



(12) **Oversettelse av
europeisk patentskrift**

(11) **NO/EP 2150617 B1**

NORGE

(19) NO
(51) Int Cl.

C12N 15/85 (2006.01)
C12N 15/86 (2006.01)

Patentstyret

(21)	Oversettelse publisert	2015.03.02
(80)	Dato for Den Europeiske Patentmyndighets publisering av det meddelte patentet	2014.10.22
(86)	Europeisk søknadsnr	08756678.2
(86)	Europeisk innleveringsdag	2008.06.04
(87)	Den europeiske søknadens Publiseringsdato	2010.02.10
(30)	Prioritet	2007.06.04, US, 933213 P
(84)	Utpekte stater	AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR
(73)	Innehaver	Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill Road, Tarrytown, NY 10591, US-USA
(72)	Oppfinner	CHEN, Gang, 2478 Mill Pond Road, Yorktown Heights, NY 10598, US-USA BABB, Robert, 292 Voorhis Avenue, River Edge, NJ 07661, US-USA FANDL, James, P., 40 Amanda's Way, Lagrangeville, NY 12540, US-USA
(74)	Fullmektig	Jesper Levin A/S, Enrum Slot, Vedbæk Strandvej 341, DK-2950 VEDBÆK, Danmark

(54)	Benevnelse	Regioner med forsterket ekspresjon og stabilitet
(56)	Anførte publikasjoner	WO-A-02/081632 SHMERLING DORON ET AL: "Strong and ubiquitous expression of transgenes targeted into the beta-actin locus by cre/lox cassette replacement" GENESIS THE JOURNAL OF GENETICS AND DEVELOPMENT, vol. 42, no. 4, August 2005 (2005-08), pages 229-235, XP002496257 ISSN: 1526-954X BOUHASSIRA E E ET AL: "Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange" BLOOD, AMERICAN SOCIETY OF HEMATOLOGY, US, vol. 90, no. 9, 1 January 1997 (1997-01-01), pages 3332-3344, XP002143981 ISSN: 0006-4971

Beskrivelse

BAKGRUNN

Området for oppfinnelsen

[0001] Oppfinnelsen tilveiebringer å eksprimere rekombinante proteiner i eukaryote celler. Spesielt gjelder oppfinnelsen framgangsmåter for forbedret ekspresjon av proteiner i eukaryote celler ved å anvende ekspresjonsforsterkende nukleotidsekvenser. Oppfinnelsen gjelder sekvenser med regioner med forsterket ekspresjon og stabilitet (EESYR) som letter forsterket og stabil ekspresjon av rekombinante proteiner i eukaryote celler, og framgangsmåter for å bruke slike sekvenser.

Beskrivelse av relatert teknikk

[0002] Utviklingen av ekspresjonssystemer er et viktig mål for å tilveiebringe en pålitelig og effektiv kilde av et gitt protein for forskning og terapeutisk bruk. Rekombinant proteinekspresjon i pattedyrceller er ofte foretrukket for å produsere terapeutiske proteiner på grunn av for eksempel evnen pattedyrs ekspresjonssystemer har til å modifisere rekombinante proteiner høvelig etter translasjon.

[0003] Flere vektorer er tilgjengelig for ekspresjon i pattedyrverter, som hver inneholder ulike kombinasjoner av *cis*- og, i noen tilfeller, *trans*-regulerende elementer for å oppnå høye nivåer av rekombinant protein med korte inkubasjonstider. Til tross for tilgjengeligheten av mange slike vektorer er ekspresjonen av et rekombinant protein som oppnås i pattedyrsystemer ofte uakseptabelt lav eller på annen måte utilfredsstillende. Dessuten krever det å utvikle en cellelinje som pålitelig eksprimerer tilstrekkelig høye nivåer av et ønsket protein, ofte tidkrevende klonings- og amplifikasjonstrinn. Følgelig finnes det i teknikken et behov for forbedrede pattedyrekspresjonssystemer.

[0004] Schmerling Doron et al. beskriver sterk og ubikvitær ekspresjon av transgener målrettet inn i beta-aktin-locus ved hjelp av loxP-Cre-kassettskifting (Journal of Genetics and Development, vol.42, s. 229–235 (2005)).

KORT SAMMENDRAG

[0005] Oppfinnelsen tilveiebringer en isolert celle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5, og minst ett rekombinasegjenkjenningssete som befinner seg i en nukleotidsekvens i SEKV ID NR: 1,

3 eller 5. Oppfinnelsen tilveiebringer også en CHO-celle som omfatter en ekspresjonsforsterkende sekvens som er nukleotidene 6 400–6 500 i SEKV ID NR: 5, videre omfatter minst ett rekombinasegjenkjenningssete flankert 5' eller 3' av et humant gen av interesse (GOI), der det humane GOI-et opsjonelt er flankert 5' av det første rekombinasegjenkjenningssetet, og CHO-cellen videre omfatter et andre rekombinasegjenkjenningssete som flankerer 3' det humane GOI-et. Oppfinnelsen tilveiebringer videre en isolert celle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5 og minst ett eksogent tillagt GOI i den ekspresjonsforsterkende sekvensen.

[0006] Oppfinnelsen tilveiebringer i tillegg en framgangsmåte for å lage et protein av interesse, som omfatter: (a) å tilveiebringe en vertcelle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5, eller nukleotidene 6 400–6 500 i SEKV ID NR: 5, (b) å sette inn i vertcellen, innenfor den ekspresjonsforsterkende sekvensen, et gen av interesse (GOI) som er funksjonelt knyttet til en promotor; (c) å opprettholde vertcellen i (a) under betingelser som gjør at GOI-et kan eksprimere et protein av interesse; og (d) å hente ut proteinet av interesse. Oppfinnelsen tilveiebringer videre bruk av oppfinnelsens isolerte celle eller oppfinnelsens isolerte CHO-celle for å lage et antistoff eller fragment av dette.

[0007] Her er også beskrevet en isolert nukleotidsekvens som omfatter en ekspresjonsforsterkende sekvens valgt blant en sekvens i SEKV ID NR:1-6, eller et ekspresjonsforsterkende fragment av dette.

[0008] De ekspresjonsforsterkende sekvensene som er beskrevet her, omfatter en ekspresjonsforsterkende sekvens med SEKV ID NR:5 som befinner seg i en posisjon innenfor SEKV ID NR:5 valgt blant nukleotider som strekker seg over posisjonene med nummer 10-13 515; 20-12 020; 1 020-11 020; 2 020-10 020; 3 020-9 020; 4 020-8 020; 5 020-7 020; 6 020-6 920; 6 120-6 820; 6 220-6 720; 6 320-6 620; 6 420-6 520; 6 460-6 500; 6 470-6 490; og 6 475-6 485. Andre ekspresjonsforsterkende sekvenser som er beskrevet her, er valgt fra gruppen som består av nukleotidene 5 000-7 400 i SEKV ID NR:5; 5 000-6 500 i SEKV ID NR:5; 6 400-7 400 i SEKV ID NR:5; og nukleotidene 6 400-6 500 i SEKV ID NR:5.

[0009] Rekombinasjongjenkjenningssetet kan være posisjonert som beskrevet ovenfor og tilveiebringe at den ekspresjonsforsterkende sekvensen omfatter en sekvens som er minst 90 prosent identisk, mer foretrukket minst 95 prosent identisk, mest foretrukket minst 99 prosent identisk, med den ekspresjonsforsterkende sekvensen i SEKV ID NR:5 eller et ekspresjonsforsterkende fragment av denne.

[0010] De ekspresjonsforsterkende sekvensene kan videre omfatte minst ett rekombinasegjenkjenningssete som omfatter en sekvens uavhengig valgt blant et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frt-sete. Rekombinasegjenkjenningssetet kan være innenfor den ekspresjonsforsterkende sekvensen. Rekombinasegjenkjenningssetet kan være umiddelbart tilstøtende i 5'-retningen mot det terminale nukleotidet i 5'-enden, eller umiddelbart tilstøtende i 3'-retningen mot det terminale nukleotidet i 3'-enden av den ekspresjonsforsterkende sekvensen.

[0011] Minst to rekombinasegjenkjennings seter kan være til stede innenfor den ekspresjonsforsterkende sekvensen. To rekombinasegjenkjennings seter med motsatt orientering kan være til stede innenfor den ekspresjonsforsterkende sekvensen. Tre rekombinasegjenkjennings seter kan være til stede innenfor den ekspresjonsforsterkende sekvensen.

[0012] Videre er her beskrevet en isolert nukleotidsekvens som omfatter en ekspresjonsforsterkende sekvens som er minst 80 prosent identisk, fortrinnsvis minst 90 prosent identisk, mer foretrukket minst 95 prosent identisk, mest foretrukket minst 99 prosent identisk med en ekspresjonsforsterkende sekvens i SEKV ID NR:1–6 eller et ekspresjonsforsterkende fragment av denne. Den ekspresjonsforsterkende sekvensen kan oppvise den nevnte identiteten med en sekvens i SEKV ID NR:5 som beskrevet ovenfor.

[0013] Her er beskrevet en isolert eukaryot celle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR:1-6, eller et ekspresjonsforsterkende fragment av dette. Den ekspresjonsforsterkende sekvensen kan omfatte en ekspresjonsforsterkende sekvens i SEKV ID NR:5 som beskrevet ovenfor.

[0014] Den eukaryote cellen kan være en muse-, rotte-, hamster-, eller human celle. Den eukaryote cellen kan være en CHO-celle.

[0015] Den eukaryote cellen kan videre omfatte minst én rekombinasegjenkjenningssekvens innenfor den ekspresjonsforsterkende sekvensen. Den minst ene rekombinasegjenkjenningssekvensen kan være uavhengig valgt blant et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frt-sete. Rekombinasegjenkjenningssetet kan være umiddelbart tilstøtende i 5'-retningen mot det terminale nukleotidet i 5'-enden, eller umiddelbart tilstøtende i 3'-retningen mot det terminale nukleotidet i 3'-enden av den ekspresjonsforsterkende sekvensen.

[0016] Minst to rekombinasegjenkjennings seter kan være til stede innenfor den ekspresjonsforsterkende sekvensen. To rekombinasegjenkjennings seter med motsatt orientering kan være til stede innenfor den ekspresjonsforsterkende sekvensen. Tre rekombinasegjenkjenningssekvenser kan være til stede, og én av de tre rekombinasegjenkjenningssekvensene er i en orientering motsatt de to gjenværende rekombinasegjenkjenningssekvensene.

[0017] Som beskrevet her kan rekombinasegjenkjenningssetet i den ekspresjonsforsterkende sekvensen i SEKV ID NR:5 befinne seg i en posisjon innenfor SEKV ID NR:5 valgt blant nukleotider som strekker seg over posisjonene med nummer 10-13 515; 20-12 020; 1 020-11 020; 2 020-10 020; 3 020-9 020; 4 020-8 020; 5 020-7 020; 6 020-6 920; 6 120-6 820; 6 220-6 720; 6 320-6 620; 6 420-6 520; 6 460-6 500; 6 470-6 490; og 6 475-6 485. Rekombinasegjenkjenningssetet kan være i en sekvens som er valgt fra gruppen som består av nukleotidene 5 000-7 400 i SEKV ID NR:5; 5 000-6 500 i SEKV ID NR:5; 6 400-7 400 i SEKV ID NR:5; og nukleotidene 6 400-6 500 i SEKV ID NR:5. Rekombinasegjenkjenningssetet kan befinne seg innenfor nukleotidene 6 400–6 500 i SEKV ID NR:5. Rekombinasegjenkjenningssetet kan settes inn før, etter, eller

innenfor «act»-tripletten av nukleotidene 6 471 til 6 473 i SEKV ID NR:5 i en ekspresjonsforsterkende sekvens i SEKV ID NR:5.

[0018] Rekombinasjongjenkjenningssetet kan være posisjonert som beskrevet ovenfor, med det forbehold at den ekspresjonsforsterkende sekvensen omfatter en sekvens som er minst 90 prosent identisk, mer foretrukket minst 95 prosent identisk, mest foretrukket minst 99 prosent identisk, med den nukleotidene 5 218 til 6 048 i SEKV ID NR:5 eller et ekspresjonsforsterkende fragment av dette.

[0019] Cellen kan være en CHO-celle der rekombinasegjenkjenningssetet settes inn i CHO-cellegenomet ved eller innenfor «act»-tripletten av nukleotidene 6 471 til 6 473 i SEKV ID NR:5.

[0020] Et første GOI kan settes innenfor den ekspresjonsforsterkende sekvensen i SEKV ID NR:5 som beskrevet ovenfor, og det første GOI-et opsjonelt funksjonelt knyttet til en promotor, der det promotor-tilknyttede GOI-et (eller GOI-et) er flankert 5' av et første rekombinasegjenkjenningssete og 3' av et andre rekombinasegjenkjenningssete. Et andre GOI kan settes inn 3' på det andre rekombinasegjenkjenningssetet, og det andre GOI-et flankert 3' av et tredje rekombinasegjenkjenningssete.

[0021] GOI-et kan være funksjonelt knyttet til en promotor som er i stand til å drive ekspresjon av GOI-et, der promotoren omfatter en eukaryot promotor som kan reguleres av en aktivator eller hemmer. Den eukaryote promotoren kan være funksjonelt knyttet til en prokaryot operator, og den eukaryote cellen kan opsjonelt videre omfatte et prokaryot repressorprotein.

[0022] Én eller flere valgbare markører kan være inkludert mellom det første og det andre og/eller det andre og det tredje rekombinasegjenkjenningssetet. Det første og/eller det andre genet av interesse og/eller den ene eller flere valgbare markørene kan være funksjonelt knyttet til en promotor, der promotoren kan være den samme eller ulik. Promotoren kan omfatte en eukaryot promotor (så som for eksempel en CMV-promotor), opsjonelt kontrollert av en prokaryot operator (så som for eksempel en tet-operator). Cellen kan videre omfatte et gen som koder en prokaryot repressor (så som for eksempel en tet-repressor).

[0023] Cellen kan videre omfatte et gen som er i stand til å eksprimere en rekombinase. I en spesifikk utførelsesform er rekombinasen en Cre-rekombinase.

