982). Glykosyloitumatonta prourokinaasia (saruplaasi) valmistetaan käyttäen synteettistä geeniä [Brigelius-Flohé et al., Appl. Microbiol. Biotech., 36, 640-649 (1992)].

Kudosplasminogeeniaktivaattori on veressä ja kudoksissa esiintyvä proteiini, jonka molekyylipaino on 72 kilodaltonia. Tämä plasminogeeniaktivaattori koostuu viidestä domeenista: aminopäätteen puoleisesta Finger-domeenista, kasvutekijädomeenista, Kringel 1- ja Kringel 2-domeenista sekä seriiniproteaasidomeenista. Toisin kuin prourokinaasi, tPA kykenee pilkkomaan plasminogeenia vasta sitten, kun se on sitoutunut fibriiniin. Samoin kuin prourokinaasi, tPA muuttuu aktiiviseen muotoonsa Kringel 2-domeenin ja seriiniproteaasidomeenin välisen, plasmiinin katalysoiman pilkkoutumisen seurauksena. Tällöin tämä kudosplasminogeeniaktivaattori sitoutuu fibriiniin, ei kuitenkaan fibrinogeeniin, jolloin plasminogeeni aktivoituu tukosspesifisesti. Toisin kuin kaksiketjuisen urokinaasin tapauksessa yleinen plasminogeeniaktivoituminen vältetään suureksi osaksi [Collen ja Lijnen, Blood, 78, 3114-3124 (1991)].

80-luvun alusta lähtien sydänlihasinfarktin aktiivinen hoitaminen tukosten hajottajilla on osoittautunut tehokkaaksi ja tarkoituksenmukaiseksi. Useissa tutkimuksissa on osoitettu, että sydäninfarktipotilaiden hoitaminen streptokinaasilla, APSAC:lla, UK:lla, yhdistelmätekniselläprourokinaasilla tai tPA:lla on johtanut huomattavasti pienempään kuolleisuuteen ilman hoitoa jätettyihin potilaisiin verrattuna. Näiden aineiden hoitavan vaikutuksen parantamiseksi alalla on valmistettu geeniteknisiä menetelmiä käyttäen joukkoa kudosplasminogeeniaktivaattorin ja prourokinaasin johdanaisia. Paremman fibrinolyyttisen aktiivisuuden ja vähäisempien sivuvaikutusten ohella mielenkiinto kohdistuu sellaisten muotojen kehittämiseen, joita muotoja voidaan antaa boluksena (väkevöitynä annoksena). Yhteenveto plasminogeeniaktivaattoreiden parantamiseksi tehdyistä yrityksistä on löydettävissä julkaisuista Thrombosis and Haemostasis, 66, 88-110 (1991) sekä Trends in Biotech., 9, 86-90 (1991).

Plasminogeeniaktivaattoreiden vaikutuksen parantamiseksi tukoksia hajottavassa hoidossa ja erityisesti niiden biologisen puoliintumisajan pidentämiseksi alalla on valmistettu kudosplasminogeeniaktivaattorin häviämä- ja korvausmuunnoksia, joista on esimerkiksi poistettu Finger- ja kasvutekijädomeeni tai joissa seriiniproteaasidomeeni on vaihdettu urokinaasin seriiniproteaasidomeeniin [Collen et al., Thromb. Haemostasis 65, 174-180 (1991); Fromage et al., Fibrinolysis 5, 187-190 (9191); Lu et al., Blood, 78, 125-131 (1991)]. Tällöin todettiinkin, että Finger- ja kasvutekijädomeenin hävitys pidensi tPA-muunnoksen biologista puoliintumisaikaa [Lijnen ja Collen, Thromb. Haemostasis 66, 94-95 (1991)]. Eräs häviämä- ja korvausmuunnos, joka muodostui tPA:n kummastakin Kringel-domeenista ja UK:n seriiniproteaasidomeenista, oli selvästi pitemmän puoliintumisaikansa takia parempi kuin alkuperäiset, eli muuntamattomat plasminogeeniaktivaattorit tukoksien hajotusta ajatellen. Näillä plasminogeeniaktivaattoreiden muunnoksilla oli kuitenkin vain vähäistä fibriinispesifisyyttä [Lu et al., Blood, 78, 125-131 (1991)].

Alalla on tehty lukuisia yrityksiä valmistaa sellaisia plasminogeeniaktivaattoreita, joiden fibriinispesifisyys on parantunut. Verenvuotovaaran vähentämiseksi tällaisten vaikuttavien aineiden tulisi aktivoida plasminogeenia mahdollisimman tarkoin vain tukoksen läheisyydessä, plasminogeenien systeemistä aktivoitumista kuitenkaan aiheuttamatta. Niinpä alalla tunnetaan esimerkiksi tPA:n sellainen muunnos, jossa Kringel 1-domeeni on vaihdettu lähtömolekyylin Kringel 2-domeeniin. Tällä muunnoksella on tosin parantunut affiniteetti N-päätteen lysiinijäännösten suhteen, muttei fibriinin suhteen. Tämä muunnos ei ollut eläinkokeessa tukosten hajotusta ajatellen alkuperäistä kudosplasminogeeniaktivaattoria vaikutuskykyisempi [Collen et al., Thromb. Haemostasis, 65, 174-180 (1991)]. Muut tunnetut muunnokset, jotka muodostuvat tukosspesifisten vasta-aineiden ja plasminogeeniaktivaattoreiden välisistä fuusioista, ovat eläinkokeessa vaikutuskykyisempiä kuin alkuperäiset plasminogeeniaktivaattorit [Lijnen ja Collen, Throm. Haemostasis, 66, 88-110 (1991)]. Eräällä Desmodus retundus-lepakosta eristetyllä plasminogeeniaktivaattorilla on erittäin suurta fibriinispesifisyyttä [Gardell et al., J. Biol. Chem. 264, 17947-17952 (1989)]. Tällä plasminogeeniaktivaattorilla päästään eläinkokeessa tPA:han verrattuna tukoksen parempaan hajoamiseen ja sen puoliintumisaika on pitempi ja plasminogeenin systeeminen aktivoituminen on vähäisempää [Gardell et al., Circulation 84, 244-253 (1991); Mellot et al., Arterioscl. Thromb. 12, 212-221 (1992)].

Hoidettaessa infarktipotilaita plasminogeeniaktivaattoreilla ei hoidon onnistuminen riipu kuitenkaan vain tukoksen hajoamisesta, vaan myös siitä, kuinka hyvin avatun verisuonen uudelleentukkeutuminen kyetään estämään. Eri havainnot viittaavat siihen, että tukoksen hajotessa tukokseen sitoutunut trombiini vapautuu jälleen aktiivisena entsyyminä, joka voi aiheuttaa verisuonten tukkeutumisen uudelleen [Szczeklik et al., Arterioscl. Thromb., 12, 548-553 (1992); Eisenberg, Circulation, 84, 2601-2603 (1991)]. Tukoksia hajottavien aineiden vaikutusta voidaan todellakin parantaa antamalla samanaikaisesti tai edel-

täkäsin trombiini-inhibiittorina toimivaa hepariinia. Samoin verisuonten uudelleensulkeutumista voidaan torjua hajotushoidossa antamalla argatrobaania, hirugeenia tai C-proteiinia [Schneider, Thromb. Res., 64, 677-689 (1990); Yao et al., Am. Physiol., 262 (Heart Circ. Physiol., 31, H 374 - H 379 (1992); Gruber et al., Circulation 84, 2454-2462 (1991)]. Tämän lisäksi tunnettua on se, että sydäninfarktipotilaiden kuolleisuutta voidaan vähentää merkittävästi vertailuryhmään nähden antamalla edeltäkäsin hepariinia ja jälkikäteen prourokinaasia (vertailuryhmälle annetaan prourokinaasia antamalla edeltäkäsin hepariinia) [Tebbe et al., Z. Kardiol., 80, Suppl. 3, 32 (1991)].

Eräs tehokkaimmista trombiini-inhibiittoreista on iilimadosta Hirudo medicinales eristetty hirudiini, jonka karboksipäätteen puoleinen puolikas sitoutuu spesifisesti trombiinin niinkutsuttuun anionien sitoutumiskohtaan. Tietyt, hirudiinimolekyylin aminopäätteen puoleisessa puolikkaassa läsnäolevat aminohapot salpaavat pääsyn trombiinissa olevaan substraatin sitoutumiskoloon [Rydel et al., Science, 249, 277-280 (1990)]. Tämän lisäksi tunnettua on myös se, että myös pienemmät hirudiinijohdannaiset voivat inhiboida trombiinia, joista johdannaisista voidaan mainita erityisesti hirulogimolekyylit, jotka on kuvattu julkaisussa Maraganore et al. Biochemistry, 29, 7095-7101 (1990) [Krstenansky et al., J. Med. Chem. 30, 1688-1691 (1987); Yue et al., Protein Engineering 5, 77-85 (1992).

Hirudiinin käyttö yhdessä plasminogeeniaktivaattorin kanssa tukoksista johtuvien verisuonisairauksien hoitoon on kuvattu eurooppalaisissa patenttihakemuksissa EP 328 957 ja EP 365 468. Hirudiinijohdannaisten käyttö hoitosovellutuksissa yhdessä tukoksia hajottavan aineen kanssa tunnetaan kansainvälisen patenttihakemuksen WO 91/01142 perusteella.

Trombiinia voi myös inhiboida peptidi, joka on peräisin ihmisen trombiinireseptorin aminopäätteen sekvenssistä [Vu et al., Nature, 253, 674-677 (1991)]. Trombiinireseptori sisältää aminopäätteen alueessa trombiinia sitovan sekvenssin ja sen vieressä

trombiinin pilkkomiskohdan. Reseptorin trombiinia sitova alue muistuttaa rakenteeltaan hyvin paljon hirudiinin karboksipäätteen puoleista aluetta. Trombiini aktivoi tätä reseptoria pilkkomalla tämän reseptorisekvenssin. Reseptorin ja trombiinin välisestä affiniteetista johtuen reseptorin eräs kappale, joka käsittää sitoutumisalueen ja muokkautuneen pilkkoutumiskohdan, toimii trombiinin inhibiittorina.

Trombiinia voi myös inhiboida eräs peptidi, joka on peräisin hemadiinin aminohapoista 41-57 [Strube et al., J. Biol. Chem. 268, 8590-8595 (1993)].

Oheinen keksintö perustui tehtävään kehittää vaikuttavia aineita, joilla voidaan käsitellä tukoksten sulkemia verisuonia, ja jotka saavat aikaan erittäin lyhyessä ajassa tukoksen täydellisen hajoamisen ja estävät samalla verisuonen tukkeutumisen uudelleen tukoksen onnistuneen hajotuksen jälkeen. Edelleen näillä vaikuttavilla aineilla tulee voida välttää plasminogeenin systeeminen aktivoituminen.

Keksinnön puitteissa todettiin, että tietyt bifunktionaaliset urokinaasimuunnokset täyttävät tällaisille vaikuttaville aineille asetetut suuret vaatimukset.

Näin ollen oheisen keksinnön kohteena ovat bifunktionaaliset urokinaasimuunnokset, joilla on yleinen kaava I

$$M4\text{-}X_1\text{-}Y_1$$

missä

M4 tarkoittaa kuvion 1 mukaisen glykosyloitumattoman prourokinaasin aminohapposekvenssiä ^{47}Ser - ^{411}Leu,

X_1 tarkoittaa suoraa sidosta M4:n ja Y_1:n välillä tai peptidiä, jolla on sekvenssi

Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

tai

Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

tai
Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly
tai
peptidisekvenssiä, jolla on yleinen kaava II
Ser-X_2-X_3-X_4-X_5-X_6-X_7
missä
X_2 on Pro tai Leu, X_3 on Val tai Pro, X_4 on Lys, Val, Arg, Gly tai Glu, X_5 on Ala, Val, Gly, Leu tai Ile, X_6 on Phe, Trp, Tyr tai Val ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä,
ja

Y_1 on peptidi, jolla on sekvenssi
Y_2-Arg-Pro-Y_3-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y_4
tai
Y_2-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu
tai
Y_2-Arg-Pro-Ser-Ser-Glu-Phe-Glu-Glu-Phe-Glu-Ile-Asp-Glu-Glu-Glu-Lys,
missä
Y_2 on Pro tai Val, Y_3 on Leu tai suora sidos Pro:n ja Gly:n välillä ja Y_4 on Gln tai hydroksyyliryhmä.

Sellaisissa yleisen kaavan I mukaisissa bifunktionaalisissa urokinaasimuunnoksissa, joissa Y_1 on peptidi, jonka sekvenssi on

Y_2-Arg-Pro-Y_3-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y_4

missä Y_2 on Pro tai Val, Y_3 on Leu tai suora sidos Pro:n ja Gly:n välillä ja Y_4 on Gln tai hydroksyyliryhmä, X_1 on edullisesti sellainen yleisen kaavan II mukainen peptidisekvenssi

Ser-X_2-X_3-X_4-X_5-X_6-X_7

missä X_2 on Pro tai Leu, X_3 on Val, X_4 on Lys, Val tai Arg, X_5 on Ala, Val tai Gly, X_6 on Phe, Trp, Tyr tai Val ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä. Näissä bifunktionaalisis-

sa urokinaasimuunnoksissa erityisen edullisia ovat ne yleisen kaavan II mukaiset peptidisekvenssit, joissa X_2 on Pro tai Leu, X_3 on Val, X_4 on Lys tai Val, X_5 on Ala tai Val, X_6 on Phe, Trp tai Tyr ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä, ja erityisesti sellainen sekvenssi, jossa X_7 on suora sidos X_6:n ja Y_1: välillä.

Sellaisissa yleisen kaavan I mukaisissa bifunktionaalisissa urokinaasimuunnoksissa, joissa Y_1 on peptidi, jonka sekvenssi on

Y_2-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu

missä Y_2 on Pro tai Val, X_1 on edullisesti sellainen yleisen kaavan II mukainen peptidisekvenssi

Ser-X_2-X_3-X_4-X_5-X_6-X_7

missä X_2 on Pro tai Leu, X_3 on Val, X_4 on Lys tai Val, X_5 on Ala tai Val, X_6 on Phe tai Trp ja X_7 on suora sidos X_6:n ja Y_1:n välillä.

Tunnettuihin plasminogeeniaktivaattoreihin ja tunnettuihin, plasminogeeniaktivaattorista ja trombiinin inhibiittorista muodostettuihin seoksiin verrattuna keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten tunnusomaisena piirteenä on voimakkaampi fibrinolyyttinen vaikutus yhdistyneenä odottamattoman hyviin trombiinia inhiboiviin ominaisuuksiin. Tämän lisäksi keksinnön mukaiset polypeptidit kuluttavat yllättävällä tavalla selvästi pienempiä määriä plasman fibrinogeenia. Tästä seurauksena olevan, merkittävästi suuremman fibriinispesifisyyden, erityisesti myös plasminogeeniaktivaattorista ja trombiinin inhibiittorista muodostettuihin tunnettuihin seoksiin verrattuna, ansiota on se, että veren hyytymiskyky muuttuu vain vähän ja hallitsemattomien verenvuotojen vaara fibrinogeenin systeemisen hajoamisen mahdollisena komplikaationa on saatu mahdollisimman pieneksi. Näin ollen keksinnön mukaisten urokinaasimuunnosten suuri fibriinispesifisyys tekee mahdolliseksi antamisen boluksena siten, että verenvuotojen vaara on selvästi pienempi kuin

annettaessa tunnettuja, trombolyyttisesti vaikuttavia aineita boluksena.

Yleisen kaavan I mukaiset bifunktionaaliset urokinaasimuunnokset ovat toksikologisesti vaarattomia niin, että niitä voidaan antaa sellaisenaan sopivissa farmaseuttisissa valmisteissa potilaille, jotka kärsivät tukoksien sulkemista verisuonista.

Näin ollen oheisen keksinnön kohteena ovat tukoksia hajottavat aineet (trombolyytit), jotka sisältävät vaikuttavana aineena yleisen kaavan I mukaista bifunktionaalista urokinaasimuunnosta.