[0024] En isolert eukaryot celle også beskrevet her som omfatter en ekspresjonsforsterkende sekvens som er minst 80 prosent, fortrinnsvis minst 90 prosent, mer foretrukket minst 95 prosent, mest foretrukket minst 99 prosent identisk med en ekspresjonsforsterkende sekvens i SEKV ID NR: 1–6 eller et ekspresjonsforsterkende fragment av dette. Den ekspresjonsforsterkende sekvensen kan være en sekvens i SEKV ID NR:5 som beskrevet ovenfor.

[0025] Her er beskrevet en eukaryot vertcelle som omfatter en ekspresjonsforsterkende sekvens valgt fra SEKV ID NR:1-6 eller et ekspresjonsforsterkende fragment av dette, som omfatter et første rekombinasegjenkjenningssete fulgt av en første eukaryot promotor, et

første markørgen, en andre eukaryot promotor, et andre markørgen, et andre rekombinasegjenkjenningssete, en tredje eukaryot promotor, et tredje markørgen, og et tredje rekombinasegjenkjenningssete. Den ekspresjonsforsterkende sekvensen kan være i SEKV ID NR:5 som beskrevet ovenfor.

[0026] Det første, andre og tredje rekombinasegjenkjenningssetet kan være ulike. Rekombinasegjenkjenningssetene kan være valgt blant et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frt-sete.

[0027] Det første markørgenet kan være et legemiddelresistensgen. Legemiddelresistensgenet kan være et puromycinresistensgen. Det andre og tredje markørgenet kan kode to ulike fluorescerende proteiner. De to ulike fluorescerende proteinene kan være valgt blant *Discosoma coral* (DsRed), grønt fluorescerende protein (GFP), forsterket grønt fluorescerende protein (eGFP), cyan-fluorescerende protein (CFP), forsterket cyan-fluorescerende protein (eCFP), og gult fluorescerende protein (YFP). De to ulike fluorescerende proteinene kan være eCFP og DsRed.

[0028] Den første, andre og tredje promotoren kan være den samme. Den første, andre og tredje promotoren kan være ulike. Den første promotoren kan være ulik den andre og tredje promotoren, og den andre og tredje promotoren er den samme. Den første promotoren kan være en SV40-senpromotor, og den andre og tredje promotoren er hver en human CMV-promotor.

[0029] En eukaryot vertcelle også beskrevet her, som omfatter en ekspresjonsforsterkende sekvens valgt fra SEKV ID NR:1-6, eller et ekspresjonsforsterkende fragment av dette, minst ett rekombinasegjenkjenningssete i den ekspresjonsforsterkende sekvensen, og minst ett gen av interesse (GOI) i den ekspresjonsforsterkende sekvensen. Den ekspresjonsforsterkende sekvensen kan være en sekvens i SEKV ID NR:5, som beskrevet ovenfor.

[0030] Cellen beskrevet her kan omfatte et første rekombinasegjenkjenningssete fulgt av en første promotor funksjonelt knyttet til et første GOI. Det første GOI-et kan være fulgt av et andre rekombinasegjenkjenningssete. Det andre rekombinasegjenkjenningssete kan være fulgt av en andre promotor funksjonelt knyttet til et andre GOI. Det andre GOI-et kan være fulgt av et tredje rekombinasegjenkjenningssete. Minst én markør kan være funksjonelt knyttet til en tredje promotor og befinne seg mellom det andre rekombinasegjenkjenningssetet og den andre promotoren. Det første rekombinasegjenkjenningssetet kan være orientert i en motsatt orientering av det andre og tredje rekombinasegjenkjenningssetet. Den første og andre promotoren kan være eukaryote promotorer funksjonelt knyttet til en prokaryot operator. Den første og andre promotoren kan være CMV-promotorer funksjonelt knyttet til tet-operatorsekvenser. Cellen kan videre omfatte et gen som er i stand til å eksprimere en prokaryot repressor. Den prokaryote repressoren kan være en tet-repressor. Cellen kan omfatte et gen som er i stand til å eksprimere en Cre-rekombinase.

[0031] Et første og et andre markørgen kan befinne seg mellom det andre rekombinasegjenkjenningssetet og den andre promotoren, og en IRES er mellom det første og andre markørgenet. Det første kodonet (ATG) på det første markørgenet kan

være umiddelbart 5' til det andre rekombinasegjenkjenningssetet, og det andre kodonet på det første markørgenet er umiddelbart 3' til det andre rekombinasegjenkjenningssetet. Det første markørgenet kan inneholde et intron, og det andre rekombinasegjenkjenningssetet befinner seg i intronet, slik at den amino-terminale halvdelen av det første markørgenet og 5'-halvdelen av intronet befinner seg 5' av det andre rekombinasegjenkjenningssetet, og 3'-halvdelen av intronet og den karboksy-terminale halvdelen av det første markørgenet er umiddelbart 3' til det andre rekombinasegjenkjenningssetet.

[0032] Det første, andre og tredje rekombinasegjenkjenningssetet kan være ulike. Rekombinasegjenkjenningssettene kan være valgt blant et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frt-sete.

[0033] Det første markørgenet kan være et legemiddelresistensgen. Legemiddelresistensgenet kan være et hygromycinresistensgen. Det andre markørgenet kan kode et fluorescerende protein. Det fluorescerende proteinet kan være valgt blant DsRed, GFP, eGFP, CFP, eCFP og YFP.

[0034] Den første, andre og tredje promotoren kan være den samme. Den første, andre og tredje promotoren kan være ulike. Den tredje promotoren kan være ulik den første og andre promotoren, og den første og andre promotoren den samme. Den tredje promotoren kan være en SV40-senpromotor, og den første og andre promotoren er hver en human CMV-promotor.

[0035] Den første og andre promotoren kan være funksjonelt knyttet til en prokaryot operator. Operatoren kan være en tet-operator.

[0036] Verticellelinjen kan ha et eksogent tillagt gen som koder en rekombinase integrert i genomet sitt, funksjonelt knyttet til en promotor. Rekombinasen kan være en Cre-rekombinase. Verticellen kan ha et gen som koder et regulerende protein integrert i genomet sitt, funksjonelt knyttet til en promotor. Det regulerende proteinet kan være et tet-repressorprotein.

[0037] Det første GOI-et og det andre GOI-et kan kode en lett kjede, eller fragment av dette, av et antistoff, eller en tung kjede, eller fragment av dette, av et antistoff. Det første GOI-et kan kode en lett kjede av et antistoff, og det andre GOI-et kan kode en tung kjede av et antistoff.

[0038] Her beskrives en framgangsmåte for å lage et protein av interesse, som omfatter: (a) å tilveiebringe en verticelle som omfatter en ekspresjonsforsterkende sekvens valgt fra en sekvens i SEKV ID NR:1–6; (b) å føre inn i verticellen, innenfor den ekspresjonsforsterkende sekvensen, et gen av interesse (GOI) som er funksjonelt knyttet til en promotor; (c) å opprettholde verticellen i (a) under betingelser som gjør at GOI-et kan eksprimere et protein av interesse; og (c) å hente ut proteinet av interesse.

[0039] Cellen kan være en CHO-celle, og nukleotidsekvensen er en ekspresjonsforsterkende sekvens i SEKV ID NR:5 som beskrevet ovenfor.

[0040] GOI-et kan settes inn i cellen ved hjelp av en målrettet vektor for homolog

rekombinasjon, der den målrettede vektoren omfatter en 5'-homologiarm homolog med en sekvens til stede i minst én av SEKV ID NR:1-6, et GOI, og en 3'-homologiarm homolog med en sekvens til stede i minst én av SEKV ID NR:1-6. Konstruktet kan videre omfatte to, tre, fire, eller fem eller flere gener av interesse. Ett eller flere av genene av interesse kan være funksjonelt knyttet til en promotor.

[0041] GOI-et kan settes inn ved hjelp av en integraseteknologi, for eksempel integraseteknologi som gjør bruk av att-seter, så som Invitrogen's Gateway™ og Multisite Gateway™ kloningssystemer som gjør bruk av att-seterekombinasjon med bakteriofag lambda.

[0042] Den ekspresjonsforsterkende sekvensen kan omfatte ett eller flere rekombinasegjenkjenningsseter som beskrevet ovenfor, og GOI-et settes inn i den ekspresjonsforsterkende sekvensen gjennom handlingen til en rekombinase som gjenkjenner rekombinasegjenkjenningssetet.

[0043] Den ekspresjonsforsterkende sekvensen kan omfatte to rekombinasegjenkjenningsseter.

[0044] Den ekspresjonsforsterkende sekvensen kan omfatte et første, et andre og et tredje rekombinasegjenkjenningssete. Det første, andre og tredje rekombinasegjenkjenningssetet kan være ulike. Det første, andre og tredje rekombinasegjenkjenningssetet kan være i ulik orientering. Det første setet kan være 5' til det andre setet, og det andre setet 5' til det tredje setet. Det andre og tredje setet kan være i motsatt orientering i forhold til det første setet.

[0045] Et første og et andre GOI settes inn i den ekspresjonsforsterkende sekvensen. Det første GOI-et kan settes inn mellom det første og det andre rekombinasegjenkjenningssetet, og det andre GOI-et settes inn mellom det andre og det tredje rekombinasegjenkjenningssetet. Rekombinasegjenkjenningssetene kan være uavhengig valgt blant et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frrt-sete.

[0046] Det første GOI-et flankert av rekombinasegjenkjenningsseter på en første vektor og det andre GOI-et flankert av rekombinasegjenkjenningsseter på en andre vektor kan settes inn i en celle som omfatter en ekspresjonsforsterkende sekvens som omfatter tre rekombinasegjenkjenningsseter i et enkelt trinn.

[0047] GOI-et kan være funksjonelt knyttet til en eukaryot promotor. Den eukaryote promotoren kan være funksjonelt knyttet til en prokaryot operator. Den eukaryote promotoren kan være en CMV-promotor, og den prokaryote operatoren en tet-operator.

[0048] Cellen kan omfatte et gen som er i stand til å eksprimere en prokaryot repressor. Den prokaryote repressoren kan være en tet-repressor.

[0049] Cellen kan omfatte et gen som er i stand til å eksprimere en rekombinase. Rekombinasen kan være en Cre-rekombinase.

[0050] En eukaryot celle er også beskrevet her, der den eukaryote cellen omfatter minst

én ekspresjonsforsterkende sekvens i SEKV ID NR:1-6 og minst ett eksogent tilsatt gen i den ekspresjonsforsterkende sekvensen. Det eksogent tilsatte genet kan være funksjonelt knyttet til en eksogent tilsatt promotor.

[0051] Den ekspresjonsforsterkende sekvensen kan være en sekvens i SEKV ID NR:5, og det minst ene eksogent tilsatte genet integrert eller innsatt i den ekspresjonsforsterkende sekvensen er et humant gen. Den eukaryote cellen kan være en CHO-celle, det eksogent tilsatte genet er et humant gen, og det humane genet er funksjonelt knyttet til en eksogent tilsatt eukaryot promotor.

[0052] Her er også beskrevet en målrettet vektor for homolog rekombinasjon, der den målrettede vektoren omfatter en 5'-homologiarm, et GOI og en 3'-homologiarm, der hver homologiarm er homolog med en sekvens i én av SEKV ID NR:1-6. 5'- og 3'-homologiarmene kan være homologe med sekvenser i SEKV ID NR:5. Den målrettede vektoren kan omfatte to, tre, fire, eller fem eller flere gener av interesse. GOI-et kan være funksjonelt knyttet til en promotor. Hvert av de to, tre, fire, eller fem eller flere av genene av interesse kan være funksjonelt knyttet til en promotor.

[0053] Her er videre beskrevet en ekspresjonsvektor som omfatter en ekspresjonsforsterkende nukleotidsekvens valgt blant SEKV ID NR: 1–6 eller et ekspresjonsforsterkende fragment av dette. Den ekspresjonsforsterkende sekvensen kan være i SEKV ID NR:5 som beskrevet ovenfor.

[0054] Vektoren kan videre omfatte en promotor. Promotoren kan være en human CMV-promotor.

[0055] Vektoren kan videre omfatte et kloningssete for et eksprimerbart gen av interesse (GOI). Nukleotidsekvensen valgt blant SEKV ID NR: 1-6 eller ekspresjonsforsterkende fragment av dette kan befinne seg 3' i forhold til kloningssetet for det eksprimerbare GOI-et. Nukleotidsekvensen valgt blant SEKV ID NR: 1-6 eller ekspresjonsforsterkende fragment av dette kan befinne seg 5' i forhold til kloningssetet for det eksprimerbare GOI-et.

[0056] Vektoren kan omfatte en nukleotidsekvens valgt blant SEKV ID NR:1-6 eller et ekspresjonsforsterkende fragment, en human CMV-promotor, et GOI, og en termineringssekvens, der nukleotidsekvensen valgt blant SEKV ID NR: 1-6, eller det ekspresjonsforsterkende fragmentet av dette, befinner seg 5' i forhold til CMV-promotoren. Vektoren kan videre omfatte et intron valgt blant et CMV-MIE-intron og et β -globin-intron fra kanin.

[0057] Her er også beskrevet en ekspresjonsvektor som omfatter en ekspresjonsforsterkende nukleotidsekvens som er minst 80 prosent identisk, fortrinnsvis minst 90 prosent identisk, mer foretrukket minst 95 prosent identisk, mest foretrukket minst 99 prosent identisk med en sekvens valgt blant SEKV ID NR: 1–6 eller et ekspresjonsforsterkende fragment av dette.

[0058] Her beskrives videre en framgangsmåte for å lage et protein av interesse, som omfatter: (a) å sette inn i en vertcelle en ekspresjonsvektor som omfatter en

ekspresjonsforsterkende sekvens valgt blant en sekvens i SEKV ID NR:1-6 og et GOI som koder et protein av interesse, der GOI-et er funksjonelt knyttet til en promotor og funksjonelt knyttet til den ekspresjonsforsterkende sekvensen; (b) å dyrke vertcellen i (a) under betingelser som gjør at GOI-et kan eksprimere; og (c) å hente ut proteinet av interesse.

[0059] Sekvensen med regioner med forsterket ekspresjon og stabilitet kan være en ekspresjonsforsterkende sekvens i SEKV ID NR:1-6. Den ekspresjonsforsterkende sekvensen kan være i SEKV ID NR:5 som beskrevet ovenfor.

[0060] Det rekombinerende proteinet kan være valgt fra gruppen som består av en underenhet av et immunoglobulin eller fragment av dette, og en reseptor eller ligandbindende fragment av dette. Det rekombinerende proteinet kan være valgt fra gruppen som består av en antistofflekkjede eller antigenspesifikt fragment av dette, og en antistofftungkjede eller antigenspesifikt fragment av dette.

[0061] Den ekspresjonsforsterkende sekvensen kan være plassert i den indikerte orienteringen som indikert i SEKV ID NR:1-6, eller i det motsatte av orienteringen indikert i SEKV ID NR:1-6.

[0062] Ethvert av aspektene og utførelsesformene av oppfinnelsen kan brukes sammen med ethvert annet aspekt eller utførelsesform av oppfinnelsen, med mindre noe annet er spesifisert eller åpenbart ut ifra sammenhengen.

[0063] Andre gjenstander og fordeler vil bli åpenbare ved en gjennomgang av den følgende detaljerte beskrivelsen.

KORT BESKRIVELSE AV FIGURENE

[0064]

Fig. 1. Skjematisk diagram av et retroviralt konstrukt, pTE252, som brukes til å sette nukleinsyrekonstruktet inn i et cellegenom. LTR: long terminal repeat (lang enderepetisjon); LoxP: Cre-rekombinasegjenkjenningsssekvens: ATAACCTTCGTATAATGTATGCTATACGAAGTTGT (SEKV ID NR:7); Lox511: en mutasjon av LoxP-sekvenser: ATAACCTTCGTATAATGTATACTATACGAAGTTAG (SEKV ID NR:8); Lox511-sekvensen gjenkjennes av Cre-rekombinasen, men Lox511-setet rekombinerer ikke med et LoxP-sete. GFP: Grønt fluorescerende protein; CMV MIE: humant cytomegalovirus større umiddelbar tidlig promotor; npt: neomycin fosfotransferase; bla: beta laktamase; IRES: internt ribosomalt inngangssete.

Fig. 2 illustrerer et plasmid konstrukt som brukes til å identifisere en EESYR.

Fig. 3A og 3B viser en sammenstilling av EESYR-sekvenser fra kinesisk hamster-ovarium (CHO), mus, menneske og rotte for et fragment av SEKV ID NR:5.

Fig. 4 viser en sammenstilling av SEKV ID NR:5 med sekvenser fra mus, menneske og rotte.

Fig. 5 illustrerer at en EESYR, funksjonelt knyttet til et gen av interesse (GOI), oppviser forsterket ekspresjon over et GOI som ikke er funksjonelt knyttet til en EESYR.

Fig. 6 viser EESYR-er sammenliknet i sin relative evne til å forsterke ekspresjon av et funksjonelt tilknyttet GOI.

Fig. 7 illustrerer kloningsegenskaper hos celler når det gjelder EESYR-funksjonalitet.

Fig. 8 illustrerer at EESYR-celler gjennomgår spesifikk og effektiv rekombinasjon.

Fig. 9 illustrerer sjeldenheten av tilfeldige integrasjonshendelser i EESYR-celler.

Fig. 10 illustrerer testing av *cis*-fungerende elementer ved hjelp av EESYR-sekvenser.

Fig. 11 illustrerer testing av promotorer, introner og UTR-er ved hjelp av et EESYR-system.

Fig. 12 illustrerer optimalisering av proteinekspresjon ved hjelp av et EESYR-system.

DETALJERT BESKRIVELSE

[0065] Før de foreliggende framgangsmåtene beskrives, skal det forstås at denne oppfinnelsen ikke er begrenset til spesifikke framgangsmåter og eksperimentelle vilkår som beskrives, ettersom slike framgangsmåter og vilkår kan variere. Det skal også forstås at terminologien som brukes her, bare har som formål å beskrive spesifikke utførelsesformer, og ikke har til hensikt å være begrensende, ettersom den foreliggende oppfinnelsens omfang bare vil begrenses av de medfølgende kravene.

[0066] Slik de brukes i denne patentteksten og de medfølgende kravene, omfatter entallsformene i ubestemt og bestemt form også et flertall av det henviste med mindre sammenhengen klart tilsier noe annet. Dermed omfatter for eksempel en henvisning til «en framgangsmåte» én eller flere framgangsmåter og/eller trinn av den typen som beskrives her, og/eller som vil være åpenbare for fagpersonen som leser denne patentteksten.

[0067] Såframt de ikke er definert annerledes eller spesifisert annerledes, har alle tekniske

og vitenskapelige begreper som brukes her, samme betydning som de vanlig forstås av den vanlige fagpersonen innenfor det området denne oppfinnelsen hører til.

[0068] Selv om alle framgangsmåter og materialer som er like eller ekvivalente med de som beskrives her, kan brukes i utøvelsen eller testingen av den foreliggende oppfinnelsen, beskrives nå spesifikke framgangsmåter og materialer.

Definisjoner

[0069] DNA-regioner er funksjonelt sammenknyttet når de er funksjonelt relatert til hverandre. For eksempel er en promotor funksjonelt knyttet til en kodesekvens dersom promotoren er i stand til å delta i transkripsjonen av sekvensen; et ribosombindende sete er funksjonelt knyttet til en kodesekvens dersom det befinner seg i en posisjon der det tillater translasjon. Generelt kan funksjonell tilknytning omfatte, men krever ikke, kontiguitet. I tilfellet med sekvenser som sekretledere er kontiguitet og riktig plassering i en leseramme typiske egenskaper. En EESYR er funksjonelt knyttet til et GOI der den er funksjonelt relatert til genet av interesse, for eksempel der dens nærvær fører til forsterket ekspresjon av GOI-et.

[0070] Når termen «forsterket» brukes til å beskrive forsterket ekspresjon, omfatter den en forsterkning på fra minst omkring 1,5 ganger til minst omkring 2 ganger forsterkning i ekspresjon over det som typisk observeres ved tilfeldig integrasjon i et genom, for eksempel sammenliknet med en pool av tilfeldige integranter av en enkelt kopi av det samme ekspresjonskonstruktet. Ganger ekspresjonsforsterkning observert ved hjelp av sekvensene fra oppfinnelsen er i sammenlikning med et ekspresjonsnivå av samme gen, målt under vesentlig samme forhold, i fravær av en sekvens fra oppfinnelsen. Slik det brukes her, brukes uttrykket «ekspresjonsforsterkende» synonymt med «forsterket ekspresjon og stabilitet» når det gjelder en region eller en sekvens. En «region med forsterket ekspresjon og stabilitet», her også kalt en «EESYR», er en region eller en sekvens som oppviser mer effektiv rekombinasjon, innsettingsstabilitet og høyere nivå av ekspresjon enn det som typisk observeres ved tilfeldig integrasjon i et genom.

[0071] Forsterket rekombinasjonseffektivitet omfatter en forsterkning av evnen en locus har til å rekombinere (for eksempel ved hjelp av rekombinasegjenkjenningsseter). Forsterkning betegner en effektivitet over tilfeldig rekombinasjon, som typisk er 0,1 %. En foretrukket forsterket rekombinasjonseffektivitet er omkring 10 ganger over tilfeldig, eller omkring 1 %. Med mindre det er spesifisert, er ikke den krevde oppfinnelsen begrenset til en spesifikk rekombinasjonseffektivitet.

[0072] Der uttrykket «eksogent tilsatt gen» eller «eksogent tilsatt GOI» anvendes med henvisning til en EESYR, viser uttrykket til ethvert gen som ikke er til stede i EESYR-en slik EESYR-en finnes i naturen. For eksempel kan et «eksogent tilsatt gen» i en CHO-EESYR (f.eks. en EESYR som omfatter en sekvens i SEKV ID NR:5) være et hamstergen som ikke finnes i CHO-EESYR-en i naturen (dvs. et hamstergen fra en annen locus i hamstergenomet), et gen fra en annen art (f.eks. et humant gen), et kimært gen (f.eks.

menneske/mus), eller et annet gen som ikke er funnet å eksistere i CHO-EESYR-en i naturen.

[0073] Prosent identitet når en EESYR beskrives er ment å omfatte homologe sekvenser som oppviser den angitte identiteten langs regioner med kontiguøs homologi, men tilstedeværelsen av mellomrom, slettinger, eller innsetninger som ikke har noen homolog i den sammenliknede sekvensen, er ikke tatt hensyn til ved beregning av prosent identitet. Ved forklaring av bruken av «prosent identitet» i denne sammenhengen vil det henvises til følgende sekvenssammenlikning:

EESYR	5595	AGATTCTGTGGGCTCTGAGGCAACTTGACCTCAGCCAGATGGTATTTGAATAACCTGCTC	5654
Rat Ch3	5619	AGATTCTGTGGGTCTGAGACAACCTTGACTTCAGCCAGATGGCATTGAATAAC-----	5672
M Mus Ch2	6618	AGATTCTGTGGGTCTGAGACAACCTTGACTTTAGCCAGATGGTATTTGAGTAATCTGGG-	6676
H. Sap. Ch20	6620	AGATTCTGTGGGCTTTGGGACAGCTTGACTTCAACTAGATGGTATTTGAATAATCTGCT-	6678
***** *			

[0074] Slik det brukes her, vil ikke en bestemmelse av «prosent identitet» mellom «EESYR»-sekvensen ovenfor (for en CHO-celle-EESYR eller fragment av dette) og en rottehomolog («Rat Ch3») omfatte en sammenlikning mellom CHO-sekvensene 5649 til 5654, ettersom rottehomologen ikke har noen homolog sekvens å sammenlikne i en sammenstilling (dvs. at CHO-EESYR-en har en innsetting ved det punktet, eller at rottehomologen har et mellomrom eller en sletting, alt etter som). I sammenlikningen ovenfor vil dermed prosent identitet-sammenlikningen strekke seg fra «AGATTC» i 5'-enden til «AATAAC» i 3'-enden. I det tilfellet skiller rottehomologen seg bare ut ved at den har en «T» i CHO-EESYR-posisjon 5607, en «A» i CHO-EESYR-posisjon 5614, en «T» i CHO-EESYR-posisjon 524, og en «C» i CHO-EESYR-posisjon 5637. Ettersom sammenlikningen er over 54 kontiguøse baser i en lengde på 60 basepar, med bare fire ulikheter (som ikke er mellomrom, slettinger, eller innsetninger), er det over 90 % identitet mellom de to sekvensene (CHO og rotte) fra CHO-EESYR-posisjon 5595 til CHO-EESYR-posisjon 5654 (ettersom «prosent identitet» ikke omfatter fradrag for mellomrom, slettinger og innsetninger).

Generell beskrivelse

[0075] Oppfinnelsen er i det minste delvis basert på oppdagelsen av at det finnes sekvenser i et genom som oppviser mer effektiv rekombinasjon, innsettingsstabilitet, og et høyere nivå av ekspresjon enn andre regioner eller sekvenser i genomet. Oppfinnelsen er også i det minste delvis basert på avdekkingen av at når slike ekspresjonsforsterkende sekvenser er identifisert, kan et egnet gen eller konstrukt tilsettes eksogent i eller nær sekvensene, og at det eksogent tilsatte genet kan eksprimeres på en fordelaktig måte. Slike sekvenser, kalt regioner med forsterket ekspresjon og stabilitet (EESYR), kan lages slik at de omfatter rekombinasegjenkjennings seter for plassering av gener av interesse for å skape cellelinjer som er i stand til å eksprimere proteiner av interesse. EESYR-er kan også inkluderes som i ekspresjonskonstrukter som for eksempel ekspresjonsvektorer.

Ekspresjonsvektorer som omfatter EESYR-er kan brukes til å eksprimere proteiner transient, eller kan integreres i et genom ved tilfeldig eller målrettet rekombinasjon som for eksempel homolog rekombinasjon eller rekombinasjon formidlet av rekombinaser som gjenkjenner spesifikke rekombinasjons seter (f.eks. Cre-lox-formidlet rekombinasjon). Ekspresjonsvektorer som omfatter EESYR-er kan også brukes til å vurdere effektiviteten av andre DNA-sekvenser, for eksempel *cis*-fungerende reguleringssekvenser.

[0076] CHO-EESYR-en beskrevet i detalj her ble identifisert ved tilfeldig integrasjon av DNA som omfattet lox-seter, inn i et CHO-cellegenom, fulgt av utvelgelse for å identifisere sekvenser der ekspresjon var forsterket. Tilfeldig integrasjon og innsetting av lox-setet ble oppnådd ved hjelp av et retroviralt konstrukt. Utvelgelse og filtrering ble oppnådd ved hjelp av legemiddelresistensmarkører og detekterbare markeringer (f.eks. fluorescerende proteiner med FACS-filtrering), ved bruk av rekombinasjonsframgangsmåter som benyttet setespesifikk rekombinasjon (f.eks. lox-seter og Cre-rekombinase). Utvelgelse fortsatte til minst 1,5 til 2 ganger forsterket ekspresjon over ekspresjon observert ved tilfeldig integrasjon av et ekspresjonskonstrukt inn i CHO-cellegenomet. Etter identifisering av EESYR-en ble rekombinasegjenkjenningsseter (i eksempelet som er gitt, lox-seter) opprettholdt i EESYR-en for å sette inn ekspresjonskassetter som omfatter et eksprimerbart GOI, sammen med andre ønskede elementer som for eksempel promotorer, forsterkere, markører, operatorer osv.

[0077] En illustrasjon av et plasmidkonstrukt brukt til å identifisere en EESYR beskrevet i denne patentsøknaden er vist i fig. 2. Plasmidredningskonstruktet omfatter en ekspresjonskassetten drevet av en promotor, der kassetten er flankert på 5'- og 3'-enden med rekombinasegjenkjenningsseter (representert ved kule og strek i fig. 2). Det er vist innsetting i en EESYR-locus, der innsettingen fører til at plasmidredningskonstruktet kommer i stedet for en ekspresjonskassetten som omfatter en promotor og en markør, der ekspresjonskassetten i EESYR-locusen er flankert på sin 5'- og 3'-ende av rekombinasegjenkjenningsseter (se fig. 2).

[0078] Her er beskrevet sammensetninger og framgangsmåter for å stabilt integrere en nukleinsyresekvens i en eukaryot celle, der nukleinsyresekvensen er i stand til forsterket ekspresjon ved å være integrert i eller nær en EESYR. Her er beskrevet celler som inneholder en rekombinasegjenkjenningsssekvens i eller nær en EESYR, bekvemt for å sette inn et GOI, for å oppnå ekspresjon av et protein av interesse fra GOI-et. Her er også beskrevet sammensetninger og framgangsmåter for å bruke EESYR-er i forbindelse med ekspresjonskonstrukter, for eksempel ekspresjonsvektorer, og for å tilsette en eksogen EESYR i en eukaryot celle av interesse.