Tukoksien sulkemien verisuonten kuten sydäninfarktin, aivoinfarktin, ääreisvaltimoiden akuutin tukkeutumisen, keuhkoveritulpan sekä syvällä jaloissa ja lantioissa sijaitsevien laskimotukosten hoitamiseen tarvitaan 0,1-1 mg/kg keksinnön mukaista polypeptidiä. Näitä bifunktionaalisia urokinaasimuunnoksia voidaan antaa laskimon sisäisesti ja erityisesti bolusinjektiona.

Keksinnön mukaiset trombolyytit sisältävät vähintään yhden bifunktionaalisen urokinaasimuunnoksen lisäksi apuaineita, esimerkiksi kantaja-aineita, liuottimia, laimentimia, väriaineita ja sideaineita. Näiden apuaineiden sekä niiden käytettävien määrien valinta riippuvat siitä, kuinka lääkettä on tarkoitus antaa, eikä tämä valinta aiheuta vaikeuksia alan asiantuntijalle.

Bifunktionaalisten urokinaasimuunnosten valmistaminen tapahtuu geeniteknisillä menetelmillä. Näin ollen keksinnön kohteena ovat myös plasmidit, joita voidaan käyttää yleisen kaavan I mukaisten bifunktionaalisten urokinaasimuunnosten tuottamiseen, ja joiden operoni käsittää säädeltävän promoottorin, ribosomien sitoutumiskohtana toimivan Shine-Dalgarno-sekvenssin, aloituskodonin, yleisen kaavan I mukaisen bifunktionaalisen urokinaasimuunnoksen synteettisen rakennegeenin sekä tämän rakennegeenin jälkeen yhden tai kaksi päättäjää.

Säädeltäväksi promoottoriksi soveltuu erityisesti trp-promoottori tai tac-promoottori. Päättäjänä käytetään edullisesti Tn 10:stä peräisin olevaa trp A-päättäjää ja/tai tet A/orf L-päättäjää.

Keksinnön mukaisten plasmidien säätelyalueessa Shine-Dalgarno-sekvenssin ja aloituskodonin välinen etäisyys on 6-12, edullisesti 8-10 nukleotidiä.

Keksinnön mukaisten plasmidien ilmentäminen toteutetaan Escherichia coli-kannoissa, erityisesti K 12-ryhmään kuuluvista Escherichia coli-kannoissa, joista esimerkkeinä voidaan mainita E. coli K 12 JM 101 (ATCC 33876), E. coli K 12 JM 103 (ATCC 39403), E. coli K 12 JM 105 (DSM 4162) ja E. coli K 12 DH 1 (ATCC 33849). Keksinnön kohteena olevia, yleisen kaavan I mukaisia bifunktionaalisia urokinaasimuunnoksia saadaan bakteerisoluissa suurina saantoina sulkeumakappaleina, joissa proteiini on denaturoituneessa muodossa. Sulkeumakappaleiden eristämisen jälkeen denaturoitu proteiini poimutetaan uudelleen toivottuun tertiäärirakenteeseen proteiinikemiallisilla menetelmillä redoxjärjestelmän avulla.

Esimerkki 1

Keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten valmistaminen, eristäminen ja puhdistaminen

a) Kloonaustoimenpiteet

Ilmentämisplasmidit keksinnön mukaisten polypeptidien tuottamiseksi geeniteknisesti Escherichia coli-bakteerissa valmistettiin sinänsä tunnetulla tavalla. Yksittäisten valmistusvaiheiden eteneminen on esitetty kuvioissa 2 sekä 2a-2p. Plasmidivalmistuksen lähtötuotteina käytettiin plasmideja pBluescript KS II+ (yhtiö Stratagene, Heidelberg), pUC8 (yhtiö Pharmacia, Freiburg) ja pGR201. pGR201 on samanlainen kuin plasmidi pBF160, joka on kuvattu julkaisuissa EP 408 945 ja Appl. Microbiol. Biotechn., 36, 640-649 (1992). Restriktioendonukleaasit BanII, BamHI, ClaI, HindIII, NcoI, NdeI, NheI ja NotI

sekä DNA:ta muokkaavat entsyymit kuten alkalinen fosfataasi, T4-ligaasi, T4-kinaasi ja T7-polymeraasi saatiin yhtiöistä Pharmacia, Stratagene, Boehringer (Mannheim) ja Gibco (Eggenstein). Plasmideissa tapahtuvia muutoksia plasmidien valmistuksen aikana seurattiin restriktioanalyysillä ja DNA-sekvenssimäärityksillä. DNA-sekvenssimääritykset toteutettiin valmistajan ohjeiden mukaisesti joukolla yhtiön Pharmacia reagensseja. Plasmideja valmistettaessa käytettiin erilaisia oligodeoksiribonukleotidejä (oligot), joiden sekvenssit sekä niiden tunnukset on esitetty seuraavassa taulukossa 1.

Taulukko 1

Oligotunnus	Oligosekvenssi (kirjoitettu 5'-->3')
O105	TATGAGCAAAACTTGCTACGAAGGTAACGGTCACTTCTA CCGTGGTAAGGCTTCTACCGACAC
O106	CATGGTGTCGGTAGAAGCCTTACCACGGTAGAAGTGACC GTTACCTTCGTAGCAAGTTTTGCTCA
O220	CGGTTAAGGCTTTCCCGAGGCCTGGTGGTGGTGGTAACG GTGACTTCGAAGAAATCCCGGAAGAGTACCTGTGATAGG ATCAA
O221	CTAGTTGATCCTATCACAGGTACTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAG CCTTAACCGGGCT
O222	CGCCGAGCCCGCCGAGCCCGCCGGGTGGTTTCCCGAGGC CTGGTGGTGGTGGTAACGGTGACTTCGAAGAAATCCCGG AAGAGTACCTGTGATAGGATCAA
O223	CTAGTTGATCCTATCACAGGTACTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAC CACCCGGCGGGCTCGGCGGGCTCGGCGGGCT
O224	CGCCGGGTGGTTTCCCGAGGCCTGGTGGTGGTGGTAACG GTGACTTCGAAGAAATCCCGGAAGAGTACCTGTGATAGG ATCAA
O225	CTAGTTGATCCTATCACAGGTACTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAC CACCCGGCGGGCT
O226	CGCCGAGCCCGCCGAGCCCGCCGGGTGGTTTCGGTCCGA GGCCTGGTGGTGGTGGTAACGGTGACTTCGAAGAAATCC CGGAAGAGTACCTGTGATAGGATCAA
O227	CTAGTTGATCCTATCACAGGTACTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGACCGA AACCACCCGGCGGGCTCGGCGGGCTCGGCGGGCT
O265	CACCCGGCGGAGACGGCGGGCTCAGAGCCAGACCGTTTT CTTCTTTGGTGTGAGAACG
O281	CGTCCGGGTGGTGGTGGTAACGGTGACTTCGAAGAAATC CCGGAAGAATACCTGTAAG
O282	GATCCGTTCTCACACCAAAGAAGAAAACGGTCTGGCTCT GAGCCCGCCGTCTCCGCCGGGTGGTTTCCCG

O283	CTAGCTTACAGGTATTCTTCCGGGATTTCTTCGAAGTCA CCGTTACCACCACCACCCGGACGCGGGAAAC
O329	AAGAAATCCCGGAAGAATACCTGCAATAAG
O330	CGGTTAAGGCTTGGGGACCGCGGCCGCTGGGTGGTGGTG GTAACGGTGACTTCG
O331	ACCACCACCCAGCGGCCGCGGTCCCCAAGCCTTAACCGG GCT
O332	CTAGCTTATTGCAGGTATTCTTCCGGGATTTCTTCGAAG TCACCGTTACC
O333	CGGTTAAGGCTTTCGGACCGC
O334	GGCCGCGGTCCGAAAGCCTTAACCGGGCT
O335	CGGTTCGGGCTTTCGGTCCGC
O336	GGCCGCGGACCGAAAGCCCGAACCGGGCT
O337	CGGTTAAGGCTTACGGACCGC
O338	GGCCGCGGTCCGTAAGCCTTAACCGGGCT
O339	CGGTTGTTGCTTTCGGTCCGC
O340	GGCCGCGGACCGAAAGCAACAACCGGGCT
O341	CGGTTCGGGCTTTCCCGC
O342	GGCCGCGGGAAAGCCCGAACCGGGCT
O343	CGGTTAAGGCTTACCCGC
O344	GGCCGCGGGTAAGCCTTAACCGGGCT
O347	CGGTTGTTGCTTTCCCGC
O348	GGCCGCGGGAAAGCAACAACCGGGCT
O381	CGGTTAAGGCTTGGCCGC
O383	GGCCGCGGCCAAGCCTTAACCGGGCT
O384	CGGTTAAGGCTTTCCCGC
O385	GGCCGCGGGAAAGCCTTAACCGGGCT
O386	CGGTTGTAGTTTTCCCGC
O387	CGGTTGAAGTTTTCCCGC
O388	GGCCGCACTACAACTACAACCGGGCT
O389	CGGTTGTAGTTGTAGTGC
O390	GGCCGCGGGAAAACTTCAACCGGGCT
O391	GGCCGCGGGAAAACTACAACCGGGCT
O392	CTAGCTTATTCGTTTTTTTCTTCGTCTTCCCAGAACGGT TCGTATTTGTCGTTCGGGTTCCGCAGCAGGAAC

0393	GGCCGTTCCTGCTGCGGAACCCGAACGACAAATACGAAC CGTTCTGGGAAGACGAAGAAAAAAACGAATAAG
0453	TGGTTAAAGCTTTCCCGC
0454	GGCCGCGGGAAAGCTTTAACCAGGCT
0455	TGGTTGTTGCTTTCCCGC
0456	GGCCGCGGGAAAGCAACAACCAGGCT
0465	GGCCGCGGGAACAGAGCCAGACCGTTTTCTTCTTTGGTG TGAGAACG
0466	GATCCGTTCTCACACCAAAGAAGAAAACGGTCTGGCTCT GTTCCCGC
0467	CGGTTAAGGCTTTCCCGCGGCCGTTCCTGCTGCGGAAC
0468	TTTGTCGTTCGGGTTCCGCAGCAGGAACGGCCGCGGGAA AGCCTTAACCGGGCT
0469	CTAGCTTATTCGTTTTTTTCTTCGTCTTCCCAGAACGGT TCGTA
0470	CCGAACGACAAATACGAACCGTTCTGGGAAGACGAAGAA AAAAACGAATAAG

Nämä oligodeoksiribonukleotidit valmistettiin detrityloidussa muodossa 0,1 μmoolin määrinä yhtiön Applied Biosystems (Weiterstadt) syntetisaattorilla (malli 391) valmistajan ohjeiden mukaan, käyttäen b-syanoetyylillä suojattuja di-isopropyyliaminofosfoamidiitteja. Kulloinkin 100 pmol oligodeoksiribonukleotidia fosforyloitiin liuoksessa, joka sisälsi 50 mM tri(hydroksimetyyli)aminometaani/HCl-puskuria (Tris/HCl), 10 ml magnesiumkloridia ja 5 mM ditiotreitolia, pH-arvossa 7,5, yhdellä entsyymiyksiköllä T4-kinaasia adenosiinitrifosfaatin 10 mM:n läsnäollessa, ja muunnettiin sitten tässä samassa puskurissa kaksijuosteisiksi DNA-molekyyleiksi. Saadut synteettiset, kaksijuosteiset DNA-molekyylit puhdistettiin geelielektroforeettisesti polyakryyliamidigeelillä (5 % polyakryyliamidia), minkä jälkeen niitä käytettiin ligaatiossa vastaavalla tavalla esikäsiteltyjen plasmidien kanssa. Plasmidien esikäsittely pilkkomalla ne restriktioentsyymeillä, eristämällä vastaavat restriktiokappaleet ja defosforyloimalla 5'-päät, tätä seuraava ligaatio ja transformointi E. coli K12 JM103-kantaan sekä kaikki muut geenitekniset toimenpiteet toteutettiin sinänsä tunnetulla tavalla, joka on kuvattu teoksessa Sambrook et al., "Molecular Cloning: A Laboratory Manual", 2. painos, Cold Spring Harbor Laboratory Press, Cold Spring Harbour, USA, 1989.

b) Kestoviljelmien valmistus ja fermentointi

Yhdistelmätekniset ilmentämisplasmidit pSJ69, pSJ76, pSJ77, pSJ78, pSJ79, pSJ81, pSJ83, pSJ90, pSJ91, pSJ92, pSJ93, pSJ94, pSJ95, pSJ101, pSJ102, pSJ103, pSJ104, pSJ105, pSJ106, pSJ109, pSJ111, pSJ114 ja pSJ113 istutettiin E. coli K12 JM103-kantaan (ATCC 39403) ja levitettiin Standard-I-agarille (150 mg/l ampisilliinia) (Sambrook et al., "Molecular Cloning: A Laboratory Manual"). Kulloinkin yhtä yksittäistä, kustakin transformantista saatua pesäkettä viljeltiin Standard-I-väliaineessa (pH 7,0; 150 mg/l ampisilliinia) 20 °C:ssa siten, että optiseksi tiheydeksi (OD) aallonpituudella 578 nm saatiin 1, ja tästä viljelmästä otettiin viisi 2 ml:n annosta kestoviljelmiksi, joihin lisättiin dimetyylisulfoksidia (DMSO) (loppupitoisuus 7,5 %) ja jotka pakastettiin nestemäisellä typellä ja säilytettiin -70

°C:ssa. Bifunktionaalisten urokinaasimuunnosten saamiseksi kulloinkin 1 ml kutakin kestoviljelmää suspendoitiin 20 ml:aan standard-I-alustaan (pH 7,0; 150 mg/l ampisilliinia) ja viljeltiin 37 °C:ssa OD-arvoon 1 aallonpituudella 578 nm määritettynä.

Lopuksi saadun viljelmän koko määrä suspendoitiin 1 litraan Standard-I-alustaa (pH 7,0; 150 mg/l ampisilliinia) ja fermentoitiin ravistelukolveissa 37 °C:ssa. Indusoiminen tapahtui lisäämällä 2 ml indoliakryylihappoliuosta (60 mg 2 ml:ssa etanolia) OD-arvossa 0,5-1 aallonpituudella 578 nm määritettynä.

c) Ilmentymisen testaaminen

Ilmentymisvoimakkuuden (Ploug-yksikköä/OD millilitrassa) määrittämiseksi viljelmästä otettiin välittömästi ennen indusointia ja kulloinkin yhden tunnin välein indusoinnin jälkeen (yhteensä kuusi tuntia) solumäärä, joka vastasi yhtä ml:aa solususpensiota, jonka OD-arvo oli 1 aallonpituudella 578 nm, ja solut erotettiin sentrifugoimalla. Sentrifugoimalla erotetut solut hajotettiin lysotsyymillä [1 mg lysotsyymiä millilitrassa 50 mM Tris/HCl-puskuria, pH 8,0, 50 mM etyleenidiamiinitetraetikkahappoa (EDTA) ja 15 til.-% sakkaroosia]. Hajotetut solut saatettiin liukoiseen muotoon 4-5 M guanidiniumhydrokloridiliuoksessa, jota laimennettiin sitten siten, että pitoisuudeksi saatiin 1,2 M glunidiniumhydrokloridia, minkä jälkeen uudelleenpoimuttaminen toteutettiin 2-5 tunnin aikana lisäämällä pelkistävää ainetta (glutationia tai kysteiiniä) [Winkler et al., Biochemistry, 25, 4041-4045 (1986)]. Saadut yksiketjuiset bifunktionaaliset urokinaasimuunnokset muunnettiin plasmiinia lisäämällä vastaaviksi kaksiketjuisiksi urokinaasimuunnoksiksi, joiden aktiivisuus määritettiin kromogeenisella pyro-Glu-Gly-Arg-p-nitroanilidi-substraatilla, jota vain kaksiketjuiset aktiiviset urokinaasit pilkkovat. Keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten aktivointi plasmiinilla toteutettiin 50 mM Tris/HCl-puskurilla, joka sisälsi 12 mM natriumkloridia, 0,02 % Tween 80, pH-arvossa 7,4 ja 37 °C:n lämpötilassa. Bifunktionaalisen urokinaasimuunnoksen ja plasmiinin

välinen suhde oli noin 100-1500:1 molaarisuussuhteena ja niiden välinen suhde oli noin 8000-36 000:1 entsyymiyksikköinä laskien. Koeinkubointi tapahtui 50 mM Tris/HCl-puskurissa, joka sisälsi 38 mM natriumkloridia, pH-arvossa 8,8, siten, että läsnä oli 0,36 μM aprotiniinia (plasmiinin inhiboimiseksi) sekä 0,27 mM pyro-Glu-Gly-Arg-p-nitroanilid-substraattia 37 °C:n läsnäollessa. Bifunktionaalisen urokinaasimuunnoksen pitoisuudesta riippuen reaktio pysäytettiin 5-60 minuuttia kestäneen inkubooinnin jälkeen lisäämällä 50-prosenttista etikkahappoa ja ekstinkio mitattiin aallonpituudella 405 nm. Substraatin valmistajan (Kabi Vitrum, Schweden) ohjeiden mukaisesti tällä tavalla meneteltäessä ja aallonpituutta 405 nm käytettäessä 0,05/min oleva ekstinktiomuutos vastaa urokinaasiaktiivisuutta, joka on 25 Ploug-yksikköä millilitrassa koeliuosta. Keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten spesifiset aktiivisuudet olivat alueella 120 000 - 155 000 Ploug-yksikköä puhdistetun proteiinin yhdessä milligrammassa. Liuosten proteiinipitoisuus määritettiin yhtiön Pierce BCA-määrityksellä.

d) Eristäminen ja puhdistaminen

Kun indusoinnista oli kulunut 5-6 tuntia, kohdassa 1b) kuvatuissa olosuhteissa toteutettu fermentointi lopetettiin (optinen tiheys OD 5-6 aallonpituudella 578 nm). Solut erotettiin sentrifugoimalla. Solukasauma suspendoitiin uudestaan 200 ml:aan vettä ja ne hajotettiin suurpainehomogenisaattorissa. Hajote semtrifugoitiin uudestaan ja sakka, joka sisälsi kaiken yksiketjuisen bifunktionaalisen urokinaasimuunnoksen, liuotettiin 500 ml:aan 5 M guanidiniumhydrokloridia, joka sisälsi 40 mM kysteiiniä, 1 mM EDTA, ja jonka pH oli 8,0, ja tämä liuos laimennettiin 2000 ml:lla 25 mM Tris/HCl-puskuria, jonka pH-arvo oli 9,0. Uudelleenpoimutusreaktio oli edennyt loppuun noin 12 tunnin kuluttua.