Fysiske og funksjonelle egenskaper hos en EESYR

[0079] Nukleinsyresekvensene kalt EESYR-er ble empirisk identifisert ved sekvenser oppstrøms og nedstrøms integrasjonssetet på et nukleinsyrekonstrukt (som omfattet en ekspresjonskassetten) hos en cellelinje som eksprimerte et rapportørprotein på et høyt nivå. EESYR-nukleinsyresekvensene beskrevet her er sekvenser med en ny funksjonalitet

tilknyttet forsterket ekspresjon av en nukleinsyre (for eksempel en eksogen nukleinsyre som omfatter et GOI) som ser ut til å fungere annerledes enn det som tidligere er beskrevet for *cis*-fungerende elementer så som promotorer, forsterkere, locuskontrollregioner, stillastilknytningsregioner eller matrisetilknytningsregioner. EESYR-er ser ikke ut til å ha åpne leserammer (ORF), noe som gjør det usannsynlig at EESYR-er koder nye *trans*-aktivatorproteiner. Transfeksjonseksperimenter viste at EESYR-sekvenser oppviser noen av egenskapene til *cis*-fungerende elementer. EESYR-aktivitet detekteres ikke i transiente transfeksjonsassayer; EESYR-sekvenser ser også ut til å være distinkte fra promotor- og forsterkerelementer, som detekteres ved hjelp av disse framgangsmåtene.

[0080] Selv om EESYR-sekvenser beskrevet i detalj her ble isolert fra genomet i to cellelinjer, er EESYR-sekvenser fra disse to cellelinjene de samme. EESYR-aktivitet ble identifisert i et 6 472 kb fragment av CHO-genomisk DNA 5' i forhold til et unikt integrasjonssete hos en retroviral vektor omfattende en DsRed-rapportørkodende sekvens og i et 7 045 kb fragment av CHO-genomisk DNA 3' i forhold til integrasjonssetet. Ekspresjonsvektorer omfattende den isolerte 6 472 kb regionen og den isolerte 7 045 kb regionen og kortere fragmenter av dette var i stand til å gi CHO-celler som var transifisert med dem, høye nivåer av ekspresjon av rekombinerende proteiner.

[0081] Her er også beskrevet ekspresjonsvektorer som omfatter reversorienterte EESYR-fragmenter. Reversorienterte EESYR-fragmenter var også i stand til å gi CHO-celler som var transifisert med dem, høye nivåer av ekspresjon av rekombinerende proteiner.

[0082] Andre kombinasjoner av fragmentene beskrevet her kan også utvikles. Eksempler på andre kombinasjoner av fragmentene beskrevet her som også kan utvikles, omfatter sekvenser som omfatter flere kopier av EESYR-en beskrevet her, eller sekvenser derivert ved å kombinere den beskrevne EESYR-en med andre nukleotidsekvenser for å oppnå optimale kombinasjoner av regulerende elementer. Slike kombinasjoner kan være kontiguøst lenket eller arrangert for å tilveiebringe optimal avstandsfordeling av EESYR-fragmentene (f.eks. ved å sette inn avstandsholdende nukleotider mellom fragmentene). Regulerende elementer kan også arrangeres slik at de tilveiebringer optimal avstandsfordeling av en EESYR i forhold til de regulerende elementene.

[0083] EESYR-sekvensene beskrevet her var isolert fra CHO-celler. Homologe ekspresjonsforsterkende elementer er forventet å eksistere i celler fra andre pattedyrarter (som for eksempel mennesker; se fig. 3) samt i cellelinjer derivert fra andre vevtyper, og kan isoleres ved hjelp av velkjent teknikk, for eksempel hybridisering mellom arter eller PCR-baserte teknikker. I tillegg kan det gjøres endringer i nukleotidsekvensen presentert i SEKV ID NR:1-6 ved hjelp av setteinrettede eller tilfeldige mutageneseteknikker som er velkjent teknikk. De resulterende EESYR-variantene kan deretter testes for EESYR-aktivitet som beskrevet her. DNA som er minst ca. 80 prosent identisk, fortrinnsvis minst ca. 90 prosent identisk, mer foretrukket minst ca. 95 prosent identisk, mest foretrukket minst ca. 99 prosent identisk i nukleotidsekvens med SEKV ID NR:1-6 eller fragmenter av dette og som har EESYR-aktivitet, er isolerbare ved rutinemessig eksperimentering, og forventes å oppvise EESYR-aktivitet. For fragmenter av EESYR viser prosent identitet til den delen av den refererte native sekvensen som finnes i EESYR-fragmentet. Følgelig er homologer av EESYR og varianter av EESYR også beskrevet her. Fig. 3A og 3B viser en

sammenstilling av sekvenser fra mus, menneske og rotte med varierende homologi til et fragment av SEKV ID NR:5.

[0084] Cellepopulasjoner som eksprimerer forsterkede nivåer av et protein av interesse, kan utvikles ved hjelp av framgangsmåtene tilveiebrakt her. Det absolutte ekspresjonsnivået vil variere med det spesifikke proteinet, avhengig av hvor effektivt proteinet prosesseres av cellen. Cellepooler utviklet ved hjelp av EESYR er stabile over tid, og kan behandles som stabile cellelinjer for de fleste formål. Kloningstrinn kan utsettes til senere i utviklingsprosessen enn det som er vanlig for rekombinerende proteiner.

EESYR-er og ekspresjonsforsterkende fragmenter av disse

[0085] EESYR-ens genomiske locus bevares i genomet hos mennesker, mus og rotter. Fig. 4 viser prosent identitet blant EESYR-sekvenser. EESYR-sekvenser homologe til 13 515 kb av klonet CHO-EESYR-DNA i SEKV ID NR:5 ble identifisert blant det publiserte genomet hos mennesker, rotter og mus ved hjelp av BLAST. Sekvenser ble sammenstilt for å bestemme prosent homologt ved hjelp av MacVector (9.0). Tjuefem bp-inkremitter i sammenstillingen er plottet som prosent identitet blant EESYR-sekvenser fra CHO, menneske, mus og rotte for hvert påfølgende 25 bp-segment. Som vist i fig. 4 markerer den vertikale linjen stedet for setespesifikke rekombinasjonshendelser for å eksprimere rekombinerende proteingener av interesse. Prosent identitet av EESYR-sekvenser tilstøtende et sted for setespesifikk rekombinasjon i en EESYR er vist i panel B i fig. 4. Ti baseparinkremitter i de sammenstilte sekvensene som svarer til nt 5022-6110 i en CHO-celles EESYR-sekvens (nukleotidene 5022 til 6110 i SEKV ID NR:5) er plottet som prosent identitet blant EESYR-sekvenser fra CHO, menneske, mus og rotte for hvert påfølgende 10 bp-segment. Sekvensene ble sammenstilt ved hjelp av MacVector™ 9.0. Som vist i panel B er en signifikant identitet i sekvens til stede i dette fragmentet av EESYR-en klonet fra CHO-celler. Det skal legges merke til at sammenlikningen i fig. 4 indikerer en lengde på ca. 1400 baser, mens sekvensen i SEKV ID NR:5 inneholder 13 515 baser. Basene i fig. 4 ser ut til å strekke seg over en lengre lengde på grunn av tilstedeværelsen av mellomrom. De nevnte nukleotidintervallene er de som svarer til nummereringen i SEKV ID NR:5 med mindre annet er angitt. Nukleotidintervallet fra ca. 6200 til ca. 7600 som vist i fig. 4B svarer til nukleotidene i SEKV ID NR:5 nummerert ca. 5 200 til ca. 6 000.

[0086] Følgelig er her beskrevet et ekspresjonsforsterkende fragment av en nukleotidsekvens i SEKV ID NR:5, der det ekspresjonsforsterkende fragmentet omfatter nukleotidsekvensen indikert av posisjoner omkring residuene 5022 til ca. 6110 i SEKV ID NR:5, eller ca. 5218 til ca. 6048 i SEKV ID NR:5; eller ca. 6200 til ca. 7600, ca. 6500 til ca. 7400, eller ca. 6400 til ca. 6500 vist i panel B i fig. 4. Her er også beskrevet et ekspresjonsforsterkende fragment av en nukleotidsekvens som er minst 80 prosent identisk, fortrinnsvis minst 90 prosent identisk, mer foretrukket minst 95 prosent identisk, mest foretrukket minst 99 prosent identisk med nukleotidsekvensen som er indikert ved posisjonene ca. 6200 til ca. 7600, eller ca. 6500 til ca. 7400, eller ca. 6400 til ca. 6500 vist i panel B i fig. 4. Videre er her beskrevet vektorer som omfatter et slikt fragment, inkludert for transient eller stabil transfeksjon. Ytterligere beskrevet her er en eukaryot celle som

omfatter et slikt fragment, der fragmentet er eksogent og er integrert i cellegenomet, og celler som omfatter et slikt fragment, har minst ett rekombinasegjenkjenningssete som er inne i, umiddelbart 5', eller umiddelbart 3' til fragmentet.

[0087] Et ekspresjonsforsterkende fragment av SEKV ID NR:5 kan befinne seg i en posisjon innenfor SEKV ID NR:5 valgt blant nukleotider som strekker seg over posisjonene med nummer 10-13 515; 20-12 020; 1 020-11 020; 2 020-10 020; 3 020-9 020; 4 020-8 020; 5 020-7 020; 6 020-6 920; 6 120-6 820; 6 220-6 720; 6 320-6 620; 6 420-6 520; 6 460-6 500; 6 470-6 490; og 6 475-6 485.

[0088] I én utførelsesform anvendes EESYR-en til å forsterke ekspresjonen av et GOI, som illustrert i fig. 5. Fig. 5 viser et GOI som er funksjonelt knyttet til en promotor (med en oppstrøms markør som har sin egen promotor) integrert i en ikke-EESYR-posisjon i et CHO-cellegenom, og en FACS-utlesning som viser fordelingen av ekspresjon i en stabilt transfisert cellepopulasjon. Til sammenlikning er vist et GOI som er funksjonelt knyttet til en promotor integrert i en EESYR-posisjon i et CHO-cellegenom, og en FACS-utlesning som viser fordelingen av ekspresjon i en stabilt transfisert cellepopulasjon, er også vist. I denne utførelsesformen viser GOI-et eksprimert i EESYR-locusen en forsterket ekspresjon på omkring to ganger sammenliknet med GOI-et eksprimert i en ikke-EESYR-locus.

[0089] I ulike utførelsesformer kan ekspresjon av et GOI forsterkes ved å plassere GOI-et i en EESYR, 5' til en EESYR, eller 3' til en EESYR. Den nøyaktige avstanden mellom GOI-et og EESYR-en, der GOI-et er enten 5' eller 3' til EESYR-en, skal være slik at EESYR-en er funksjonelt knyttet til GOI-et. En EESYR er funksjonelt knyttet til GOI-et der ekspresjon av GOI-et – ved den valgte avstanden fra EESYR-en (i 5'- eller 3'-retningen) – beholder evnen til å forsterke ekspresjon av GOI-et over for eksempel ekspresjon typisk observert på grunn av en tilfeldig integrasjonshendelse. I ulike utførelsesformer er forsterkningen minst omkring 1,5 ganger til omkring 2 ganger eller mer. Fortrinnsvis er forsterkningen i ekspresjon sammenliknet med en tilfeldig integrasjon, eller tilfeldig ekspresjon, omkring 2 ganger eller mer.

[0090] Fig. 6 viser en utførelsesform der SEKV ID NR:1 («EESYR 5'») og SEKV ID NR:2 («EESYR 3'») er sammenliknet i sin relative evne til å forsterke ekspresjon av et funksjonelt tilknyttet GOI, der GOI-et er funksjonelt knyttet til en promotor også (en markør og en promotor funksjonelt knyttet til markøren er vist 5' til GOI-promotoren). Orientering av EESYR-en er vist ved retningen av pilen under termen «EESYR». Konstruktene er tilfeldig integrert i et CHO-cellegenom. Ekspresjon er relativ til det tilfeldig integrerte konstruktet som ikke omfatter noen EESYR. FACS-utlesninger som viser relativ ekspresjon, er vist til høyre. I fig. 6 anvender det første EESYR-konstruktet SEKV ID NR:1; det andre EESYR-konstruktet anvender SEKV ID NR:3; det tredje EESYR-konstruktet er SEKV ID NR:2; det fjerde EESYR-konstruktet er SEKV ID NR:4. Som vist i figuren oppviser SEKV ID NR:3 en 2,4 ganger forsterkning i ekspresjon, og SEKV ID NR:1 oppviser en 2 ganger forsterkning av ekspresjon.

[0091] EESYR-rekombinerende cellepooler oppviser kloningsegenskaper. Fig. 7 illustrerer EESYR-rekombinantpoolers kloningsegenskaper. I det tofargede FACS-plottet som representerer et toparameterhistogram av celler merket med røde eller grønne markører (røde celler er vert-CHO-celler; grønne celler er rekombinanter som eksprimerer et GOI),

viser EESYR-rekombinantpooler klyngedannelse i plottet som gjenspeiler vesentlig identisk vekst og ekspresjon, strømningscytometriprofil og Southern-analyse (ikke vist). Et histogram av ELISA-er av enkeltceller viser uniform ekspresjon i alle kloner. Kloningsstabiliteten er også høy etter tre måneder uten utvelgelse.

[0092] EESYR-rekombinanter gjennomgår spesifikk og effektiv rekombinasjon, som vist i fig. 8. Panel A viser to markører separert av en IRES og flankert av rekombinasegjenkjenningsseter, og en tredje markør som ikke er flankert av rekombinasegjenkjenningsseter som en tilfeldig integrasjonskontroll. Når det rekombineres ved en EESYR-locus som omfatter en markør flankert av to rekombinasegjenkjenningsseter, er rekombinasjonen spesifikk. Panel B viser lite tilfeldig integrasjon ved fravær av rekombinase, men effektiv og setespesifikk integrasjon ved tilstedeværelsen av rekombinase.

[0093] Tilfeldig integrasjon ved hjelp av setespesifikk rekombinasjon ved en EESYR er sjelden (se fig. 9). Fig. 9 viser at når tilfeldige integrasjonshendelser visualiseres, representerer slike hendelser bare en ørliten del av integrasjonshendelsene.

[0094] EESYR-cis-fungerende elementer kan vurderes ved hjelp av framgangsmåtene og sammensetningene beskrevet her. Som vist i fig. 10 tillater EESYR-rekombinerende celler sammenlikning av cis-fungerende elementer ekvivalent. Ettersom EESYR-rekombinanter oppfører seg som en kloningspopulasjon, kan forskjeller i genekspresjon som følge av for eksempel tilstedeværelsen eller fraværet av mistenkte cis-fungerende elementer sammenliknes direkte. Isogene cellelinjer tillater direkte sammenlikning av *cis*-fungerende elementer. Ved bruk av et EESYR-system befinner *cis*-fungerende elementer som for eksempel promotorer, introner og UTR-er seg fortrinnsvis mellom rekombinasjonssetene. Ekspresjonsoptimalisering kan også oppnås, inkludert for eksempel ekspresjonskassettorientering og kodonoptimalisering. Som et eksempel viser fig. 11 kassetter flankert av rekombinasjonsgjenkjenningsseter som inneholder en promotor, en markør, ulike cis-elementer (her introner i panel A; UTR-er i panel B), og et GOI var integrert i en EESYR (SEKV ID NR: 1 i 5'-enden, SEKV ID NR:2 i 3'-enden). Relativ ekspresjon av GOI-et er vist til høyre.

[0095] Fig. 12 viser et eksempel på hvordan proteinoptimalisering kan oppnås ved hjelp av framgangsmåtene og sammensetningene beskrevet her. Fig. 12 bekrefter at opsjonell plassering av et cDNA for et lettkjedeantistoffgen er 5' til et cDNA for et tungkjedeantistoffgen.

Proteiner av interesse

[0096] En nukleinsyresekvens som koder et protein av interesse, kan bekvemt integreres i en celle som omfatter en EESYR som har et rekombinasegjenkjenningssete, gjennom for eksempel en rekombinaseformidlet kassetutvekslingsprosess (RMCE). Et hvilket som helst protein av interesse som er egnet for ekspresjon i eukaryote celler kan brukes. For eksempel kan proteinet av interesse være et antistoff eller fragment av dette, et kimært

antistoff eller fragment av dette, en ScFv eller fragment av dette, et Fc-merket protein eller fragment av dette, en vekstfaktor eller et fragment av dette, et cytokin eller et fragment av dette, eller et ekstracellulært domene av en celleoverflatereseptor eller fragment av dette.

Nukleinsyrekonstrukter

[0097] Rekombinerende ekspresjonsvektorer kan omfatte syntetiske eller cDNA-deriverte DNA-fragmenter som koder et protein, funksjonelt knyttet til et egnet transkripsjons- og/eller translasjonsregulerende element derivert fra pattedyr-, virus- eller insektgener. Slike regulerende elementer omfatter transkripsjonspromotorer, forsterkere, sekvenser som koder egnede mRNA-ribosombindende seter, og sekvenser som kontrollerer termineringen av transkripsjon og translasjon, som beskrevet i detalj nedenfor. Ekspresjonsvektorer hos pattedyr kan også omfatte ikke-transkriberte elementer så som en replikasjonsstart, andre 5'- eller 3'-flankerende ikke-transkriberte sekvenser, og 5'- eller 3'-ikke-translaterte sekvenser så som donor- og akseptorseter for sammenbinding. Et utvelgbart markørgen for å lette gjenkjennelse av transfektanter kan også inkorporeres.

[0098] Transkripsjons- og translasjonskontrollsekvenser i ekspresjonsvektorer som kan brukes til å transifisere vertebratceller, kan tilveiebringes av viruskilder. For eksempel leveres vanlig benyttede promotorer og forsterkere fra virus så som polyoma, adenovirus 2, simian virus 40 (SV40), og humant cytomegalovirus (CMV). Virusgenompromotorer, kontroll- og/eller signalsekvenser kan benyttes til å drive ekspresjon, forutsatt at slike kontrollsekvenser er compatible med vertcellen som er valgt. Ikke-virale cellepromotorer kan også brukes (f.eks. β -globinet og EF-1 α -promotorene), avhengig av celletypen det rekombinerende proteinet skal eksprimeres i.

[0099] DNA-sekvenser derivert fra SV40-virusgenomet, for eksempel SV40-starten, tidlige og sene promotor-, forsterker-, sammenbindings- og polyadenylasjonsseter kan brukes til å tilveiebringe andre genetiske elementer som kan brukes til å eksprimere en heterolog DNA-sekvens. Tidlige og sene promotorer er særlig nyttige ettersom begge enkelt framskaffes fra SV40-viruset som et fragment som også omfatter SV40-virusreplikasjonsstarten (Fiers *et al.*, Nature 273:113, 1978). Mindre eller større SV40-fragmenter kan også brukes. Typisk er den ca. 250 bp-sekvensen som strekker seg fra Hind III-setet mot BglI-setet lokalisert i SV40-replikasjonsstarten, inkludert.

[0100] Bicistroniske ekspresjonsvektorer brukt til ekspresjon av flere transkripter er tidligere blitt beskrevet (Kim S. K. and Wold B. J., Cell 42:129, 1985; Kaufman *et al.* 1991, supra) og kan brukes i kombinasjon med en EESYR-sekvens som er beskrevet her. Andre typer ekspresjonsvektorer vil også være nyttige, for eksempel de som er beskrevet i U.S. Pat. nr. 4 634 665 (Axel *et al.*) og U.S. Pat. nr. 4 656 134 (Ringold *et al.*).

Vertceller og transfeksjon

[0101] De eukaryote vertcellene som benyttes i oppfinnelsens framgangsmåter er pattedyrvertceller som omfatter for eksempel CHO-celler og museceller. Her er beskrevet en nukleinsyresekvens som koder en EESYR-sekvens fra CHO-celle. Et integrasjonssete, for eksempel et rekombinasegjenkjenningssete, kan være plassert i en EESYR, eller 5' eller 3' til EESYR-sekvensen. Ett eksempel på et egnet integrasjonssete er et loxP-sete. Et annet eksempel på et egnet integrasjonssete er to rekombinasegjenkjenningsseter, for eksempel et loxP-sete og et lox 5511-sete. EESYR-sekvensen kan befinne seg på kromosom 6 i et CHO-cellegenom. EESYR-sekvensen kan befinne seg i en sekvens valgt fra gruppen som består av nukleinsyrer som omfatter nukleotidene 1-6473 og 4607-6473 i SEKV ID NR: 1; og 1-7045, 1-3115, 1-2245, 1-935 og 1-465 i SEKV ID NR:2.

[0102] Her er også beskrevet en pattedyrvertcelle transfisert med en ekspresjonsvektor beskrevet her. Selv om en hvilken som helst pattedyrcelle kan brukes, er vertcellen fortrinnsvis en CHO-celle.

[0103] Transfiserte vertceller omfatter celler som er blitt transfisert med ekspresjonsvektorer som omfatter en sekvens som koder et protein eller polypeptid. Eksprimerte proteiner vil fortrinnsvis være sekretert inn i kulturmediet, avhengig av nukleinsyresekvensen som er valgt, men kan holdes tilbake i cellen eller deponeres i cellemembranen. Ulike cellekultursystemer fra pattedyr kan anvendes for å eksprimere rekombinerende proteiner. Eksempler på egnede pattedyrvertcellelinjer omfatter COS-7-linjene med apenyreceller, beskrevet av Gluzman (1981) Cell 23:175, og andre cellelinjer som er i stand til å eksprimere en egnet vektor, inkludert for eksempel CV-1/EBNA (ATCC CRL 10478), L-celler, C127, 3T3, CHO, HeLa og BHK-cellelinjer. Andre cellelinjer utviklet for spesifikke utvelgelses- eller amplifikasjonsskjemaer vil også kunne brukes med framgangsmåtene og sammensetningene som er tilveiebrakt her. En foretrukket cellelinje er CHO-cellelinjen kalt K1. For å oppnå målet om storvolumproduksjon av rekombinerende proteiner bør vertcellelinjen forhåndstilpasses bioreaktormediet i det enkelte tilfellet.

[0104] Flere transfeksjonsprotokoller er kjent teknikk, og er gjennomgått i Kaufman (1988) Meth. Enzymology 185:537. Transfeksjonsprotokollen som velges vil være avhengig av vertcelletypen og GOI-ets natur, og kan være valgt på grunnlag av rutinemessig eksperimentering. De grunnleggende kravene til slike protokoller er å først sette inn DNA som koder proteinet av interesse i en egnet vertcelle, og deretter å identifisere og isolere vertceller som har inkorporert det heterologe DNA-et på en relativt stabil, eksprimerbar måte.

[0105] Én vanlig benyttet framgangsmåte for å sette inn heterologt DNA i en celle er kalsiumfosfatpresipitering, for eksempel som beskrevet av Wigler et al. (Proc. Natl. Acad. Sci. USA 77:3567, 1980). DNA som settes inn i en vertcelle ved hjelp av denne framgangsmåten, undergår ofte omarrangering, noe som gjør denne prosedyren nyttig for kotransfeksjon av uavhengige gener.

[0106] Polyetylenindusert fusjon av bakterielle protoplaster med pattedyrceller (Schaffner et al., (1980) Proc. Natl. Acad. Sci. USA 77:2163) er en annen nyttig framgangsmåte for å

sette inn heterologt DNA. Protoplastfusjonprotokoller gir ofte flere kopier av plasmid-DNA-et integrert i pattedyrvertcellegenomet, og denne teknikken krever at utvelgelses- og amplifikasjonsmarkøren er på samme plasmid som GOI-et.

[0107] Elektroporasjon kan også benyttes til å sette DNA direkte inn i cytoplasmaet til en vertcelle, for eksempel som beskrevet av Potter et al. (Proc. Natl. Acad. Sci. USA 81:7161, 1988) eller Shigekawa et al. (BioTechniques 6:742, 1988). I motsetning til protoplastfusjon krever ikke elektroporasjon at utvelgelsesmarkøren og GOI-et er på samme plasmid.

[0108] I det senere er det blitt beskrevet flere reagenser som kan brukes til å sette heterologt DNA inn i en pattedyrcelle. Dette omfatter Lipofectin™-reagens og Lipofectamine™-reagens (Gibco BRL, Gaithersburg, Md.). Begge disse reagensene er kommersielt tilgjengelige reagenser som brukes til å danne lipid-nukleinsyrekomplekser (eller liposomer) som, når de anvendes på dyrkede celler, letter opptaket av nukleinsyren i cellene.

[0109] En framgangsmåte for å amplifisere GOI-et er også ønskelig for å eksprimere det rekombinerende proteinet, og involverer typisk bruken av en utvelgelsesmarkør (gjennomgått i Kaufman *supra*). Resistens mot cytotoxiske legemidler er den egenskapen som oftest benyttes som utvelgelsesmarkør, og kan være resultatet av enten et dominant anlegg (kan f.eks. brukes uavhengig av vertcelletype) eller et recessivt anlegg (kan f.eks. brukes i visse vertcelletyper som har mangler i den aktiviteten det utvelges for). Flere amplifiserbare markører egner seg til bruk i oppfinnelsens ekspresjonsvektorer (f.eks. som beskrevet i Maniatis, Molecular Biology: A Laboratory Manual, Cold Spring Harbor Laboratory, NY, 1989; s. 16.9-16.14).

[0110] Nyttige utvelgbare markører for genamplifikasjon i legemiddelresistente pattedyrceller er vist i tabell 1 hos Kaufman, R. J., *supra*, og omfatter DHFR-MTX-resistens, P-glykoprotein og multiplert legemiddelresistens (MDR) – ulike lipofile cytotoxiske virkestoffer (f.eks. adriamycin, colchicin, vincristin), og adenosindeaminase (ADA)-Xyl-A eller adenosin og 2'-deoksykoformycin.

[0111] Andre dominante utvelgbare markører omfatter mikrobielt deriverte antibiotikaresistensgener, for eksempel neomycin-, kanamycin- eller hygromycinresistens. Disse utvelgelsesmarkørene er imidlertid ikke blitt vist å være amplifiserbare (Kaufman, R. J., *supra*,). Flere egnede utvelgelsessystemer finnes for pattedyrverter (Maniatis *supra*, s. 16.9-16.15). Kotransfeksjonsprotokoller som anvender to dominante utvelgbare markører er også blitt beskrevet (Okayama and Berg, Mol. Cell Biol 5:1136, 1985).

[0112] Nyttige regulerende elementer som er tidligere beskrevet eller kjent teknikk, kan også inkluderes i nukleinsyrekonstruktene som benyttes til å transifisere pattedyrceller. Transfeksjonsprotokollen som velges og elementene som utvelges til bruk der, vil være avhengig av typen vertcelle som brukes. Fagpersonen vil kjenne mange ulike protokoller og vertceller, og kan velge et egnet system for å eksprimere et ønsket protein ut ifra kravene til cellekultursystemet som brukes.

[0113] Andre trekk ved oppfinnelsen vil bli åpenbare i løpet av de følgende beskrivelsene

av eksemplariske utførelsesformer som er gitt for å illustrere oppfinnelsen og ikke er ment å avgrense den.

EKSEMPLER

[0114] De følgende eksemplene er presentert for å vise den ordinære fagpersonen hvordan framgangsmåtene og sammensetningene her skal lages og brukes, og er ikke ment å avgrense omfanget av hva oppfinnerne anser som sin oppfinnelse. Det er gjort forsøk på å sikre nøyaktig bruk av tall (f.eks. mengde, temperatur osv.) men det må tas høyde for noen eksperimentelle feil og avvik. Med mindre annet er angitt, er andeler vektandeler, molekylvekt er gjennomsnittlig molekylvekt, temperatur er i grader Celsius, og trykk er ved eller nær atmosfærisk.

Eksempel 1. Generering av RGC9- og RGC16-cellelinjer.