Saadut bifunktionaaliset urokinaasimuunnokset sidottiin täydellisesti silikageeliin lisäämällä 8 g silikageeliä ja sekoittamalla 2 tunnin ajan. Tällä tavalla "ladattu" silikageeli erotettiin ja pestiin asetaattipuskurilla (pH 4,0). Urokinaasi-

muunnokset eluoitiin 0,5 M tetrametyyliammoniumkloridilla (TMAC) 0,1 M asetaattipuskurissa (pH 4). Urokinaasimuunnokset saatiin puhtaassa muodossa kahden kromatografisen erotuksen jälkeen (kuparikelaattipylväs ja kationinvaihtaja). N-päätteen sekvenssianalyysillä voitiin osoittaa toisaalta yksiketjuisuus ja toisaalta toivottu aminopäätteen sekvenssi. Yksittäisissä muunnoksissa muuttuneen karboksipäätteen puoleisen alueen proteiinikemiallinen karakterisointi toteutettiin proteiinin muokatun CNBr-pilkkomisen (liuotettuna 1 ml:aan 90 % muurahaishappoa ja 1 ml:aan heptafluorivoihappoa) jälkeen pilkkomalla peptidiketju tryptofaanijäännösten jälkeen. Ennen sekvenssianalyysiä karboksipäätteen puoleinen peptidi erotettiin ja puhdistettiin HPLC-menetelmällä (High Pressure Liquid Chromatography).

Yhdelläkään eristetyllä ja taulukossa 2 esitetyllä bifunktionaalisella urokinaasimuunnoksella ei todettu suorassa aktiivisuusmäärityksessä, jossa käytettiin urokinaasin kromogeenista substraattia, minkäänlaista tai niillä todettiin vain hyvin vähäistä aktiivisuutta (alle 1200 Ploug-yksikköä milligrammassa puhdistettua proteiinia). Vasta plasmiinilla pilkkomisen jälkeen [olosuhteet on kuvattu kappaleessa 1c)] entsyymiaktiivisuus saatiin alueelle 120 000-155 000 Ploug-yksikköä milligrammassa puhdistettua proteiinia. Näin ollen kaikki urokinaasimuunnokset ilmentyivät E. coli K12 JM103-kannassa yksiketjuisena proteiinina.

Taulukko 2

Keksinnön mukaiset bifunktionaaliset urokinaasimuunnokset (bU), joilla on yleinen kaava I: M4-X_1-Y_1

bU	X_1	Y_1
M11	Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe	A^1 missä Y_2=Pro, Y_3=suora sidos Pro:n ja Gly:n välillä, Y_4=OH
M12	Ser-Pro-Val-Lys-Ala-Phe	A^1 missä Y_2=Pro, Y_3=suora sidos Pro:n ja Gly:n välillä, Y_4=OH
M13	Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe	A^1 missä Y_2=Pro, Y_3=suora sidos Pro:n ja Gly:n välillä, Y_4=OH
M14	Ser-Pro-Pro-Gly-Gly-Phe	A^1 missä Y_2=Pro, Y_3=suora sidos Pro:n ja Gly:n välillä, Y_4=OH
M15	Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly	A^1 missä Y_2=Pro, Y_3=suora sidos Pro:n ja Gly:n välillä, Y_4=OH
M16	Ser-Pro-Val-Lys-Ala-Trp-Gly	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M17	Ser-Pro-Val-Lys-Ala-Phe-Gly	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M18	Ser-Pro-Val-Arg-Ala-Phe-Gly	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M19	Ser-Pro-Val-Lys-Ala-Tyr-Gly	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M20	Ser-Pro-Val-Val-Ala-Phe-Gly	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M21	Ser-Pro-Val-Arg-Ala-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M22	Ser-Pro-Val-Lys-Ala-Tyr	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M23	Ser-Pro-Val-Val-Ala-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln

(Taulukko 2, jatkoa)

bU	X_1	Y_1
M24	Ser-Pro-Val-Lys-Ala-Trp	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M25	Ser-Pro-Val-Lys-Ala-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M26	Ser-Pro-Val-Val-Val-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M27	Ser-Pro-Val-Glu-Val-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M28	Ser-Pro-Val-Val-Val-Val	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M29	Ser-Pro-Val-Val-Ala-Phe	B^2 missä Y_2=Pro
M30	Ser-Leu-Val-Val-Ala-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M31	Ser-Leu-Val-Lys-Ala-Phe	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln
M32	Ser-Pro-Val-Lys-Ala-Phe	B^2 missä Y_2=Pro
M33	suora sidos M4:n ja Y_1:n välillä	A^1 missä Y_2=Pro, Y_3=Leu, Y_4=Gln

[1] A = Y_2-Arg-Pro-Y_3-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y_4

[2] B = Y_2-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu

Esimerkki 2

Farmakologiset tutkimukset

Trombiinia inhiboivan vaikutuksen määrittäminen

Keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten inhibiittoriaktiivisuus määritettiin mittaamalla trombiiniaika siten, että 200 µl:aan sitraatilla käsitellyn ihmisplasman 1:10-laimennosta veronaalipuskurissa sekoitettiin 50 µl trom-

biiniliuosta (0,2 yksikköä) ja 50 µl vesiliuosta, joka sisälsi 0,5-50 µg bifunktionaalista urokinaasimuunnosta. Sen jälkeen mitattiin aika, joka kului fibriinikasautuman muodostumiseen. Taulukossa 3 on esitetty mitatut estotekijät, jotka osoittavat trombiiniajan pidentymisen olosuhteissa, joissa on läsnä 10 µg jotakin keksinnön mukaista bifunktionaalista urokinaasimuunnosta. Samoin määritettiin pitoisuudesta riippuva trombiiniajan piteneminen, ja se on esitetty graafisesti kuviossa 3 bifunktionaalisten urokinaasimuunnosten M12, M23, M29, M32, M33 tapauksessa ja vertailun vuoksi M4:n tapauksessa. Toisin kuin keksinnön mukaisten bifunktionaalisten urokinaasimuunnosten tapauksessa, M4, eli kuviossa 1 esitetyn glykolysoitumattoman prourokinaasin aminohapposekvenssi ^{47}Ser-^{411}Leu, glykosyloitumaton prourokinaasi (saruplaasi) ja LUK eivät kyenneet pidentämään hyytymiseen kuluvaa aikaa edes 1 mg:n suuruisena annoksena.

Taulukko 3

Yleisen kaavan I: M4-X_1-Y_1 mukaisten ja keksinnön kohteena olevien bifunktionaalisten urokinaasimuunnosten kyky pidentää trombiiniaikaa

Bifunktionaalinen urokinaasimuunnos	Estotekijä[1]
M11	1,8
M12	4,6
M13	1,7
M14	1,8
M15	2,5
M16	3,2
M17	3,1
M18	2,9
M19	2,0
M20	2,2
M21	2,3
M22	3,7
M23	5,3
M24	6,2
M25	2,9
M26	3,2
M27	2,0
M28	2,1
M29	2,6
M30	3,4
M31	2,0
M32	3,0
M33	2,0

[1] määritettynä 10 μg:n suuruisen proteiinimäärän vaikutuksena

Estotekijä = Inhibiittorin läsnäollessa saadun trombiiniajan ja inhibiittorin puuttuessa saadun trombiiniajan osamäärä

Bifunktionaalisten urokinaasimuunnosten M12 ja M23 farmakologiset ominaisuudet eläinkokeessa

Bifunktionaalisten urokinaasimuunnosten M12 ja M23 vaikutus valtimosuonissa esiintyvien tukosten hajottamiseen selvitettiin farmakologisessa in vivo-mallissa saruplaasiin (glykosyloitumaton prourokinaasi) verrattuna. Tätä tarkoitusta varten nukutettujen kaniinien tilapäisesti eristettyyn, noin 1 cm:n pituiseen reisivaltimon lohkoon injektoitiin paikallisesti sivuhaaran kautta trombiinia ja ^{125}J-merkittyä ihmisen fibrinogeenia. Tämän seurauksena muodostui tukos, joka johti verisuonen täydelliseen tukkeutumiseen. Muodostuneen tukoksen koko määritettiin ihmisfibriinin mukana sisältyneen radioaktiivisuuden avulla ulkoista gammalaskuria käyttäen. Veren läpivirtausta mitattiin sähkömagneettisesti ja tukoksen radioaktiivisuutta määritettiin jatkuvasti kokeen koko keston ajan. Fibrinolyyttinen vaikutus kvantitoitiin siis sekä tukoksen sulkeman verisuonen läpäisevyyden (reperfuusion) että tukokseen sisältyneen, radioaktiivisesti merkityn fibriinin häviämisen perusteella. Verinäytteitä otettiin ennen keksinnön mukaisten, bifunktionaalisten urokinaasimuunnosten antamista sekä sen jälkeen, kun urokinaasimuunnosten antamisesta oli kulunut 30, 60 ja 90 minuuttia, ja näistä verinäytteistä määritettiin fibrinogeenin pitoisuus plasmassa. Kulloinkin 6 mg/kg muunnosta M12 ja M23 sekä saruplaasia annettiin laskimon sisäisesti bolusinjektiona. Koska muunnoksilla M12 ja M23 on lisäksi vielä hyytymistä torjuvaa vaikutusta, päinvastoin kuin saruplaasilla, niin neljännessä koeryhmässä saruplaasia annettiin yhdessä hyytymistä torjuvana aineena toimineen hepariinin kanssa (150 U/kg i.v.-bolus). Kukin koeryhmä käsitti 6 eläintä.

90 minuutin pituisen kokeen aikana merkityn tukosfibriinin trombolyysi oli 46 ± 11 % M12:n tapauksessa, 43 ± 12 % M23:n tapauksessa, 22 ± 5 % saruplaasin tapauksessa ja 39 ± 15 % saruplaasi-hepariini-yhdistelmän tapauksessa. M12:n ja M23:n antaminen boluksena johti kaikkien kuuden eläimen tapauksessa tukoksen sulkeman verisuonen avautumiseen; saruplaasia antamalla verisuoni saatiin läpäisemään verta 5 eläimessä kuudesta ja

sekä saruplaasia että hepariinia antamalla verisuoni saatiin läpäisemään verta 4 eläimessä kuudesta. Verisuonen läpäisevyys oli parhaimmillaan (prosentteina lähtöarvosta) 95 ± 10 % M12:n tapauksessa ja 82 ± 9 % M23:n tapauksessa, ja verisuonen läpäisevyys erosi merkittävästi 43 ± 12 % olevasta suurimmasta läpäisevyydestä saruplaasin tapauksessa. 58 ± 8 % oleva suurin reperfuusiovirtaus, johon päästiin käsittelemällä seka saruplaasilla että hepariinilla, oli toisaalta M12:lla ja M23:lla sekä toisaalta saruplaasilla saatujen tulosten välissä, eikä eroa merkittävästi näistä kummastakaan vertailuarvosta. Fibrinolyyttinen kokonaisvaikutus määritettiin reperfuusiovirtauksen pinta-alana (prosentteina lähtövirtauksesta) 90 minuutin pituisen kokeen ajan. Tämä kokonaisvaikutus oli 4502 ± 1127 % · min M12:n tapauksessa ja 4270 ± 885 % · min M23:n tapauksessa, ja se oli kummankin keksinnön mukaisen urokinaasimuunnoksen tapauksessa merkittävästi suurempi kuin saruplaasilla saatu arvo, 1519 ± 643 % · min. Saruplaasi-hepariini-yhdistelmäkäsittelyllä saatiin mitatuksi kokonaisvaikutukseksi 2217 ± 761 % · min., joka ei ole merkittävästi parempi kuin pelkällä saruplaasikäsittelyllä saatu arvo, ja joka on selvästi pienempi kuin M12:lla ja M23:lla saatu tulos. Tulokset on koottu taulukkoon 4.

Taulukko 4

Trombolyyttinen vaikutus i.v.-boluksena antamisen jälkeen; reisivaltimon tukos, nukutettu kaniini

Polypeptidi	Annos	^{125}J-fibrinolyysi, %	Suurin reperfuusiovirtaus (% lähtöarvosta)	Kumulatiivinen reperfuusiovirtaus (%·min)
M12	6 mg/kg	46 ± 11	95 ± 10*	4502 ± 1127
M23	6 mg/kg	43 ± 12	82 ± 9*	4270 ± 885*
Saruplaasi	6 mg/kg	22 ± 5	43 ± 12	1519 ± 643
Saruplaasi + Hepariini	6 mg/kg 150 U/kg	39 ± 15	58 ± 8	2217 ± 761

* p <0,05 vs. saruplaasi

Yllättäen todettiin, että sekä M12:n boluksena tapahtuneen antamisen jälkeen että M23:n boluksena tapahtuneen antamisen jälkeen fibrinogeenin pitoisuus plasmassa pieneni merkittävästi vähemmän kuin saruplaasin boluksena tapahtuneen antamisen jälkeen. Tulokset on koottu taulukkoon 5.