[0115] CHO-K1-celler (1×10^7) ble infisert med pantropisk retrovirus produsert med plasmid pTE252 (fig.1), med et loxP-sete i seg, ved en MOI på mindre enn 1 for å generere en stabil pool av celler med for det meste én retroviral innsetting per celle. Celler i den stabile poolen som eksprimerte et markørprotein ved et høyt nivå, ble utvalgt og ekspandert. Trettiseks kloner ble isolert og ekspandert ut i 36 cellelinjer. Kloner som oppviste den høyeste ekspresjons- og rekombinasjonseffektiviteten ble identifisert og klonet, der hver av klonene inneholdt minst ett rekombinasegjenkjenningssete i en forsterket ekspresjonslocus. Åtte cellepopulasjoner med den beste ekspresjons- og rekombinasjonseffektiviteten ble utvalgt, og revurdert for forsterket proteinekspresjon. To cellepopulasjoner ble utvalgt og benevnt RGC9 og RGC16. Southern blot-analyse av cellepopulasjonene fra de opprinnelige 36 cellelinjene som svarte til disse to cellelinjene, viste at en enkelt kopi av et rapportørkonstrukt var integrert i CHO-cellegenomet ved samme locus for begge cellepopulasjoner, og integrasjonens lokalitet ble bestemt å være ved trippel-«act»-en ved nukleotidposisjon 6 471 til 6 473 i SEKV ID NR:5. Minst én av disse to cellelinjene ble anvendt i eksperimenter beskrevet nedenfor.

Eksempel 2. Ekspresjon av FcFP1-protein i serumfritt produksjonsmedium

[0116] RGC38-celler ble derivert fra RGC9-celler og ble tilpasset til å vokse i suspensjon i et serumfritt produksjonsmedium. RGC38-celler ble brukt som vertceller for ekspresjonen av FcFP1-protein (Fc-fusjon-protein-1). RGC38-celler ble transisert i en ti-centimeters skål med en FcFP1-ekspresjonsvektor, pTE851 og et Cre-plasmid, pRG858. FcFP1-plasmidet har, i 5'- til 3'-retning, et LoxP-sete, en SV40-senpromotor, et hygromycinresistent gen, en IRES, en eGFP, en CMV-MIE-promotor, et gen som koder et FcFP1-protein, og et

Lox511-sete. Celler ble dyrket i F12-medium med 400 µg/ml hygromycin i to uker etter transfeksjon. Celler som eksprimerte eGFP men ikke DsRed, ble isolert ved hjelp av strømningscytometri og benevnt som RS421-1. Isolerte celler var vesentlig isogene, selv om de var derivert fra ulike morceller. RS421-1-celler ble ekspandert i suspensjonskulturer i serumfritt produksjonsmedium. FcP1-protein i kondisjonert medium med 3 dager gamle kulturer ble undersøkt i SDS-PAGE med Coomassie-blått. FcF1-protein i det kondisjonerte mediet var rikelig til stede og kunne sees uten rensing.

Eksempel 3. Regulert ekspresjon av FcFP1-protein i serumfritt produksjonsmedium

[0117] RGC49-celler ble derivert fra RGC16, ble tilpasset til å vokse i serumfritt produksjonsmedium, og inneholdt et stabilt integrert tetR-YFP-ekspresjonsplasmid, og pcDNA6/TR. TetR-proteinet tillater regulering av transkripsjon fra promotorer som omfatter en tet-operatorsekvens, ved hjelp av tetracyklin eller doksycyklin. RGC49-celler ble kotransfisert med pTE851 og pRG858. De transfiserte cellene ble utvalgt med 400 µg/ml hygromycin i to uker. Celler som eksprimerte eGFP men ikke YFP, ble isolert ved hjelp av strømningscytometri og benevnt som RS569-1. RS569-1-celler ble ekspandert i suspensjonskulturer i serumfritt produksjonsmedium ved tilstedeværelse eller fravær av doksycyklin. FcP1-protein i kondisjonert medium med 3 dager gamle kulturer ble undersøkt med SDS-PAGE og Coomassie-blått. RS569-1-cellene eksprimerte FcFP1-protein på samme måte som RS421-1 ved tilstedeværelse av 1 µg/ml doksycyklin i kulturmediet. Svært lite FcFP1-protein ble detektert fra RS569-1-cellene i fraværet av doksycyklin.

Eksempel 4. Bruk av et tredje Lox-sete (Lox2272) ved EESYR for å skape dualekspresjonskasset

[0118] Celler fra en hvilken som helst cellelinje som bærer et DsRED-gen flankert av et LoxP-sete og et Lox511-sete ved EESYR-locus transfiseres med pRG858 og en vektor som omfatter, i 5'- til 3'-retning, et LoxP-sete, en første promotor, en YFP, et Lox2272-sete, en andre promotor, en eGFP, og en Lox511. Celler som eksprimerer eGFP og YFP, men ikke DsRed, isoleres. Isolerte celler transfiseres deretter med pRG858, pRG1167 (en vektor som har, i 5'- til 3'-retning, et LoxP-sete, en SV40-senpromotor, et hygromycinresistent gen, en CMV-MIE-promotor, en DsRed og et Lox2272-sete) eller pRG1234 (en vektor som har, i 5'- til 3'-retning, et Lox2272-sete, en SV40-senpromotor, et hygromycinresistent gen, en CMV-MIE-promotor, en DsRed og et Lox511-sete). Celler som er i stand til å eksprimere noen av DsRED og eGFP, men ikke YFP, eller DsRED og YFP, men ikke eGFP, isoleres.

Eksempel 5. Antistoffekspresjon fra RGC38-vertceller

[0119] RGC38-celler ble transfisert med pTE963 og pRG858. pTE963 har, i 5'- til 3'-retning, et LoxP-sete, en SV40-senpromotor, et hygromycinresistent gen, en IRES, en eGFP, en CMV-MIE-promotor, et antistoffs lettkjedegen, en CMV-MIE-promotor, et antistoffs tungkjedegen, og et Lox511-sete. De transfiserte kulturene ble utvalgt med hygromycin ved 400 µg/ml hver i to uker. Celler som eksprimerte eGFP men ikke dsRed, ble isolert ved hjelp av strømningscytometri. De isolerte cellene ble ekspandert i suspensjon i serumfritt produksjonsmedium og ble benevnt RS533-celler. For produksjonen av antistoffer ble RS533-celler dyrket i en bioreaktor i 10 dager. Alikvoter av brukt medium fra dag seks til dag ti ble oppsamlet, og proteinsammensetningen deres ble analysert med SDS-PAGE. Tungkjede- og lettkjedepeptidet til antistoffet i det bioreaktorbrukte mediet ble lett detektert ved hjelp av Coomassie-blått.

Eksempel 6. Uthenting og subkloning av EESYR-sekvenser

[0120] En CHO-cellelinje (benevnt RGC21) som eksprimerte høye nivåer av et rapportørgen, DsRed, ble utvalgt for isolasjon av EESYR-sekvenser, ettersom Southern blot-analyse indikerte at den høye ekspresjonen av DsRed-ekspresjon observert for denne cellelinjen er drevet av en enkelt integrasjon av en ekspresjonskassetten som koder DsRed. Genomsekvenser 5' til ekspresjonskassetten ble uthentet ved å transfisere RGC21-celler med lineariserte pTE494-plasmider, en vektor som har, i 5'- til 3'-retning, et LoxP-sete, et ampicillinresistensgen, en bakteriell replikasjonsstart, et multippelt kloningssete (EcoRI, BglII, XbaI, ApaI), en CMV-MIE-promotor, et neomycin-fosfotransferasegen, en IRES, en eGFP og et Lox511-sete. Celler som eksprimerte eGFP men ikke DsRed, ble isolert ved hjelp av strømningscytometri som en pool. Genomisk DNA ble isolert, åpnet med XbaI-restriksjonsendonuklease, og selvligerte slik at det dannet pRG1106. Genomsekvenser 3' til ekspresjonskassetten ble uthentet ved å transfisere RGC21-celler med sirkulære pTE495-plasmider, en vektor som har, i 5'- til 3'-retning, et LoxP-sete, en CMV-MIE-promotor, et neomycin-fosfotransferasegen, en IRES, en eGFP, en bakteriell replikasjonsstart, et multippelt kloningssete (EcoRI, BglII, XbaI, ApaI), et ampicillinresistensgen, og et Lox511-sete. Celler som eksprimerte eGFP men ikke DsRed, ble isolert ved hjelp av strømningscytometri som en pool. Genomisk DNA ble isolert, åpnet med MfeI-restriksjonsendonuklease, og selvligerte slik at det dannet pRG1099.

Eksempel 7. Plasmidkonstruksjon for EESYR-analyse

[0121] EESYR-sekvenser ble eksidert fra enten pRG1106 eller pRG1099 som Agel-fragmenter og satt inn i Agel- og NgoMIV-setene på pTE575, et plasmid som eksprimerer FCFP2, for å gi plasmid pTE809. pTE575-plasmidet har, i 5'- til 3'-retning, en SV40-senpromotor, et hygromycinresistent gen, en CMV-MIE-promotor, og et gen som koder

FCFP2-protein. I stabilt transifiserte CHO-celler ga pTE809 og pTE575 at henholdsvis 97,22 % og 38,57 % av cellene eksprimerte detekterbare nivåer av FCFP2-protein. Gjennomsnittlig fluorescens hos FcFP2 detektert av et FITC-konjugert antistoff var 482,54 og 279,75 for kulturer transifisert med henholdsvis pTE809 og pTE575. Dermed økte inklusjonen av EESYR i ekspresjonsvektorer ekspresjonen av FCFP2-protein i stabil transfeksjon.

Eksempel 8. Uthenting og subkloning av EESYR-sekvenser

[0122] En CHO-cellelinje (benevnt RGC21) som eksprimerte høye nivåer av et rapportørgen, DsRed, ble utvalgt for isolasjon av EESYR-sekvenser, ettersom Southern blot-analyse indikerte at den høye ekspresjonen av DsRed-ekspresjon observert for denne cellelinjen er drevet av en enkelt integrasjon av en ekspresjonskassetten som koder DsRed. Genomsekvenser 5' til ekspresjonskassetten ble uthentet ved å transifisere RGC21-celler med lineariserte pTE494-plasmider, en vektor som har, i 5'- til 3'-retning, et LoxP-sete, et ampicillinresistensgen, en bakteriell replikasjonsstart, en CMV-MIE-promotor, et neomycin-fosfotransferasegen, en IRES, en eGFP og et Lox511-sete. Celler som eksprimerte eGFP men ikke DsRed, ble isolert ved hjelp av strømningscytometri som en pool. Genomisk DNA ble isolert, åpnet med XbaI-restriksjonsendonuklease, og selvligerte slik at det dannet pRG1106. Genomsekvenser 3' til ekspresjonskassetten ble uthentet ved å transifisere RGC21-celler med sirkulære pTE495-plasmider, en vektor som har, i 5'- til 3'-retning, et LoxP-sete, en CMV-MIE-promotor, et neomycin-fosfotransferasegen, en IRES, en eGFP, en bakteriell replikasjonsstart, et ampicillinresistensgen, og et Lox511-sete. Celler som eksprimerte eGFP men ikke DsRed, ble isolert ved hjelp av strømningscytometri som en pool. Genomisk DNA ble isolert, åpnet med MfeI-restriksjonsendonuklease, og selvligerte slik at det dannet pRG1099.

Eksempel 9. Plasmidkonstruksjon for EESYR-analyse

[0123] EESYR-sekvenser ble eksidert fra enten pRG1106 eller pRG1099 som Agel-fragmenter og satt inn i Agel- og NgoMIV-setene på pTE575, et plasmid som eksprimerer FCFP2, for å gi plasmid pTE809. pTE575-plasmidet har, i 5'- til 3'-retning, en SV40-senpromotor, et hygromycinresistent gen, en CMV-MIE-promotor, og et gen som koder FCFP2-protein. I stabilt transifiserte CHO-celler ga pTE809 og pTE575 at henholdsvis 97,22 % og 38,57 % av cellene eksprimerte detekterbare nivåer av FCFP2-protein. Gjennomsnittlig fluorescens hos FcFP2 detektert av et FITC-konjugert antistoff var 482,54 og 279,75 for kulturer transifisert med henholdsvis pTE809 og pTE575. Dermed økte inklusjonen av EESYR i ekspresjonsvektorer ekspresjonen av FCFP2-protein i stabil transfeksjon.

LISTE OVER SEKVENSER**[0124]**

<110> Regeneron Pharmaceuticals, Inc.

<120> Regioner med forsterket ekspresjon og stabilitet

<130> 3051A

<140> Skal tildeles

< 141> 2008-06-04

<150> 60/933 213

< 151> 2007-06-04

<160> 8

<170> FastSEQ for Windows versjon 4.0

<210> 1

< 211> 6473

< 212> DNA

< 213> Cricetulus griseus

<400> 1

```

tctagaaaca aaacaaaaaa tattaagtca ggcttggtt caggtgctgg ggtggagtgc 60
tgacaaaaat acacaaattc ctggctttct aaggcttttt cggggattca ggtattgggt 120
gatggtagaa taaaaatctg aaacataggt gatgtatctg ccatactgca tgggtgtgta 180
tgtgtgtgta tgtgtgtctg tgtgtgtgcc cagacagaaa taccatgaag gaaaaaaca 240
cttcaaagac aggagagaag agtgacctgg gaaggactcc ccaatgagat gagaactgag 300
cacatgccag aggaggtgag gactgaacca ttcaacacaa gtggtgaata gtccctgcaga 360
cacagagagg gccagaagca ctcagaactc cagggggtca ggagtgggtc tctggaggct 420
tctgcccttg gaggttcctg aggaggaggc ttccatattg aaaatgtagt tagtggcctg 480
ttccattagt acagtgacta gagagagctg agggaccact ggactgaggc ctagatgctc 540
agtcagatgg ccattgaaagc ctagacaagc acttccgggt ggaaaggaaa cagcaggtgt 600
gaggggtcag gggcaagtta gtgggagagg tcttccagat gaagtagcag gaacggagac 660
gcaactggatg gccccacttg tcaaccagca aaagcttggg tcttgttcta agaggccagg 720
gacatgacaa ggggtgatctc ggttttttaa aggctttgtg ttacctaatc acttctatta 780
gtcagatact ttgtaacaca aatgagtact tggcctgtat ttagaaactc tctgggatcc 840
tgaaaaaaca caatgacatt ctggctgcaa cactggaga ctcccagcca ggccctggac 900
ccgggtccat tcatgcaaat actcagggac agattcttca ctagtactg atgactgtc 960
ttggatgcaa atgtggcctc ttcattttac tacaagtcac catgagtcag gaggtgctgt 1020
ttgcacagtg tgactaagtg atggagtgtt gactgcagcc attcccggcc ccagcttgtg 1080
agagagatcc ttttaattg aaagtaagct caaagttacc acgaagccac acatgtataa 1140
actgtgtgaa taatctgtgc acatacaca accatgtgaa taatctgtgt acatgtataa 1200
actgtgtgaa taatctgtgt gcagccttct cttaacctact acctccagt gatcagggtt 1260
ggactgctg tgtgtactg gacctgaat gtccccaccg ctgtccctg tcttttacga 1320
ttctgacatt ttttaataat tcagcggctt cccctctgct ctgtgcctag ctataccttg 1380
gtactctgca ttttggtttc tgtgacattt ctctgtgact ctgtacatt ctacagatgac 1440
atgtgacaca gaagggtgtt cctctggaga catgtgatgt cctgtcatt agtggaatca 1500
gatgccccca aactgttgtc cagtgtttgg gaaagtgaca cgtgaaggag gatcaggaaa 1560
agaggggtgg aaatcaagat gtgtctgagt atctcatgtc cctgagtggg ccaggctgct 1620
gacttcactc cccaagtga gggaggccat ggtgagtaca cacacctcac acatactata 1680
tccaacacac acacacacac acacacacac acgcacgcac gcacgcacgc acgcacacat 1740
gcacacacac gaactacatt tcacaaacca catacgcata ttacacccca aacgtatcac 1800
ctatacatat cacacatata caccctcca cacatcacac acataccaca cccacacaca 1860
gcacacacat acataggcac acattcacac accacacata tacatttgtg tatgcataca 1920
tgcatacaca cacaggcaca cagacaccac acacatgcat tgtgtacgca cacatgcata 1980
cacacacata ggcacacatt gagcacacac atacatttgt gtacgcacac tacatagaca 2040
tatatgcatt tgtatatgca cacatgcatg cacacatata taggcacaca tagagcacac 2100
acatacatct gtgtatgcac acatgcacac accaatcaca tgggaagact caggttcttc 2160