Taulukko 5

M12:n ja M23:n boluksena tapahtuneen antamisen vaikutus plasmassa todetun fibrinogeenipitoisuuden pienenemiseen, verrattuna saruplaasiin sekä ilman hepariinia että yhdessä hepariinin kanssa; nukutettu kaniini

Polypeptidi	Annos	Plasman fibrinogeenipitoisuuden pieneneminen (% muutos lähtöarvoon verrattuna)		
		Aika antamisen jälkeen		
		30 min	60 min	90 min
M12	6 mg/kg	-19 ± 9	-20 ± 9*	-19 ± 9*
M23	6 mg/kg	-20 ± 11*	-21 ± 11*	-20 ± 11*
Saruplaasi	6 mg/kg	-64 ± 7	-66 ± 6	-67 ± 6
Saruplaasi + hepariini	6 mg/kg 150 U/kg	e.m.[1]	-46 ± 8	-45 ± 9

* p <0,05 vs. saruplaasi
[1] e.m. = ei määritetty

Näistä tuloksista nähdään, että keksinnön mukaiset bifunktionaaliset urokinaasijohdannaiset M12 ja M23 kykenevät liuottamaan verisuonen täydellisesti tukkivia valtimotukoksia ja saamaan aikaan verivirtauksen tukkeutuneissa verisuonissa. Tähän vaikutukseen päästiin antamalla M12:ta tai M23:ta yhden kerran boluksena hepariinilla käsittelemättömille eläimille. Yllättäen tähän M12:n ja M23:n saruplaasiin verrattuna voimakkaampaan fibrinolyyttiseen vaikutukseen liittyi plasman fibrinogeenin pienempää kulutusta. Tämä tarkoittaa sitä, että M12:lla ja M23:lla on merkittävästi suurempi fibriinispesifisyys saruplaasiin verrattuna. Plasman fibrinogeenin parempi säilyminen M12:n

ja M23:n läsnäollessa saruplaasiin verrattuna tarkoittaa sitä, että veren hyytymiskyky pysyy parempana, joten fibrinogeenin systeemisen hajoamisen mahdollisina komplikaatioina esiintyvien hallitsemattomien verenvuotojen vaara on pienempi. Niinpä verenvuodon tyrehdyttämiseen liittyvien sivuvaikutusten vaaraa ajatellen M12:ta ja M23:ta voidaan pitää turvallisempina kuin saruplaasia.

SEKV. ID NRO 1

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M24 (pSJ101) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAG GCT TGG CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Lys Ala Trp Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 2

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M25 (pSJ102) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                     85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
                 100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
             115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
         130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                     165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAG GCT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Lys Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 3

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M26 (pSJ103) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 GTT GTA GTT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
Val Val Val Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Glu Glu Ile Pro Glu Glu Tyr Leu Gln
385 393

SEKV. ID NRO 4

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M27 (pSJ104) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 GTT GAA GTT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
Val Glu Val Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Glu Glu Ile Pro Glu Glu Tyr Leu Gln
385 390

SEKV. ID NRO 5

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M28 (pSJ105) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT GTA GTT GTA GTG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Val Val Val Val Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 6

PITUUS: 1191 BP, 396 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M29 (pSJ106) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT GTT GCT TTC CCG CGG CCG TTC CTG CTG CGG AAC CCG AAC GAC AAA
     Val Val Ala Phe Pro Arg Pro Phe Leu Leu Arg Asn Pro Asn Asp Lys
         370                 375                 380

1152 TAC GAA CCG TTC TGG GAA GAC GAA GAA AAA AAC GAA TAA
     Tyr Glu Pro Phe Trp Glu Asp Glu Glu Lys Asn Glu
     385                 390                 395
```

SEKV. ID NRO 7

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M20 (pSJ109) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
             115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
         130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC GTG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Leu
             355                 360                 365

1104 GTT GTT GCT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Val Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 8

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M31 (pSJ111) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC GTG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Leu
             355                 360                 365

1104 GTT AAA GCT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Lys Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 9

PITUUS: 1167 BP, 388 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M33 (pSJ113) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
                 100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
             115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
         130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG TTC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Phe Pro
             355                 360                 365

1104 CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC GAA GAA ATC CCG GAA
     Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu
         370                 375                 380

1152 GAA TAC CTG CAA TAA
     Glu Tyr Leu Gln
     385
```

SEKV. ID NRO 10

PITUUS: 1167 BP, 388 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M32 (pSJ114) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 GTT AAG GCT TTC CCG CGG CCG TTC CTG CTG CGG AAC CCG AAC GAC AAA
Val Lys Ala Phe Pro Arg Pro Phe Leu Leu Arg Asn Pro Asn Asp Lys
370 375 380

1152 TAC GAA CCG TTC TGG GAA GAC GAA GAA AAA AAC GAA TAA
Tyr Glu Pro Phe Trp Glu Asp Glu Glu Lys Asn Glu
385 390 395

SEKV. ID NRO 11

PITUUS: 1101 BP, 366 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M4 (pSJ41) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu
355 360 365

SEKV. ID NRO 12

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M11 (pSJ69) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 CCG TCT CCG CCG GGT GGT TTC CCG CGT CCG GGT GGT GGT GGT AAC GGT
     Pro Ser Pro Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly
         370                 375                 380

1152 GAC TTC GAA GAA ATC CCG GAA GAA TAC CTG TAA
     Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu
     385                 390
```

SEKV. ID NRO 13

PITUUS: 1176 BP, 391 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M12 (pSJ76) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAG GCT TTC CCG AGG CCT GGT GGT GGT GGT AAC GGT GAC TTC GAA
     Val Lys Ala Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly Asp Phe Glu
         370                 375                 380

1152 GAA ATC CCG GAA GAA TAC CTG TGA
     Glu Ile Pro Glu Glu Tyr Leu
     385                 390
```

SEKV. ID NRO 14

PITUUS: 1194 BP, 397 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M13 (pSJ77) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 CCG AGC CCG CCG AGC CCG CCG GGT GGT TTC CCG AGG CCT GGT GGT GGT
     Pro Ser Pro Pro Ser Pro Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly
         370                 375                 380

1152 GGT AAC GGT GAC TTC GAA GAA ATC CCG GAA GAG TAC CTG TGA
     Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu
     385                 390                 395
```

SEKV. ID NRO 15

PITUUS: 1176 BP, 391 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M14 (pSJ78) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 CCG GGT GGT TTC CCG AGG CCT GGT GGT GGT GGT AAC GGT GAC TTC GAA
     Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly Asp Phe Glu
         370                 375                 380

1152 GAA ATC CCG GAA GAG TAC CTG TGA
     Glu Ile Pro Glu Glu Tyr Leu
     385                 390
```

SEKV. ID NRO 16

PITUUS: 1197 BP, 398 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M15 (pSJ79) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 CCG AGC CCG CCG AGC CCG CCG GGT GGT TTC GGT CCG AGG CCT GGT GGT
Pro Ser Pro Pro Ser Pro Pro Gly Gly Phe Gly Pro Arg Pro Gly Gly
370 375 380

1152 GGT GGT AAC GGT GAC TTC GAA GAA ATC CCG GAA GAG TAC CTG TGA
Gly Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu
385 390 395

SEKV. ID NRO 17

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M16 (pSJ81) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAG GCT TGG GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC
     Val Lys Ala Trp Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
         370                 375                 380

1152 TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 18

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M17 (pSJ83) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAG GCT TTC GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC
     Val Lys Ala Phe Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
         370                 375                 380

1152 TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 19

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M18 (pSJ90) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
                 100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
             115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
         130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT CGG GCT TTC GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC
     Val Arg Ala Phe Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
         370                 375                 380

1152 TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 20

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M19 (pSJ91) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
                 100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
             115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
         130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 GTT AAG GCT TAC GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC
Val Lys Ala Tyr Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
370 375 380

1152 TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln
385 390

SEKV. ID NRO 21

PITUUS: 1185 BP, 394 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M20 (pSJ92) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT GTT GCT TTC GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC
     Val Val Ala Phe Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
         370                 375                 380

1152 TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 22

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M21 (pSJ93) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT CGG GCT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Arg Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 23

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M22 (pSJ94) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

```
528  TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
     Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
                 180                 185                 190

576  AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
     Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
             195                 200                 205

624  AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
     Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
         210                 215                 220

672  CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
     Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
     225                 230                 235                 240

720  GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
     Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
                 245                 250                 255

768  AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
     Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
                 260                 265                 270

816  AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
     Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
             275                 280                 285

864  GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
     Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
         290                 295                 300

912  GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
     Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
     305                 310                 315                 320

960  GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
     Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
                 325                 330                 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
     Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
                 340                 345                 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
     Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
             355                 360                 365

1104 GTT AAA GCT TAC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
     Val Lys Ala Tyr Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
         370                 375                 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
     Glu Glu Ile Pro Glu Glu Tyr Leu Gln
     385                 390
```

SEKV. ID NRO 24

PITUUS: 1182 BP, 393 AS

LAJI: Nukleotidisekvenssi ja vastaava aminohappo-sekvenssi

TOPOLOGIA: Lineaarinen

MOLEKYYLILAJI: Proteiinin M23 (pSJ95) synteettinen geeni

ERITYISTÄ: N-pään metioniini lohkeaa proteolyyttisesti E. colissa

```
1    ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
     Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
     1               5                  10                  15

48   GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
     Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
                 20                  25                  30

96   ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG
     Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
             35                  40                  45

144  CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT
     Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
         50                  55                  60

192  CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA
     Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
     65                  70                  75                  80

240  TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG
     Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
                 85                  90                  95

288  GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT
     Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
             100                 105                 110

336  AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC
     Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
         115                 120                 125

384  GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC
     Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
     130                 135                 140

432  GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC
     Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
     145                 150                 155                 160

480  TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT
     Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
                 165                 170                 175
```

528 TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

576 AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

624 AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

672 CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

720 GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

768 AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

816 AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

864 GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

912 GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

960 GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

1008 CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

1056 TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

1104 GTT GTT GCT TTC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC
Val Val Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

1152 GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Glu Glu Ile Pro Glu Glu Tyr Leu Gln
385 390

Patenttivaatimukset

1. Menetelmä bifunktionaalisten urokinaasimuunnosten, joilla on yleinen kaava I

$$M4\text{-}X_1\text{-}Y_1$$

valmistamiseksi, jossa kaavassa

M4 tarkoittaa kuvion 1 mukaisen glykosyloitumattoman prourokinaasin aminohapposekvenssiä $^{4}7Ser$ - $^{4}11Leu$,

X_1 tarkoittaa suoraa sidosta M4:n ja Y_1:n välillä tai peptidiä, jolla on sekvenssi
Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe
tai
Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe
tai
Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly
tai
peptidisekvenssiä, jolla on yleinen kaava II
Ser-X_2-X_3-X_4-X_5-X_6-X_7
jossa
X_2 on Pro tai Leu, X_3 on Val tai Pro, X_4 on Lys, Val, Arg, Gly tai Glu, X_5 on Ala, Val, Gly, Leu tai Ile, X_6 on Phe, Trp, Tyr tai Val ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä,
ja

Y_1 on peptidi, jolla on sekvenssi
Y_2-Arg-Pro-Y_3-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y_4
tai
Y_2-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu
tai
Y_2-Arg-Pro-Ser-Ser-Glu-Phe-Glu-Glu-Phe-Glu-Ile-Asp-Glu-Glu-Glu-Lys,
jossa
Y_2 on Pro tai Val, Y_3 on Leu tai suora sidos Pro:n ja Gly:n välillä ja Y_4 on Gln tai hydroksyyliryhmä,

t u n n e t t u siitä, plasmideista pBlueskript KS II+, pUC 8 ja GR 201 valmistetaan kuvioiden 2 ja 2a-2p mukaises-

ti plasmideja, joiden operonissa on säädettävä promoottori, ribosomin sitoutumiskohtana toimiva Shine-Dalgarno-sekvenssi, aloituskodoni, synteettinen rakennegeeni bifunktionaalista urokinaasimuunnosta varten ja rakennegeenistä myötävirtaan 1 tai 2 terminaattoria, seuraavaksi saadulla plasmidilla transformoidaan Escherichia coli -kanta itsessään tunnetulla tavalla, joka indusoi rakennegeenin ilmentymisen, erotetaan bifunktionaalisen urokinaasimuunnoksen muodostunut esiasteproteiini väliaineesta ja hajotetuista bakteerisoluista, tehdään esiasteproteiini liukoisaksi ja sitten se poimutetaan uudelleen redox-järjestelmän avulla bifunktionaaliseksi urokinaasimuunnokseksi.

2. Patenttivaatimuksen 1 mukainen menetelmä, t u n n e t t u siitä, että valmistetaan yleisen kaavan I mukainen urokinaasimuunnos, jossa Y_1 on peptidi, jonka sekvenssi on

Y_2-Arg-Pro-Y_3-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y_4

3. Patenttivaatimuksen 1 mukainen menetelmä, t u n n e t t u siitä, että että valmistetaan yleisen kaavan I mukainen urokinaasimuunnos, jossa Y_1 on peptidi, jonka sekvenssi on

Y_2-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu.

4. Patenttivaatimuksen 1 ja/tai patenttivaatimuksen 2 mukainen menetelmä, t u n n e t t u siitä, että valmistetaan yleisen kaavan I mukainen urokinaasimuunnos, jossa X_1 on yleisen kaavan II mukainen peptidisekvenssi, jossa X_2 on Pro tai Leu, X_3 on Val, X_4 on Lys, Val tai Arg, X_5 on Ala, Val tai Gly, X_6 on Phe, Trp, Tyr tai Val ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä.

5. Patenttivaatimuksen 4 mukainen menetelmä, t u n n e t t u siitä, että X_4 on Lys tai Val, X_5 on Ala tai Val, X_6 on Phe, Trp tai Tyr ja X_7 on Gly tai suora sidos X_6:n ja Y_1:n välillä.

6. Patenttivaatimuksen 4 ja/tai patenttivaatimuksen 5 mukainen menetelmä, t u n n e t t u siitä, että X_7 on suora sidos X_6:n ja Y_1:n välillä.

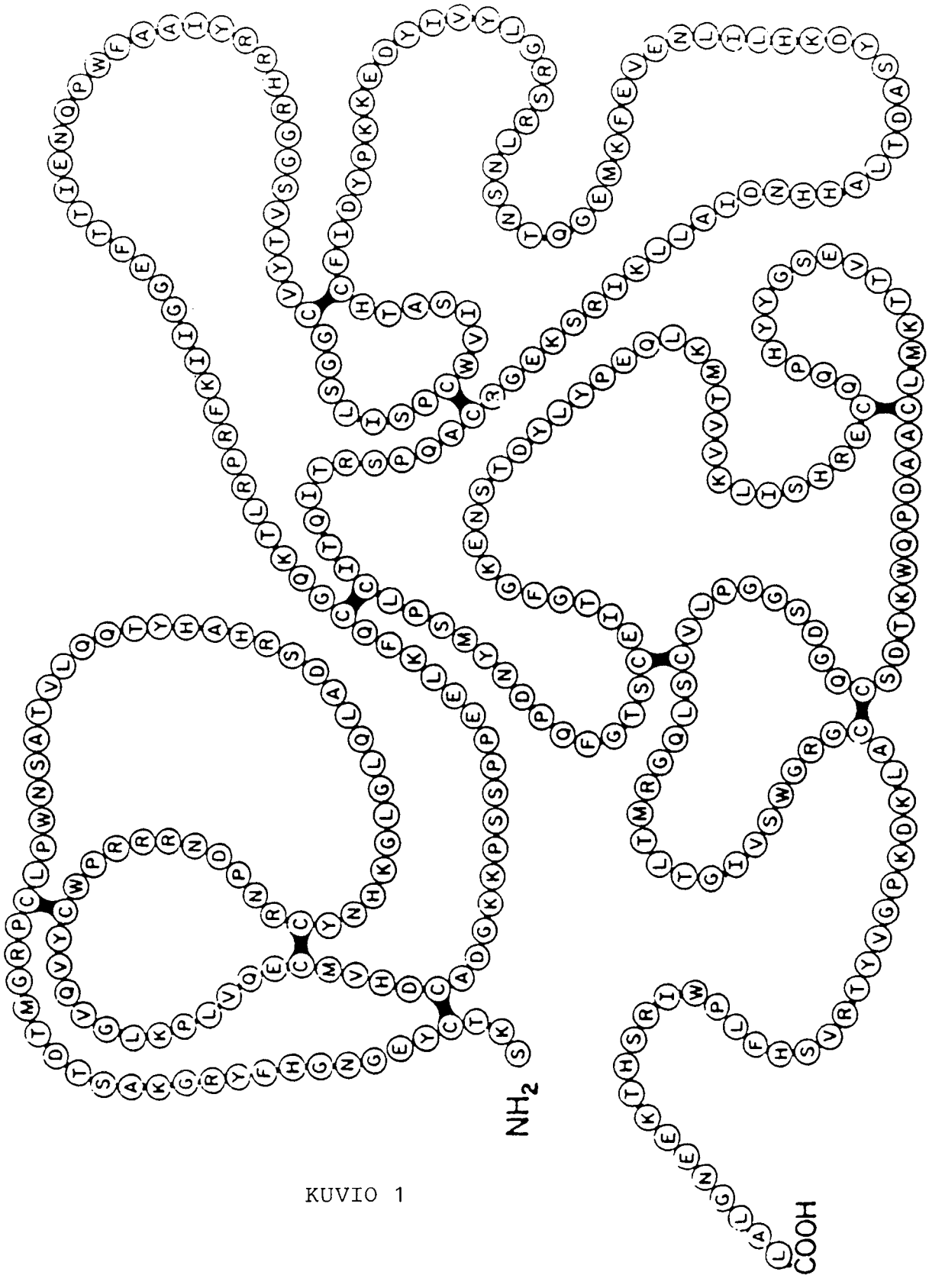
7. Patenttivaatimuksen 1 ja/tai patenttivaatimuksen 3 mukainen menetelmä, t u n n e t t u siitä, että X_1 on yleisen kaavan II mukainen peptidisekvenssi, jossa X_2 on Pro tai Leu, X_3 on Val, X_4 on Lys tai Val, X_5 on Ala tai Val, X_6 on Phe tai Trp ja X_7 on suora sidos X_6:n ja Y_1:n välillä.