```

actaagggttc	acatgaactt	agcagttcct	ggttatctcg	tgaaacttgg	aagattgctg	2220
tggagaagag	gaagcggttg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggtttaaaaa	tqagagggtc	tcaatgtgga	acccgcaggg	cgccagtlca	gagaagagac	2340
ctacccaagc	caactlgagag	caaaggcaga	gggatgaacc	tyggatgtag	tttgaacctc	2400
tgtaccagct	gggtctcatg	ctattttgtt	atatctttat	taaataattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaatgt	atgggcactg	tatgtgttct	tgggtgccgg	2520
ggaggccagg	agagggcattg	gatcctccgg	agctggcgtt	tgagacagtt	gtgacccaca	2580
gtgtggggtc	tgggaactgg	gtcttagtgt	tccgcaagtg	cagctggggc	tcttaacctc	2640
tgagccatcc	ctccagcttc	aagaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggcttttaggc	ttgtttgttc	agcaaatact	cttttcgtgt	attttgaatt	ttattttatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggcl	gggcctgaac	tcatcaaaal	2820
cctcctgtct	cagctctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
ctttagagtc	cccagccctt	ctggacactt	gttccaagta	taatatatat	atatatatat	2940
atatatatat	atatatatat	atatattgtg	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgcttttaat	ttatttttta	attaaaaatg	3060
attagctaca	tgtcacctgt	atgctctgt	atcatctata	tatccttctt	tccttctctc	3120
tctttctctc	ttcttctctc	cacccccaag	catctatttt	caaatccttg	tgccgaggag	3180
atgccaaagag	tctcgttggg	ggagatggtg	agggggcgat	acagggggaag	agcaggagga	3240
aagggggaca	gactggtgtg	ggtcttttga	gagctcagga	gaatagcagc	gatcttccct	3300
gtcccttggtg	tcacclctla	cagccaacac	cattttgttg	cctggcagaa	gagttgtcaa	3360
gctygtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctggggt	taaaaggcgt	gcctggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaataat	acccgcagag	tcacagtccc	aggttcccg	gctttctctga	3540
agcgccaggc	acgggttccc	taggaaatgg	ggccttgctt	gccaaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcgggc	gcgtcacac	aaacccgcaa	cgatttgtta	3720
aacactctcc	attgagaatc	tggagtgcgg	ttgccctcta	ctggggagct	gaagacagct	3780
agtggggqcg	gggggaggac	cgtgclagca	tccttccacg	glgctcgctg	gclgtggllc	3840
atgccgggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttgggtggaa	ctgacaatca	3900
acgaaatcac	ttcgattggt	ttcctctttt	tactggaatt	cttggatttg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
ggggaggagg	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaagattc	ggagaaagtg	caccgctagt	gaccgggagg	aggaatgccc	tattgggcat	4140
tatatccctt	gtcgtctaatt	ggaatcaaac	tcttggttcc	agcaccagg	attctgagcc	4200
tatcctattc	aagacagtaa	ctacagccca	cacggaagag	gctatacaac	tgaagaaata	4260
aaattttcac	tttatttcat	ttctgtgact	gcatgttcac	atgtagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagtagagtt	ctgagcccgt	taactggaac	aaccagaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttatttagag	acagggtctc	atactgtagt	cgaagctagc	ctccagctca	cagaaattct	4500
cctgttccgg	tttccaaagt	actggagtta	tgaagtgtgt	tttaattgaac	gctaagaatt	4560
tgtcgtattg	agaaaacctc	aagtgggttt	ggctaattcc	cacgacccca	gaggtgagg	4620
caggagggaat	gagagaattc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgagggtg	agctcaatgt	ggagactgtg	aggggtgagc	caatgtggag	actgtgaggg	4740
tgagctcaat	gtggagactg	tgagggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtgtgt	4860
tgaagtgttag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccggctc	atctglggcc	4920
ctgcaacag	aagggagggg	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagatagg	aatgagagag	gaaggaagaa	acgggaagaa	attcagactc	cttctgagt	5040
tccgccaacg	cctagtgaca	tcctgtgnac	accct.aaggt	ggcctttgtg	tggcactggc	5100
ttgggtggtc	gggaaaggca	ttttcagctt	gttgacagaac	tgccacagta	gcatgctggg	5160
tccgtgaaag	tttctgccc	ttacaagaa	gtctctacta	cttgtgacct	caccagtga	5220
aattttcttta	attgtctcct	ggtgttctgg	gttttgcatt	tttgtttcta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttggat	tgggaaagat	5340
gatttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttgggaa	aatgtcctcc	5400
cgccaccccg	aattttggct	tggcaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460
gaggaggcat	gggtccctag	gaggacagga	aycagaagga	gagagctygg	ctgacagcct	5520
gcaggcattg	cacagtttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggtctc	gaggcaactt	gacctcagcc	agatggattt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgatga	5700
aactctcact	ttgcctgagt	catgtgcggc	catgccacgg	ggtcaggctg	acactcaact	5760
caaaaacaag	tgagaaattg	aagacaatcc	gtggtggcag	ctactgggaag	ggccaccaca	5820

```

tccccagaaa gagtggagct gctaaaaaagc catttgatgat aggcacagtt atcttgaatg 5880
catggagcag agattacgga aaaatcgaga atgttaatga ggcaacattc gagttgagtc 5940
attcagtggt ggaaacccag acgcttccat cccctaaaag gaacatcttg ctctcagtc 6000
aaatggaaat aaaaattggg gcttgaattt ggcaaatgat tcagaactct gtgtaggtat 6060
tttcacacgc acagtggata attttcatgt tggagtttat ttgtgctaaa aggcagaaaa 6120
gggtaaaaag cacatcttaa gagttatgag gttctacgaa taaaaataat gttacttaca 6180
gctattcctt aattagtagc ccttccacc tgtggttaatt tcctgagata gtcagtggg 6240
aaaagatctc tccttctctt ctttctcccc cccccctcct cctcctccct cctcctccct 6300
ctcctcctc tcctcctcct cctcttctct tcttctcttg cctcttctct tctgctcct 6360
tctcctcttc ttcttcattt attctaagta gcttttaaca gcacaccaat tacctgtgta 6420
taacgggaaa acacaggctc aagcagctta gagaagattg atctgtgttc act 6473

```

<210> 2

<211> 7045

<212> DNA

<213> *Cricetulus griseus*

<400> 2

```

actagcgtgc aattcagagg tgggtgaaga taaaaggcaa acatttgagg ccatttctct 60
atttggcagc gcacttagga agtggaaacat gcctaatacta ctggtttgta ccaccttctc 120
ctataatgga ctgtttggga agctcctggg caaccgattc tggcatctca ttggtcagag 180
gcctgttaaa tggtagctct atttgcaaag aaggctgtaa cttgtagctt taaaagcctc 240
tcctcaagaa agaagggaga aaggatatgg ctagacatat ctaatagact taaccactgt 300
gaaaagcctt agtatgaatc agatagaacc tatttttaac tcagttttga aaaaaataat 360
ctttatattt atttgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt 420
gaaccacatg tagcaggtgc tggaggaggc cagaagaggg caccagatct cctggaactg 480
acaccacaca tggttatgag ctgcctgatg tgggtgctgg gaactgaact ctctgttct 540
gcaagagcag caactgttct cttaactgat gagccatctc tcagcccccc ccataattt 600
taattgttca ttttagtaaa ttttattcat aatcaattat cacagtataa aacaatgatt 660
ttatatatat catatacata tcaaggatga cagtgaaggg gatatgtgtg tgtgtgtgtg 720
tgtgtgtgtg tgtgtgtgtg tgtgttattt gtgtgtgtgc ttttaagaa ggtgccatag 780
tcaactgcatt tctctgaagg atttcaaagg aatgagacat gtctgtctgc caggaaacct 840
atcttctctt ttgggaatct gacccaaatg aggtattctg aggaactgaa tgaagagctc 900
aagtacagat gtcttaaac caaatgtgct gtctagagaa agtcaacgtc atcagtgagc 960
tgaggagaga tttactgagc ggaagacaag cgctcttga tttagtggtc tcgaacagtc 1020
acggctgtgg agtggagcct gtgctcaggt ctgaggcagt ctttgctagc cagctgtgat 1080
gagcagtgaa gaaaggggtg agatggaggc aggtgggag cagggtctatg gttcagacta 1140
ggtatcgtga gcacaccagc tgggtgactt gtggtctgtg ggtcaggcgt tgtaaacgcc 1200
ctcagggtca ggcagtcaca ttgcttgaag ctgaatgggt gaggcaacac agagagtga 1260
aagaaggcaa agtaccacct ctccccgac ccaggctcact tctgggttat agctgagact 1320
ccggacagca tgcaaccagc tgggttagagc ttcagggaaa acttgatgct tgcattgtgc 1380
tatgaaatgt gattcggtag atctggagaa aatttataat gctggctcag tcaagcactg 1440
aacaaggtta ccttggcttt gggagctaca tgacattgac ttgtagtcag actttttttt 1500
ttctgccgcg caattcccag ataaccaata tggaggtcca atattaatta taaatgctcg 1560
gctgatagct caggcttgtt actagctaac tcttccaact taaatgaacc catttctatt 1620
atctacattc tgccacgtga ctttaacctg tacttctctt tctctctct tgtctgactc 1680
tgcccttctg ctccccagag tctttagtct ggttctctct cctaacctta tctgcccag 1740
ctgctgacca agcatttata attaatatta agtctcccag tgagactctc atccaggag 1800
gacttgggtg cctccccctc ctcatggcca tctgtgtct cctcttctct cgttcccc 1860
tctcttctc gctcttctc ctccaccct ccttccatag tattgatggc aagggtgttc 1920
tagaatggag gagtgcccat aggcattgaa agaaaccagt taggatgctc tgtgaggggt 1980
tgtaatcata agcgatggac acaattcaag ccacagagtg aagacggaag gatgcactgt 2040
gctctagagc aacttctggg gcagaatcac aggtgagtt tctgacttga gggcgaagag 2100
gccacgagga agggagttag tttgtctgag ctagaagcta cggcccactc cttggtagca 2160
gacctgccc caagcatgct ttgttaatca tgtgggatct gatttctctc taaatctatg 2220
ttcaactctt aagaaaatgt gaattctcac attaaaattt agatatacgt cttttgggtg 2280
ggggggtgta aaaaatctct aagaatatgg atttctggg gccggagaga tggctcagag 2340
gttaagagaa ctgggtgtct ttctagacat tctgagttca attcccagca accacatggt 2400
ggctcacaac catctgtaat gcgacctggt gccatctctc gacatgcatt gatcacatgca 2460
ggcagaaagc tgtatacata gtaaatgat aaatctttt taaaaagag tatggattct 2520
gccgggtgtt ggtggcgcac gcccttaatc ccagactct ggaggcagag gcaggtggat 2580