8. Plasmidit, joita voidaan käyttää patenttivaatimusten 1-7 mukaisen bifunktionaalisen urokinaasimuunnoksen tuottamiseen, t u n n e t u t siitä, että ne on valittu plasmidien pSJ 69, pSJ 76, pSJ 77, pSJ 78, pSJ 79, pSJ 81, pSJ 83, pSJ 90, pSJ 91, pSJ 92, pSJ 93, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 104, pSJ 105, pSJ 106, pSJ 109, pSJ 111, pSJ 114 ja pSJ 113 joukosta.

9. Patenttivaatimuksen 8 mukaiset plasmidit, t u n n e t u t siitä, että ne on valittu plasmidien pSJ 76, pSJ 81, pSJ 83, pSJ 90, pSJ 91, pSJ 92, pSJ 93, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 105, pSJ 106, pSJ 109, pSJ 111 ja pSJ114 joukosta.

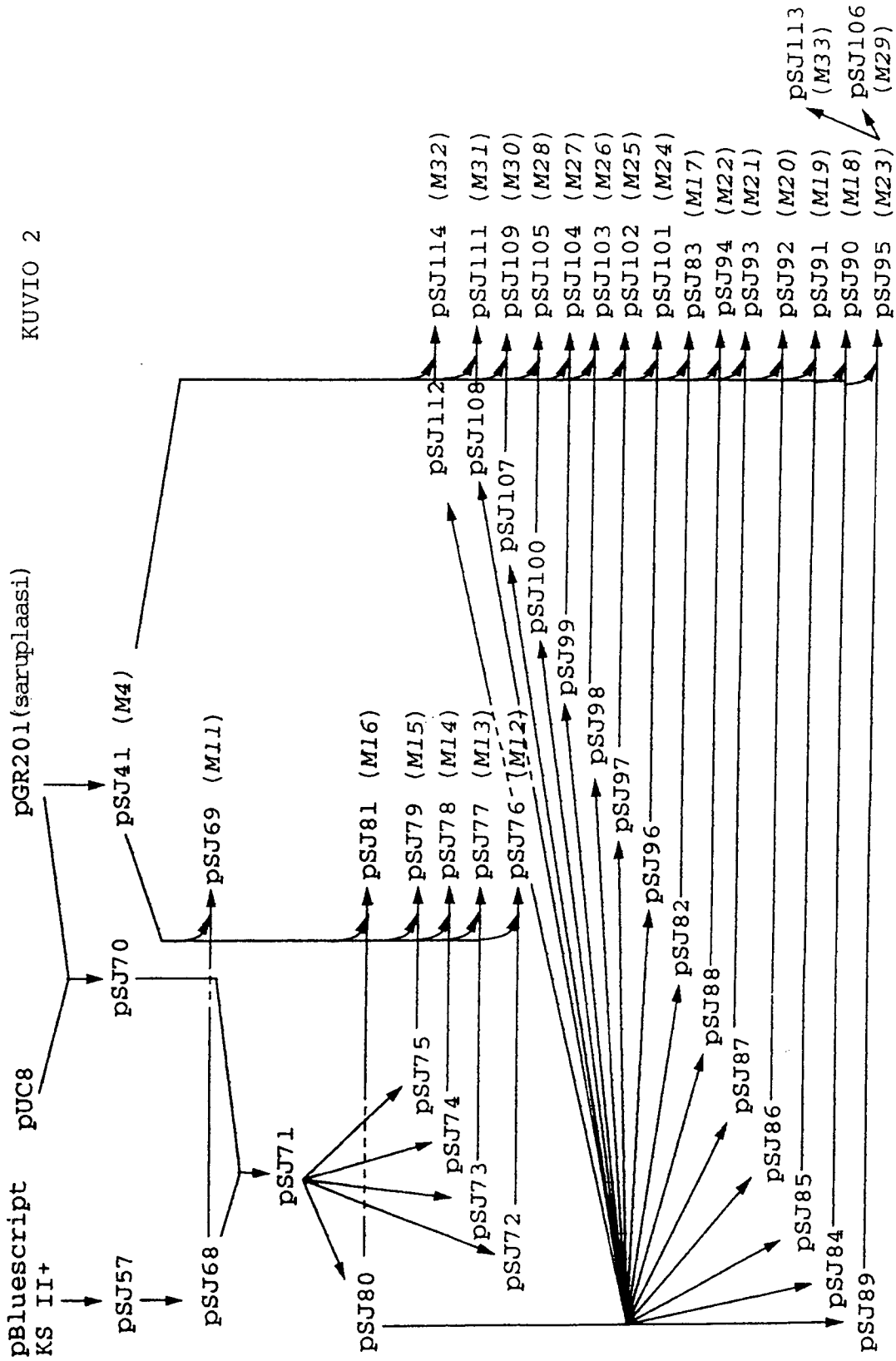
10. Patenttivaatimuksen 8 ja/tai patenttivaatimuksen 9 mukaiset plasmidit, t u n n e t u t siitä, että ne on valittu plasmidien pSJ 76, pSJ 81, pSJ 83, pSJ 91, pSJ 92, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 106, pSJ 109, pSJ 111 ja pSJ 114 joukosta.

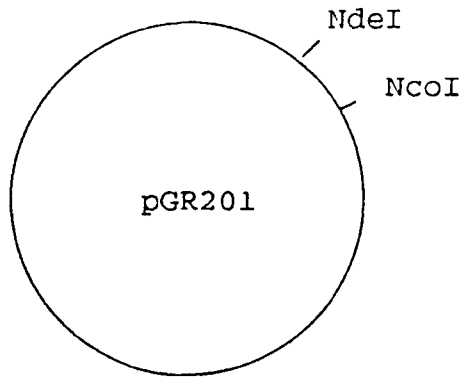
11. Yhden tai useamman patenttivaatimuksen 8-10 mukaiset plasmidit, t u n n e t u t siitä, että ne on valittu plasmidien pSJ 76, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 106, pSJ 109, pSJ 111 ja pSJ 114.



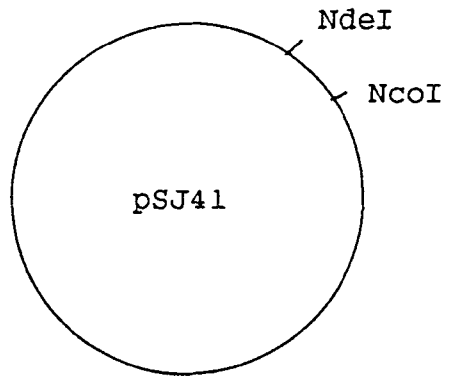
KUVIO 1

KUVIO 2

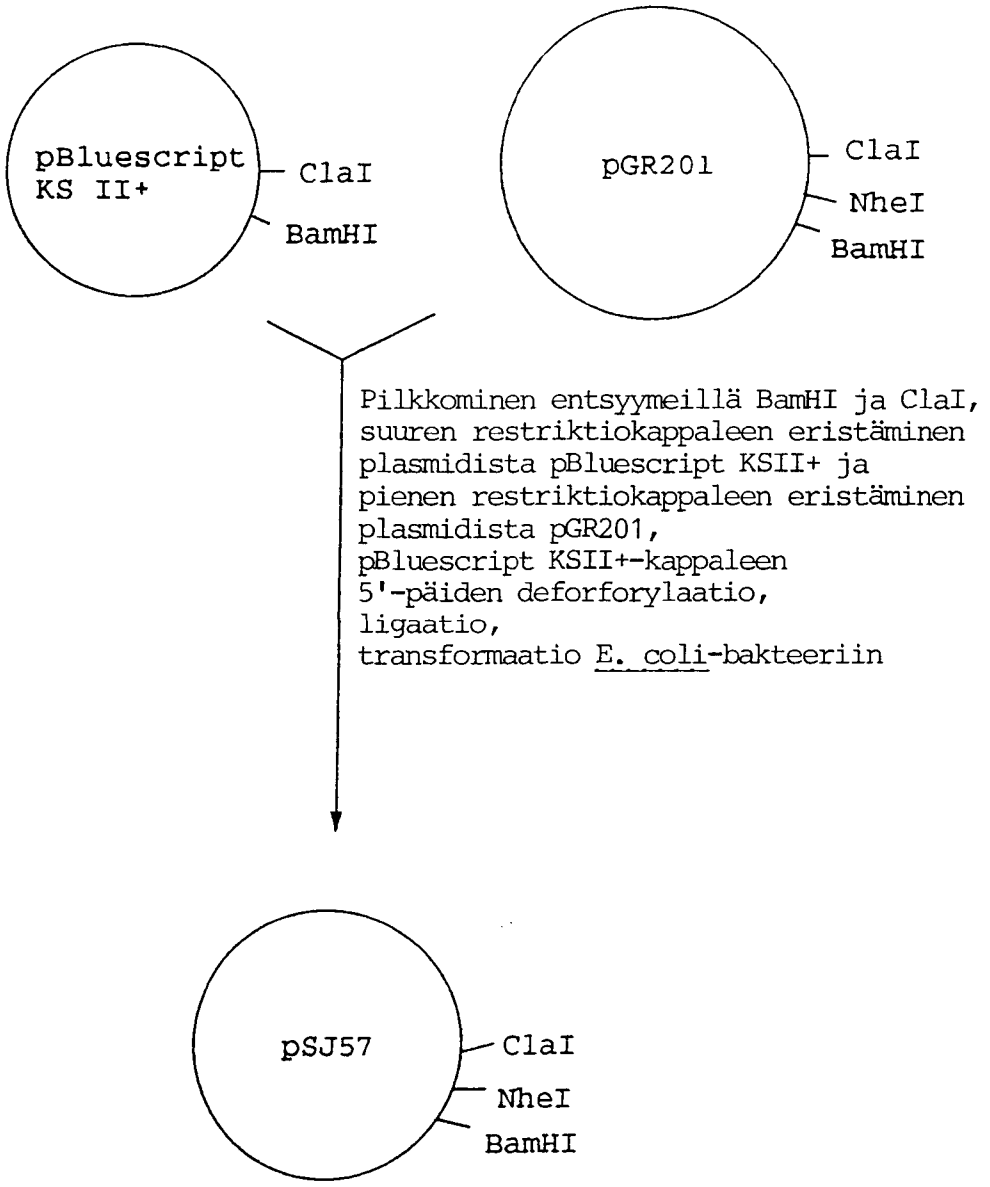

pBluescript KS II+
pSJ57
pSJ68
pUC8
pGR201(saruplaasi)
pSJ70
pSJ41 (M4)
pSJ69 (M11)
pSJ71
pSJ80
pSJ81 (M16)
pSJ75
pSJ79 (M15)
pSJ74
pSJ78 (M14)
pSJ73
pSJ77 (M13)
pSJ72
pSJ76 (M12)
pSJ112
pSJ114 (M32)
pSJ108
pSJ111 (M31)
pSJ107
pSJ109 (M30)
pSJ100
pSJ105 (M28)
pSJ99
pSJ104 (M27)
pSJ98
pSJ103 (M26)
pSJ97
pSJ102 (M25)
pSJ96
pSJ101 (M24)
pSJ82
pSJ83 (M17)
pSJ88
pSJ94 (M22)
pSJ87
pSJ93 (M21)
pSJ86
pSJ92 (M20)
pSJ85
pSJ91 (M19)
pSJ84
pSJ90 (M18)
pSJ89
pSJ95 (M23)
pSJ113 (M33)
pSJ106 (M29)



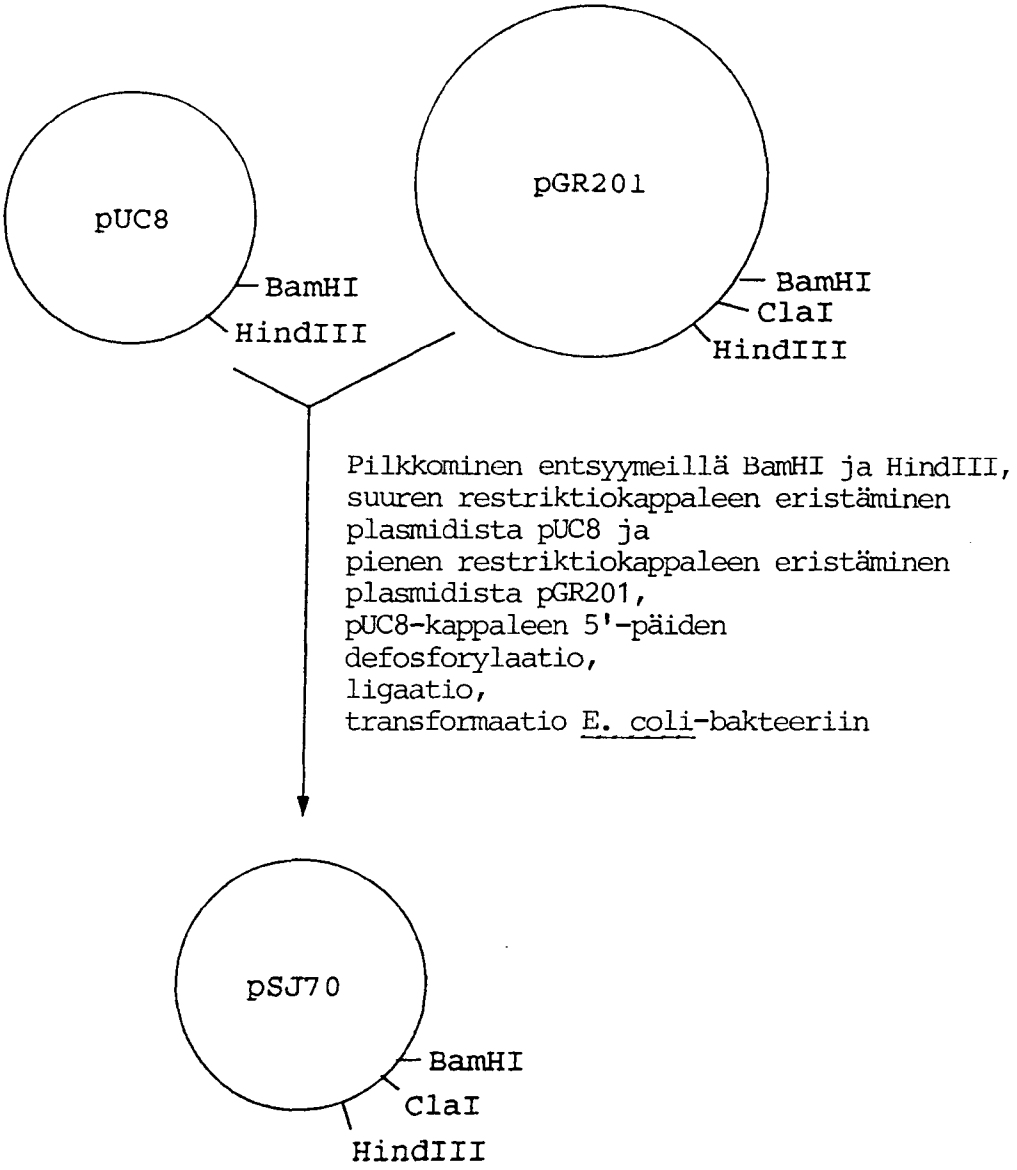
Pilkkominen entsyymeillä NdeI ja NcoI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligosta 0105 ka 0106
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin



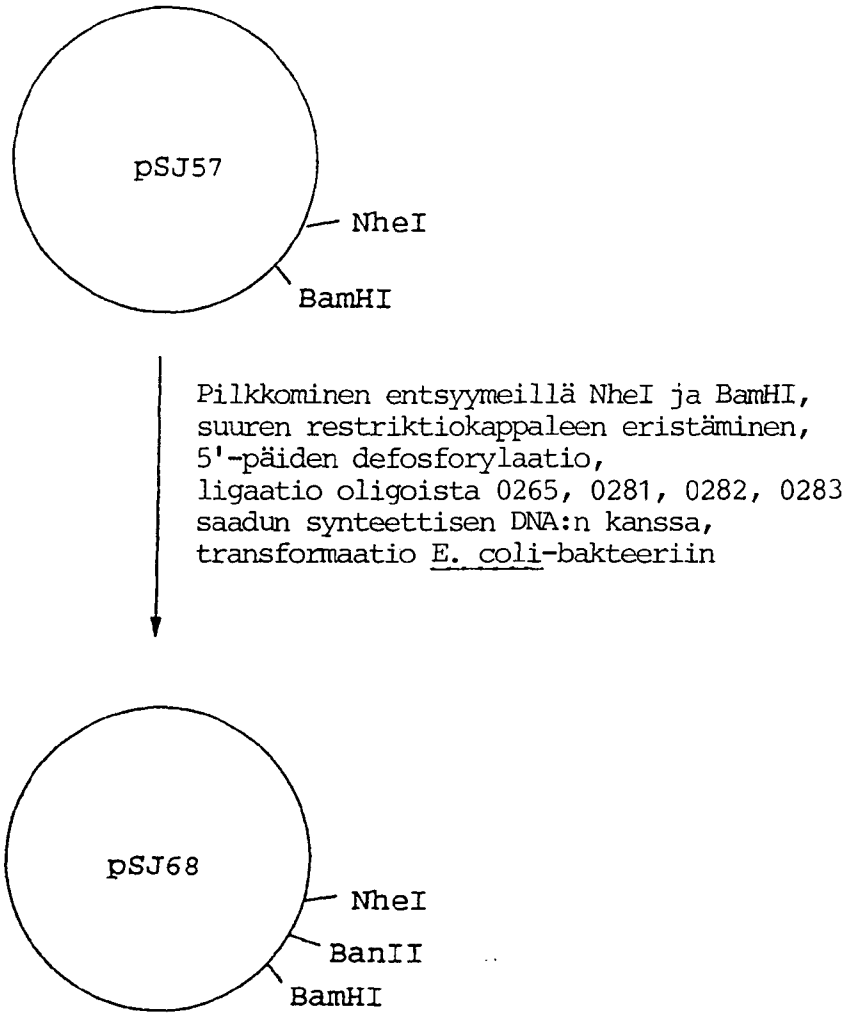
KUVIO 2a


pBluescript
KS II+
ClaI
BamHI
pGR201
ClaI
NheI
BamHI
Pilkkominen entsyymeillä BamHI ja ClaI,
suuren restriktiokappaleen eristäminen
plasmidista pBluescript KSII+ ja
pienen restriktiokappaleen eristäminen
plasmidista pGR201,
pBluescript KSII+-kappaleen
5'-päiden deforforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ57
ClaI
NheI
BamHI

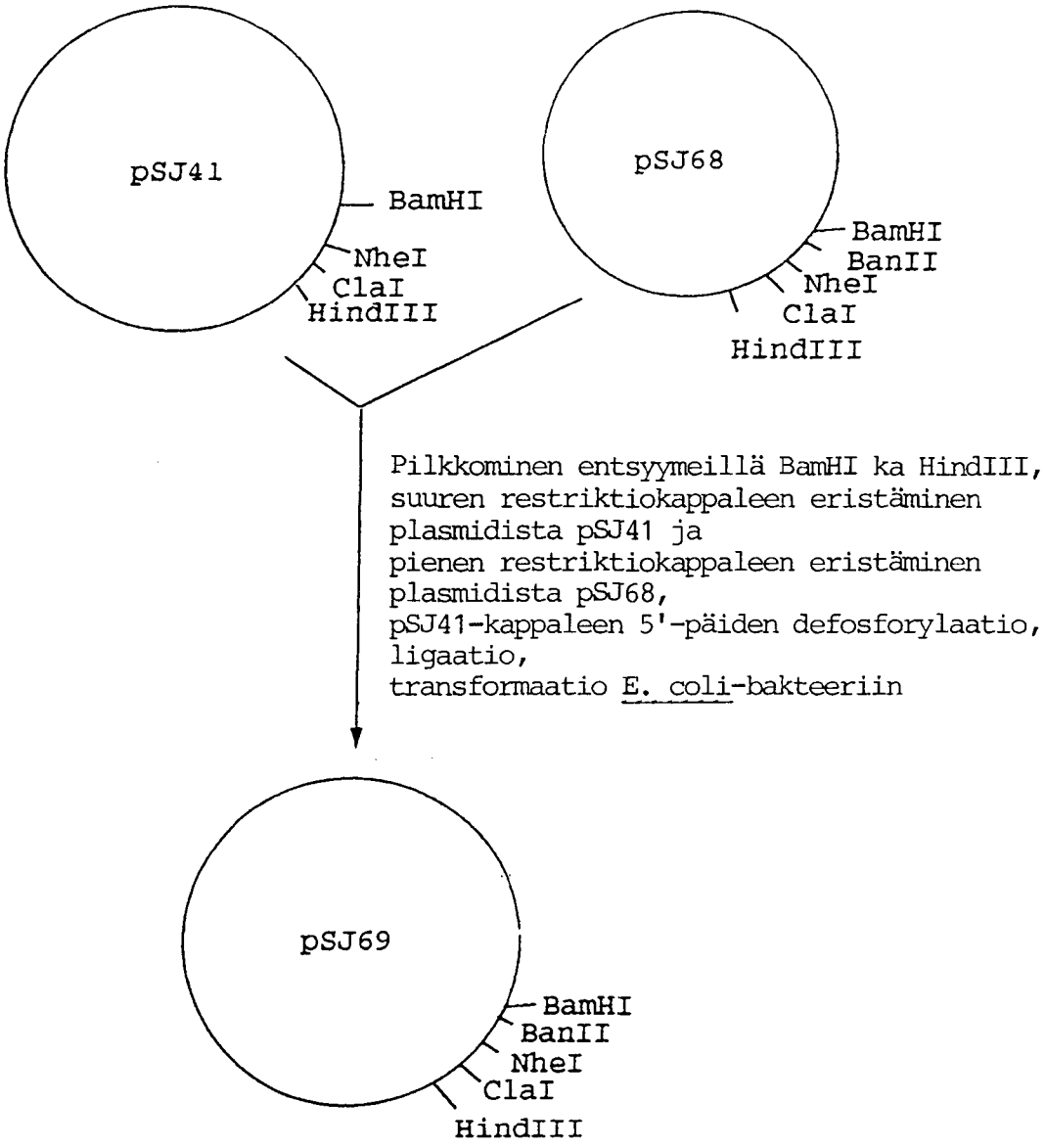
KUVIO 2b


pUC8
BamHI
HindIII
pGR201
BamHI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pUC8 ja
pienen restriktiokappaleen eristäminen
plasmidista pGR201,
pUC8-kappaleen 5'-päiden
defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ70
BamHI
ClaI
HindIII

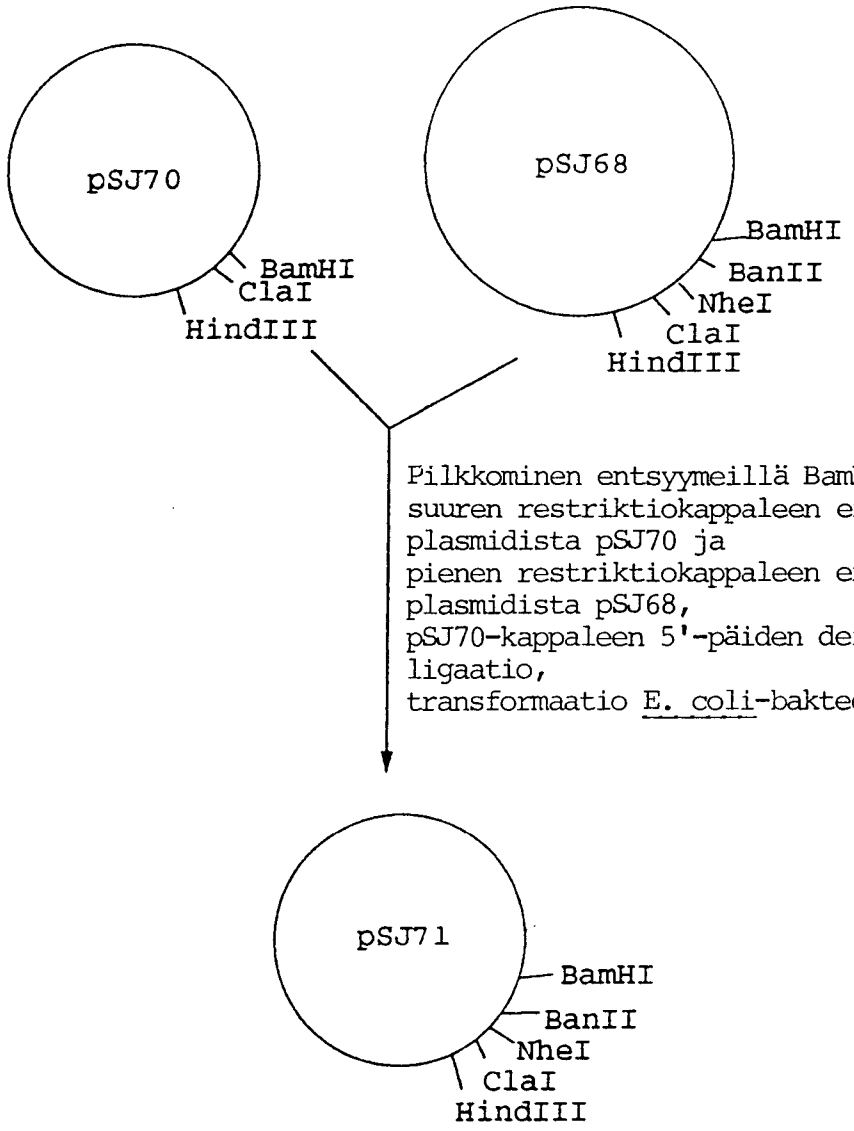
KUVIO 2c


pSJ57
NheI
BamHI
Pilkkominen entsyymeillä NheI ja BamHI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligoista 0265, 0281, 0282, 0283
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin
pSJ68
NheI
BanII
BamHI

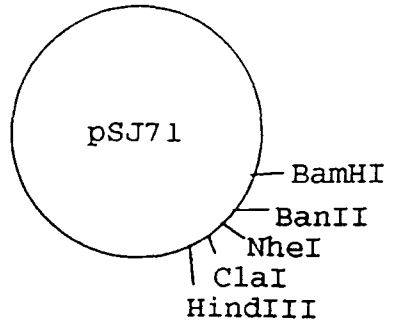
KUVIO 2d


pSJ41
BamHI
NheI
ClaI
HindIII
pSJ68
BamHI
BanII
NheI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ka HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pSJ41 ja
pienen restriktiokappaleen eristäminen
plasmidista pSJ68,
pSJ41-kappaleen 5'-päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ69
BamHI
BanII
NheI
ClaI
HindIII

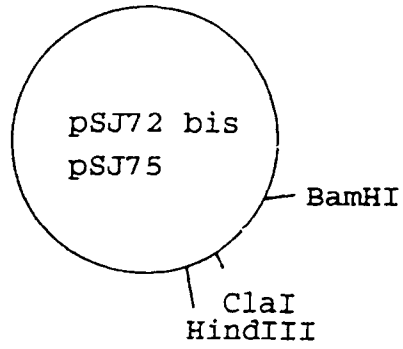
KUVIO 2e


pSJ70
BamHI
ClaI
HindIII
pSJ68
BamHI
BanII
NheI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja ClaI,
suuren restriktiokappaleen eristäminen
plasmidista pSJ70 ja
pienen restriktiokappaleen eristäminen
plasmidista pSJ68,
pSJ70-kappaleen 5'-päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ71
BamHI
BanII
NheI
ClaI
HindIII

KUVIO 2f

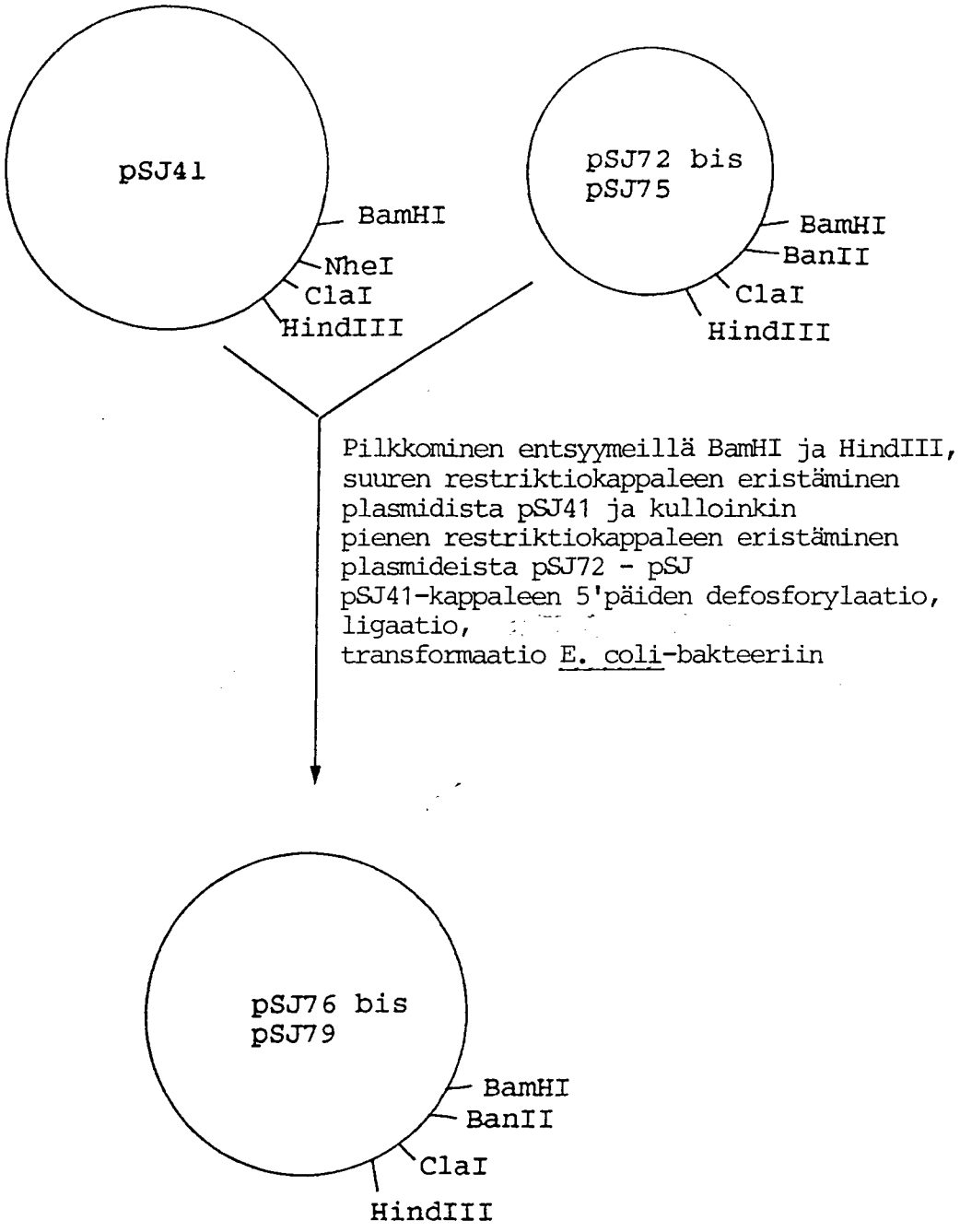


Pilkkominen entsyymeillä BanII ja NheI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin

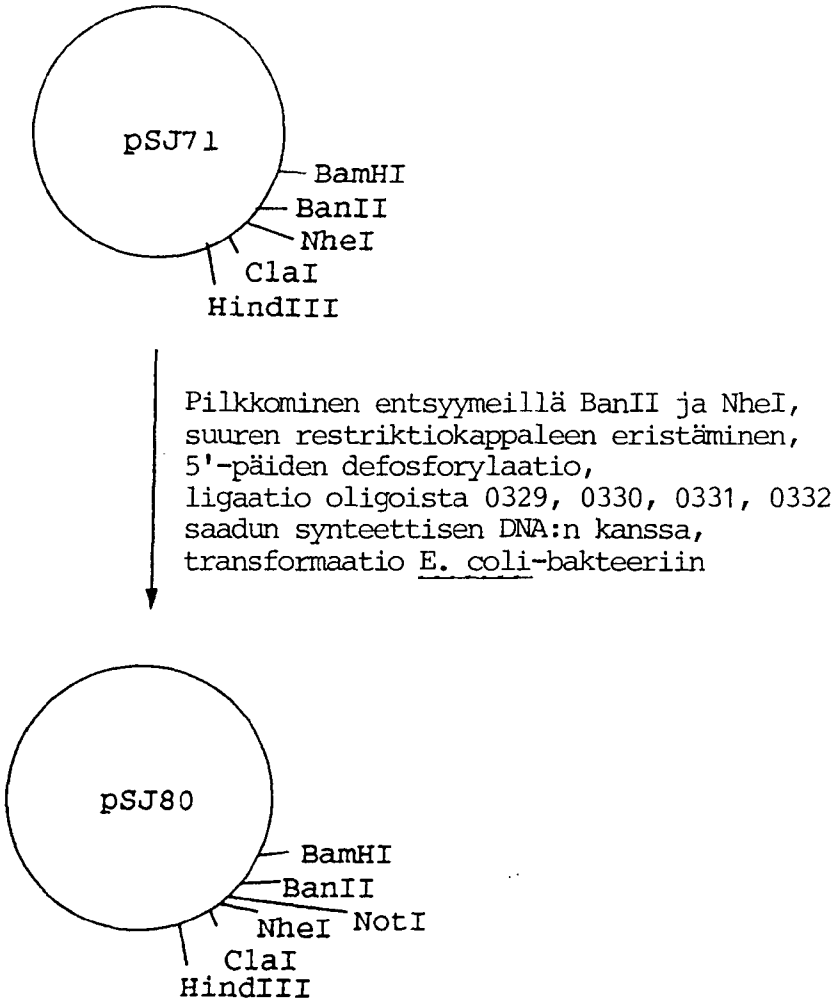


plasmidi	oligot
pSJ72	0220, 0221
pSJ73	0222, 0223
pSJ74	0224, 0225
pSJ75	0226, 0227

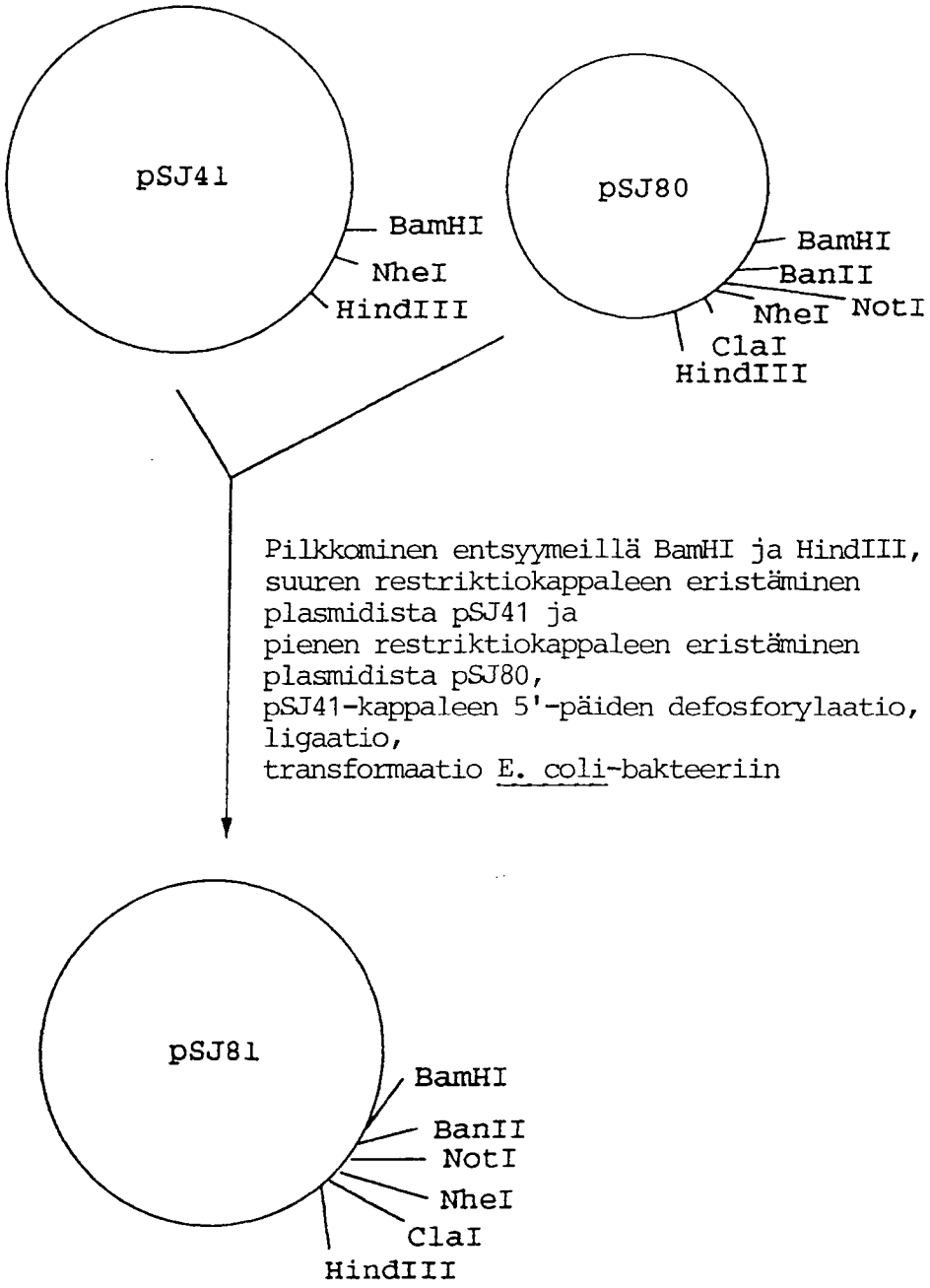
KUVIO 2g


pSJ41
BamHI
NheI
ClaI
HindIII
pSJ72 bis
pSJ75
BamHI
BanII
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pSJ41 ja kulloinkin
pienen restriktiokappaleen eristäminen
plasmideista pSJ72 - pSJ
pSJ41-kappaleen 5'päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ76 bis
pSJ79
BamHI
BanII
ClaI
HindIII

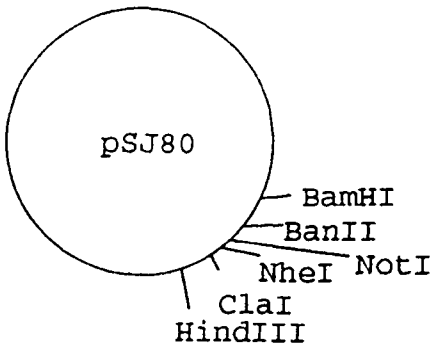
KUVIO 2h


pSJ71
BamHI
BanII
NheI
ClaI
HindIII
Pilkkominen entsyymeillä BanII ja NheI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligoista 0329, 0330, 0331, 0332
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin
pSJ80
BamHI
BanII
NheI
NotI
ClaI
HindIII

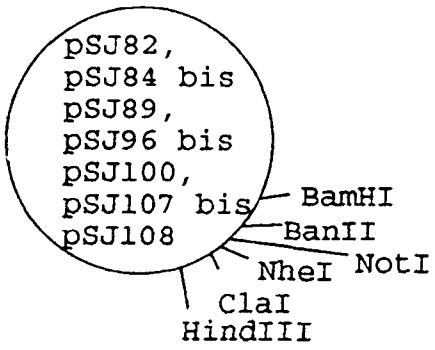
KUVIO 2i


pSJ41
BamHI
NheI
HindIII
pSJ80
BamHI
BanII
NheI
NotI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pSJ41 ja
pienen restriktiokappaleen eristäminen
plasmidista pSJ80,
pSJ41-kappaleen 5'-päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ81
BamHI
BanII
NotI
NheI
ClaI
HindIII

KUVIO 2j

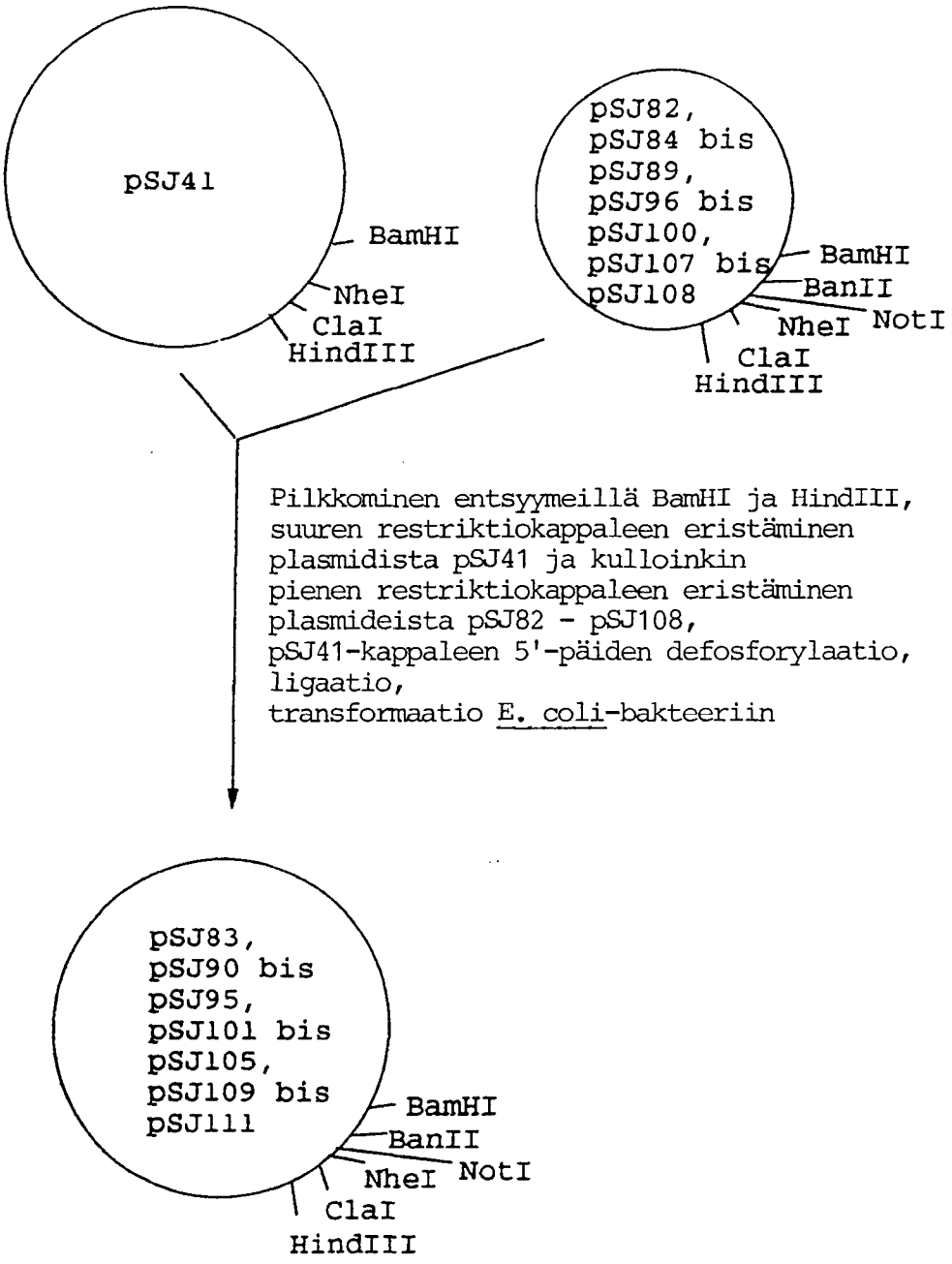


Pilkkominen entsyymeillä BanII ja NotI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin

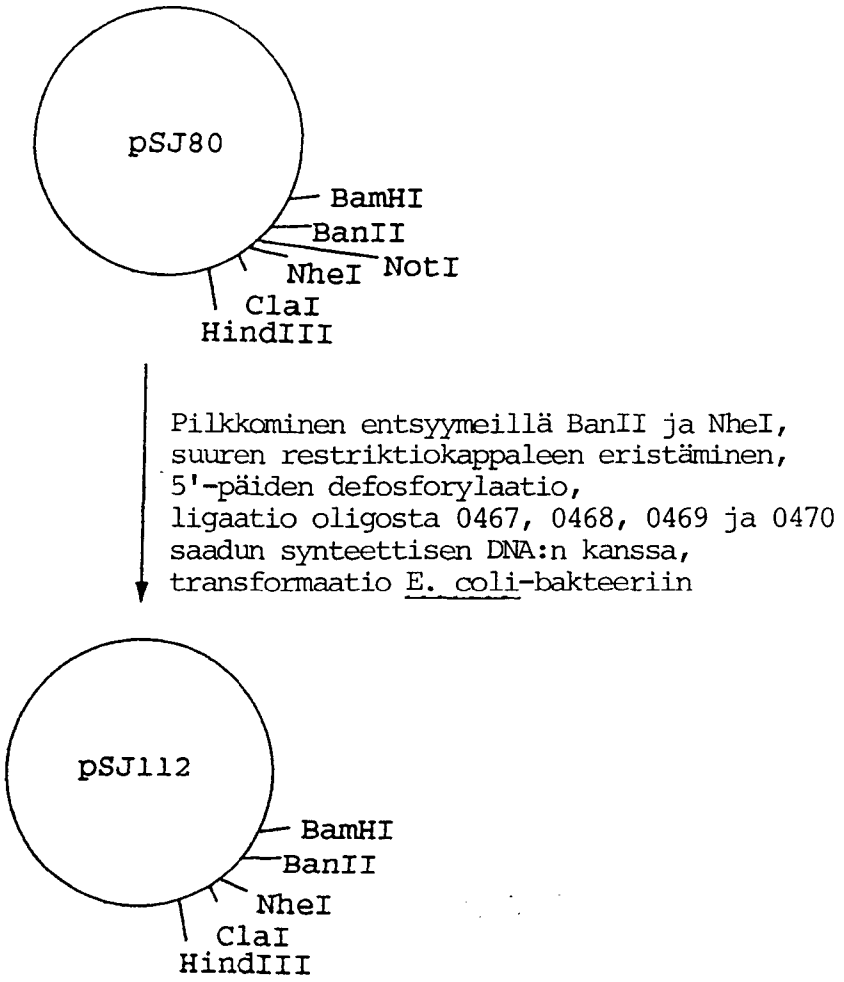


plasmidi	oligot
pSJ82	O333, O334
pSJ84	O335, O336
pSJ85	O337, O338
pSJ86	O339, O340
pSJ87	O341, O342
pSJ88	O343, O344
pSJ89	O347, O348
pSJ96	O381, O383
pSJ97	O384, O385
pSJ98	O386, O391
pSJ99	O390, O387
pSJ100	O388, O389
pSJ107	O455, O456
pSJ108	O453, O454

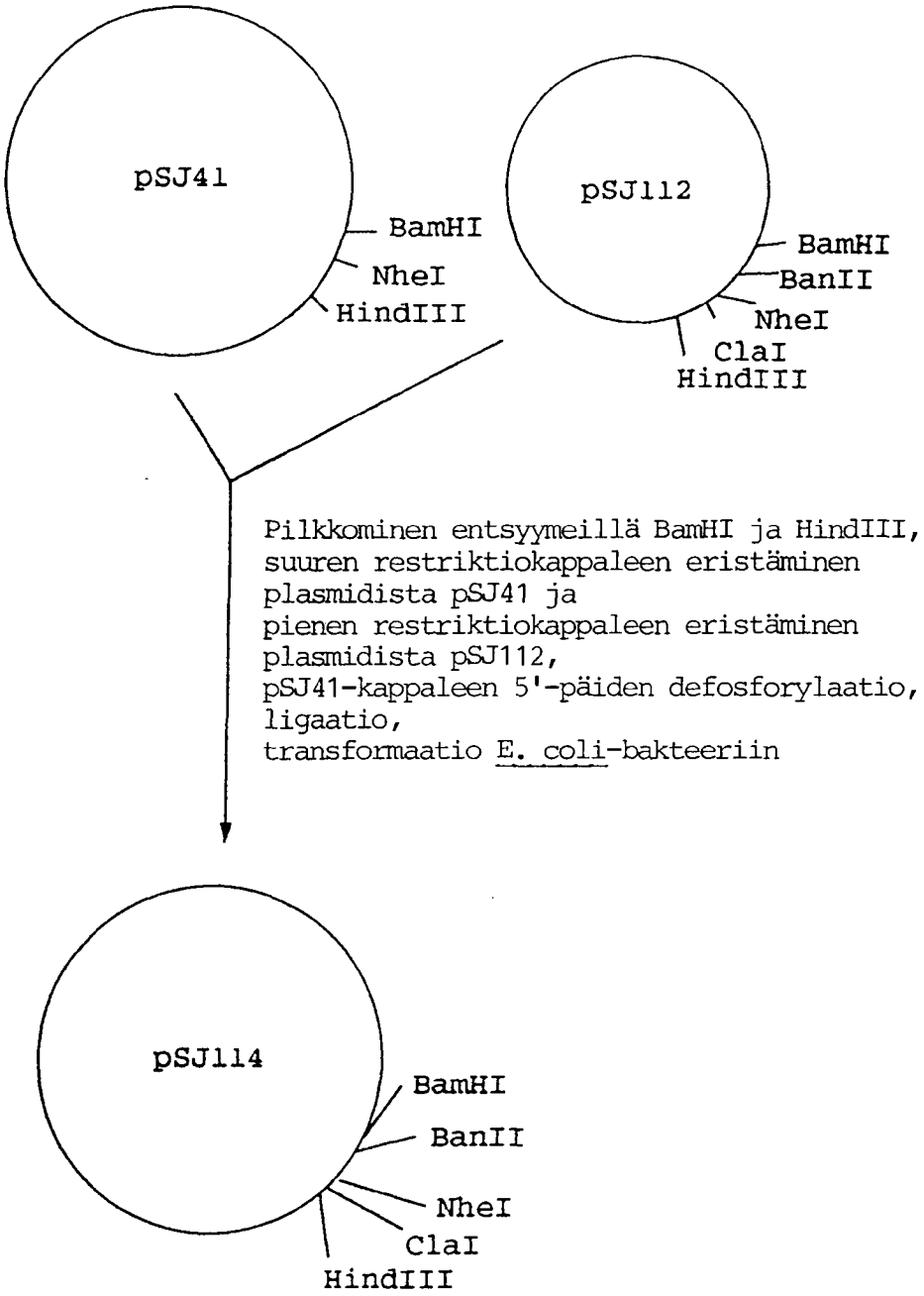
KUVIO 2k


pSJ41
BamHI
NheI
ClaI
HindIII
pSJ82,
pSJ84 bis
pSJ89,
pSJ96 bis
pSJ100,
pSJ107 bis
pSJ108
BamHI
BanII
NheI
NotI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pSJ41 ja kulloinkin
pienen restriktiokappaleen eristäminen
plasmideista pSJ82 - pSJ108,
pSJ41-kappaleen 5'-päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ83,
pSJ90 bis
pSJ95,
pSJ101 bis
pSJ105,
pSJ109 bis
pSJ111
BamHI
BanII
NheI
NotI
ClaI
HindIII

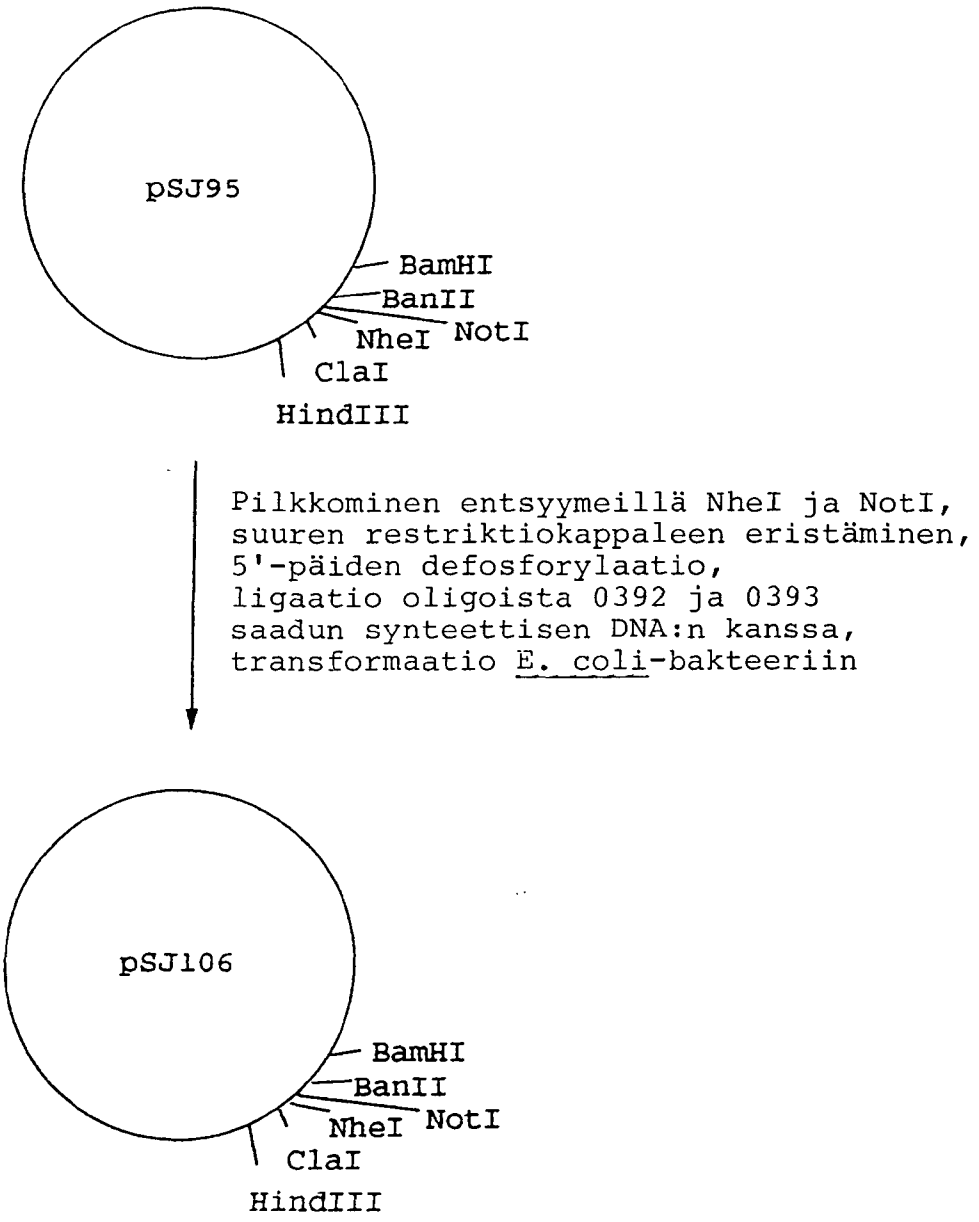
KUVIO 21


pSJ80
BamHI
BanII
NheI
NotI
ClaI
HindIII
Pilkkominen entsyymeillä BanII ja NheI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligosta 0467, 0468, 0469 ja 0470
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin
pSJ112
BamHI
BanII
NheI
ClaI
HindIII

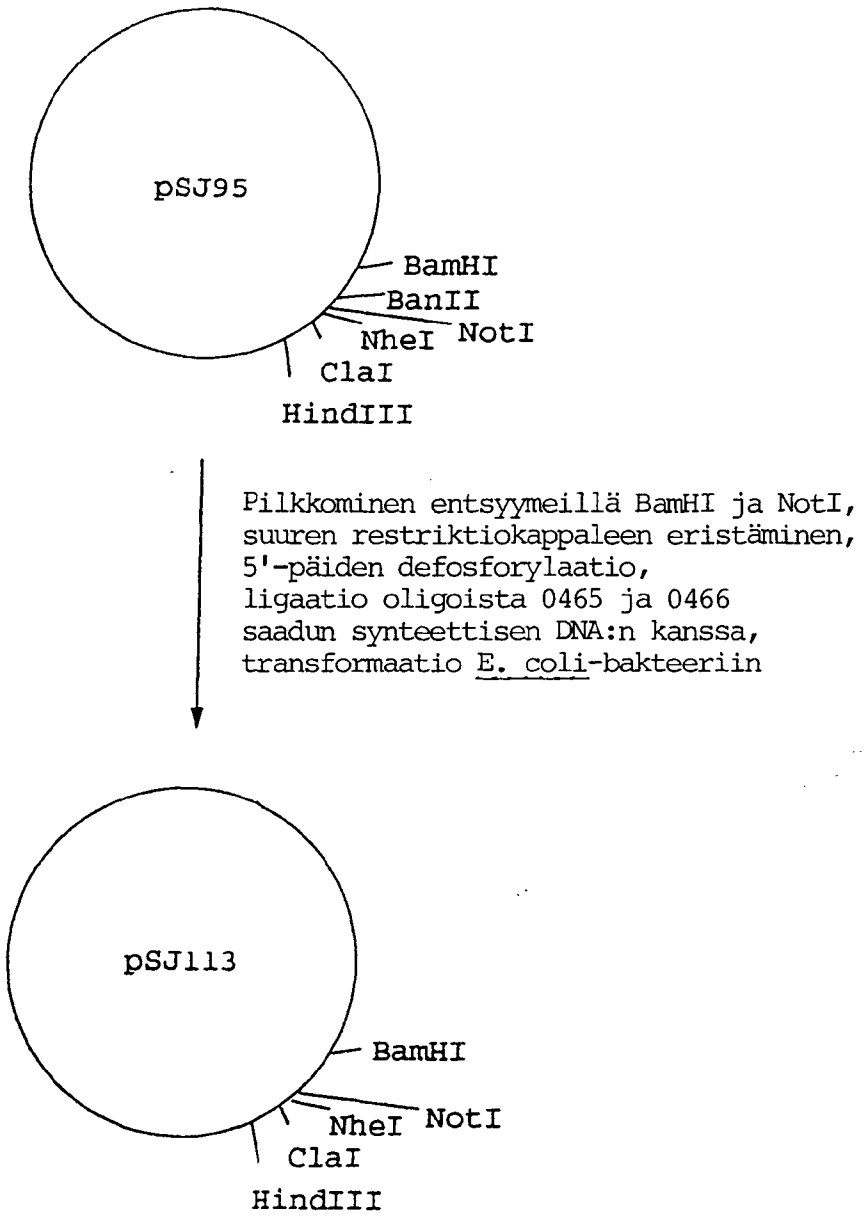
KUVIO 2m


pSJ41
BamHI
NheI
HindIII
pSJ112
BamHI
BanII
NheI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja HindIII,
suuren restriktiokappaleen eristäminen
plasmidista pSJ41 ja
pienen restriktiokappaleen eristäminen
plasmidista pSJ112,
pSJ41-kappaleen 5'-päiden defosforylaatio,
ligaatio,
transformaatio E. coli-bakteeriin
pSJ114
BamHI
BanII
NheI
ClaI
HindIII

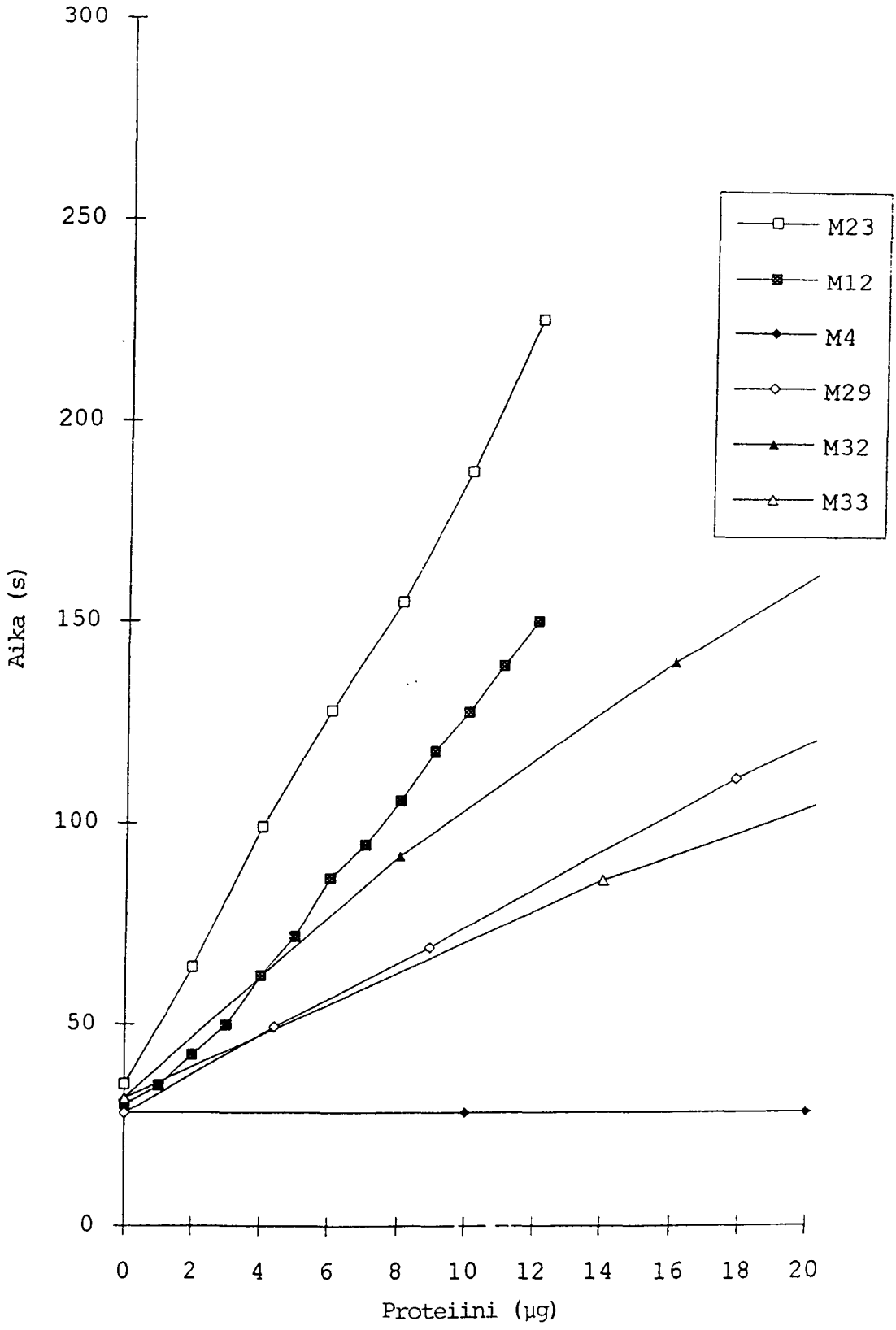
KUVIO 2n


pSJ95
BamHI
BanII
NheI
NotI
ClaI
HindIII
Pilkkominen entsyymeillä NheI ja NotI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligoista 0392 ja 0393
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin
pSJ106
BamHI
BanII
NheI
NotI
ClaI
HindIII

KUVIO 2o


pSJ95
BamHI
BanII
NheI
NotI
ClaI
HindIII
Pilkkominen entsyymeillä BamHI ja NotI,
suuren restriktiokappaleen eristäminen,
5'-päiden defosforylaatio,
ligaatio oligoista 0465 ja 0466
saadun synteettisen DNA:n kanssa,
transformaatio E. coli-bakteeriin
pSJ113
BamHI
NheI
NotI
ClaI
HindIII

KUVIO 2p



KUVIO 3: Trombiiniajan piteneminen pitoisuuden funktiona