```

```

ctctgtgagt tcgagaccag cctgggtctat aagagctagt tccaggacag cctccaaagc 2640
cacagagaaa cctgtgtctcg aaaaaccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2700
gtatggattc taagaaagcc gtaacagctg gagclgtgta cggagtlcag cgtgglaact 2760
gaagaacaga cattcatgat gaaacacccc aggatTTTTa cttagtatct agtttccatt 2820
gttgttttga gaccggtctt tatgtctctc aggetggcct caaactgctg atcttcccg 2880
ctctacctct caagtccctg gactacttgg ctcataaaac agtttttgtc gggtccctg 2940
aagttatggt tgtacaaacc gtgggggtca atatactcac ttgggcagag agagaaggtc 3000
tgaatcccag acaatgactg catctcagga cagttgggaa gaggacaatg gcagaaggac 3060
ttagaaaaga tagactggag ctgcagaata gacaggaac agagaaacaa aacaggaagc 3120
ttgtatcca gggccactct ggagtcctgt ggcaagatgg aagcgggcta ggggaatata 3180
tttqtgctac tglgtgtglg tgtgtgtglg tgtgtgtgtg tgtgtgtgtg caalgcctat 3240
caatgttgaa ggggaatat ctgcagaata gacagagtga atggtgccct ttatcattg 3300
cctagagaaa ggaatggccc ctgcagaata acatccttcc atagagtttc aggtaatga 3360
ctaaagttaa ggagaaataa acatccttcc atagagtttc aggtaatga accccacagt 3420
tcactgtgct cgtggtggag gcctggccaa cagttaaaaa gatttagacac ggacaaagtc 3480
tgaaggaaac acctcgaata ggaagaggag agccacctca ttctgtaact ttctcagg 3540
gaaagtggtt ccaagagtggt gaataaatgg tcaaaagggg gatttttaac taggaaaacg 3600
atctcctgtt tcacttgcga aactggaggt tgatttgggg cataggacaa tagattttag 3660
gctttgcaaa aagctgtttc aaagcagaga aatggaatag agacaattal gtacgagga 3720
gggagggtag ggcgaagatg gagacagaga agtggaagct gactttaggg aagaggaaca 3780
tagaccacag gggcgggggc gggggcaggg gcggggggcg gggctcaaag gaggcagtgg 3840
gaacgttgct agtgttcgca gcgtaaagct gaatgtgcaa gcgtctttgt gggtgtgtgac 3900
caggagttag gtggtggtc ttgtgtgtgc ttgtaatccc agtcttttag gtttccacac 3960
tgttccacag tgggtgtgat ttccctcgg agagcatgag ggctctgctt tccccacatc 4020
ctccccagcg ttctgttgta ttgttttcca agatgttagt ggggtgagaca aagcctctct 4080
gttgatttgc ctttaacagg tgacaaaaaa agctcaacca ggagacattt ttgctttctt 4140
ggaaggtaat gctcccatgt agagcaatgg gacctctc taaggtaggg ctactcttgc 4200
aqlttgcacc cagctcttcl tgacacagaa ggaagttagt gggcaagcaa gactgtttgc 4260
ttcttgcat ggacacattc tgacacaaaa ggctcaggag gggagaaggg tgtttgatgt 4320
ttagcactca ggaaggcccc tgatgcatct gtgattagct gtctccatct gtggagcaga 4380
cacggactaa ctaaaaacca gtgtttttaa attgtcaagg ctttaagggt aggaaattga 4440
cttattgtgc tgggccatac gtagagcaag tgctctgcat tgggccaaac cccggctctg 4500
gtttctaggg accagaatgg cctagaacta actcacaatc ctcccatcc aggtctcagg 4560
tgctagaatg aaccactata ccagcctgcc tgcctgecta cctgccttcc taaattttaa 4620
atcatgggga gtaggggaga ctacacttat cttagttagg gtttctattg ctgtgaagag 4680
acaccatgag catggcaact ctataaagg aaacatttta gttgggtggc agtttcagag 4740
gttttagtac attgtcatca tggctgggaa catgatggca tgcagacaga catgggtgctg 4800
gagaaaggga tgagagtcct acatcttgca ggcaacagga cctcagctga gacactggct 4860
ggtacctga gcataggaaa cctcacagcc caccctcaca gtgacatatt tcttcaaca 4920
aagccatccc tctaatagt gccactccct atgagatgac agggccaatt acattcaaac 4980
tgctataaca ctttaagta ttttattttt attattgtaa attatgtatg tagctgggtg 5040
gtggcagccg aggtgcacgc ctttaatccc agcacttggg aggcagaggg agatggatct 5100
ctgtgagttc aagaccagcc tgggtctataa gagctagttag caaggaagga tatacaaaga 5160
acagttctag gatagccttc aaagccacag agaagtgctg tcttgaaaac caaaaattgt 5220
gctgggacct gtctctgctt tgggtgtctc ccaactcccc agagctggac tcttgggtcaa 5280
cactgaatca gctgcaaaat aaactcctgg allectctcl lgtaacagga gcccgaglc 5340
agcgccccc ttgtcttctc gcaggattgc catagacttt ttctgtgtgc ccaccattcc 5400
agactgaagt agagatggca gtggcagaga ctgggaaggg tgcaacgaaa acaggaagt 5460
attgcacctt gggaatagtc tggaaatgaa gcttcaaaac ttgcttcatg ttcagtgtga 5520
cacagactca ctcccaggtt gactcacacg tgtaaatatt cctgactatg tctgactgc 5580
ttttatctga tgcttcttcc ccaaatgccc aagtgtacaa ggtgagggaa tcaccttgg 5640
attcagagcc cagggtcgtc ctcttaacc ttgacttgtc tttctcgggc agcctctgac 5700
acctctccc ccattttctc tctagaagg tctgagcaga gttggggcac gctcatgtcc 5760
tgatacactc cttgtcttcc tgaagatcta acttctgacc cageaagatg gctaagggtg 5820
tgaagtgttt gacatgaaga cttggtctta agaactggag caggggaaaa aagtcggatg 5880
lggcagcaug taccgaaat cccagaactg gggaggtaga gacggatgag tgcccggggc 5940
tagctggctg ctcagccagc ctagctgaat tgccaaattc caactcctat tgaaaaacct 6000
ttaccaaaca aacaacaaaa caaataataa caacaacaac aacaacaaac taccatac 6060
aaggtggggc gctcttggct cttgaggaat gactcaccce aaccacaaagc ttgccacagc 6120
tgttctctgg cctaataagg gtgggggtgg gcagagaca gagacagaga gagacatgac 6180
ttcctgggct gggctgtgtg ctctaggcca ccaggaactt tctgtcttg ctctctgtct 6240

```

```

ggcacagcca gagcaccagc acccagcagg tgcacacacc tccctccgtg cttcttgagc 6300
aaacacaggt gccttggtct gtctattgaa ccggagtaag ttcttgaga tgtatgcattg 6360
gaaacaacat tgtcctgggt ttatttctac tgttggtgata aaaaccgggg aactccagga 6420
agcagctgag gcagaggcaa atgcaaggaa tctgcctcc tagcttgctc cccatggctt 6480
gccgggctctg ctttctgcaa gcccttctct ccccatgggc atgcctgaca tgaacagcgt 6540
ttgaaatgct ctcaaagtgc actttcaaag aaggcttctc tgatcttgct aactaaatca 6600
gaccatgttt caccgtgcat tatctttctg ctgtctgtct gtctgtctgt ctgtctatct 6660
gtctatcatc tatcaatcat ctatctatct atcttctatt tatctaccta tcattcaatc 6720
atctatcttc taactagtta tcatttatit atttggttac ttactttttt tatttgagac 6780
agtatttctc tgagtgcagc ccttggtgtg cctggaaccc attctgtaac caggctgtcc 6840
tcaaactcac agagatccaa ctgctctgct ctctctggtg ctgggggttaa agacgtgcac 6900
caccaacgcc ccgctctatc atctatttat gtacttatta ttcagtcatt atctatctc 6960
taactatcca tcatctgtct atccatcatc tatctatcta tctatctatc tatctatcta 7020
tctatcatcc atctataatc aattg 7045

```

<210> 3

< 211> 6473

< 212> DNA

< 213> *Cricetulus griseus*

<400> 3

```

agtgaacaca gatcaatctt ctctaagctg cttgagcctg tgttttcccg ttatacacag 60
gtaattggtg tgctgtttaa agctacttag aataaatgaa gaagaaaggg agaaggaggc 120
agaggagaag gagcaaagaa agaaggaaag ggggaggagg ggagaggagg gagggaggga 180
gggaggaggg gagaggaggg gagggggaga aagaagagaa ggagagatct ttccccact 240
gactatctca ggaattacc acaggtggaa gggggtacta attaagggaat agctgtaagt 300
aacattatit ttattcgtag aacctcataa ctcttaagat gtgcttttta ccttttctg 360
ccttttagca caaataaact ccaacatgaa aattatccac tgtgcgtgtg aaaataccta 420
cacagagttc tgaatcattt gccaaattca agccccaatt tttatttoca ttttgactga 480
gagcaagatg ttcttttag gggatggaag cgtctgggtt tcccacactg aatgactcaa 540
ctcgaatgtt gcctcattaa cattctcgat ttttccgtaa tctctgctcc atgcattcaa 600
gataactgtg cctatcacia atggcttttt agcagctcca ctctttctgg ggatgtgggtg 660
gcccttccag tagctgccac caaggattgt ctccaattt tcaettgttt ttgagttgag 720
tgtcagcctg acccctgggc atggccgcac atgactcagg caaagtgaga gtttcatcac 780
taaacgtggc tctgtttgct atgtctgttt tccctctaag agcaggttat tcaaatacca 840
tctggctgag gtcaagttgc ctgagagccc acagaatctc taccaggtc cctgttggtat 900
cctaaaaaac tcagtcatgc tgtaatctcc ttctgaaact gtgcaatgcc tgcaggtctg 960
cagcccaagt ctctctctct gcttctgtc ctctaggac cccatgcctc ctcaaacgtc 1020
cacgtgtttc ttgctctct accacgggtg ccaagccaaa attcgggtgg gcgggaggac 1080
attttcccaa gtgctctgtt ccttctttt ccttttgaca ccccagataa atcatctttc 1140
ccaatccaac acagcccccac tgtgtctttt gggacttcat gacatcaccc aggaatgtat 1200
ccttagaaac aaaaatgcaa aaccagaaac accagagagac aattaaagaa attttcactg 1260
gtgaggtcac aagtagtaga gacttcttgt taacgggcag aaactttcac ggaccacaga 1320
tgctactgtg gcagttctgc aacaagctga aaatgcctt cccgaccacc caagccagtg 1380
ccacacaaag gccaccttag ggtgtgcaca ggatgtcact aggcgttggc ggaactcagg 1440
aaggagctctg aatttctctc cgtttctctc ttctctctc attccctatc ttagcttctg 1500
tctctctttc ctctctctct tccccccct tctcctctc ctctcgttg caggggccaca 1560
gatggaccgg gagacctcaa gcatgtcaaa tcaactaact gctctaccac tcaaccacac 1620
cctgcctgct attgttacta ctactattat tatcttgata caggtctcca cattgagctc 1680
accctcacag tctccacatt gagctcaccc tcacagctc cacattgagc tcaccctcac 1740
agtctccaca ttgagctcac cctcacagtc tccacattga gctcaccctc acagtctcca 1800
cattgagctc accctgtggc tctggcaaac cttgaattct ctcatctc ctgcctcagc 1860
ctctgggggtc gtgggggatta gccaaaccca cttgagggtt tcttcaatca gcaaattctt 1920
agcgttcaat taacacacac tcataactcc agtactttgg aaaccgggaa aggagaattt 1980
ctgtgagctg gaggttagct tggactacag tatgagaccc tgtctctaaa taaatacaca 2040
aagaaatctc accaagggtc tccctctctc agcaagctct aactgtgggt ggagttctgg 2100
gttggtccag ttaacgggtc cagaactcta ctgccagca catcagcccc tagacacagg 2160
tggtctctca catgtgaaca tgcagtcaca gaaatgaaat aaagtgaata ttttatttct 2220
tcagttgtat agcctctctc gtgtgggtct tagttactgt cttgaatagg ataggctcag 2280
aatccttggg gctggaacca agagtttgat tccattagac gacagggaat ataatgccc 2340
atagggcatt cctcctccg gtcactagcg gtgcacttcc tccgaactct tgtcatgttg 2400

```

```

aattagaaaa gttagtattt tctccatcc ctccccctcc tccccctctc cctctctccc 2460
ctccccctcc cctccccctc tctccccctc cctccccctc cctctctgat cccccccatc 2520
tatcaaatcc aagaattcca gtaaaaaagag gaaaaacaatc gaagtgattt cgttgattgt 2580
caqgttccacc aaagcaagac ttgactltag lccgcggttt cggttcccg cagtcaccac 2640
agccagcgag caccgtggaa ggatgctagc acggctctcc ccccgcccc actagctgtc 2700
ttcagctccc cagttagagg caaccgcact ccagattctc aatggagagt gttacacaa 2760
tcgttgccggg ttgtgtgag cgcgcgcgt tccagagaca cttcttcttt tctttttttc 2820
catttcaccc cagtggcaac gcagagtgcc agatcattca ggccgtttgc agggcaagcc 2880
gtgggagcct ggcaagcaag gccccatttc ctagggaacc cgtgcctggc gcttcaggaa 2940
agcacgggaa cctggcactg tgactctgag ggtattattt tgcagaactc tttattaaac 3000
gggagtttca agtccagctg gagacgacca ggcagcgcc ttaaccccag agtcacacac 3060
aggtgccttt tcttggggcc agattggggt tgtgtggcag acctgcqacc agcttgacaa 3120
ctcttctgcc aggccacaaa atggtgttgg ctgtaagagg tgacaccagg gacagggaag 3180
alcgctgcta ttctctgag ctctccaaag acccacacca gtctgtcccc ctttctctct 3240
gctcttcccc tgtatcgccc cctcaccatc tcccccaacg agactcttgg catctcctcg 3300
gcacaaggat ttgaaaatag atgcttgggg gtgagaagaa gaagagagaa agagagagaa 3360
ggaaggaaag atatatagat gatacagacg catacagtg atacgtagt aatcattttt 3420
aattaaaaaa taaattaaaa gcaaatcaag gatatatatg atacccttag agcaagtgtc 3480
tcatacacac acaaacacac acacacaata tatatatata tatatatata tatatatata 3540
tatatatata ttatacttgg aacaagtgtc cagaagggct ggggactcta aagtgcctgt 3600
caaagccagg ctcacatcag taatcttctc acctggtaga ctgagacagg aggattttga 3660
tgagttcagg cccagcctga gctgcagaat gtgattctat cccaaaaaaq taaaataaaa 3720
taaaattcaa aalacacgaa aagagtattt gctgaacaaa caagcctaaa gccctggatc 3780
ccttccccca tgtcctaaga aaataagttt ctgaaagctg gagggatggc tcagaggtta 3840
agagccccag ctgcacttgc ggaacactaa gaccagttc ccagacccca cactgtgggt 3900
cacaaactgt tcaaacgcca atccatgcc tctcctggcc tccaccgcaa 3960
ccaagaacac atacagtgcc catacattta tgcaagcaag gtattcacgc acataaaaact 4020
aaaagaatat ttaataaaga tataacaaaa tagcatgaag ccagctgggt acagaggttc 4080
aaactacatc ccaggttcat cctctgcctc ttgctctcag ttggcttggg taggtctctt 4140
ctctgaactg gcgccttgcg ggttccacat tgagaccctc tcatttttaa acctacttct 4200
tctgggcggg gtttaattgct gccagggtc aagccaaacg tctcctctct ccacagcaat 4260
cttccaagtt tcacgagata accaggaact gctaagttca tgtgaacctt agtgaagaac 4320
ctgagctctc ccattgtcatt qgtgtgtgca tglgtgcata cacaatgta tgtgtgtgct 4380
clalglgtgc ctatgtatgt gtgcatgcat gtgtgcatat acaaatgcat atatgtctat 4440
gtagtgtgcg tacacaaatg tatgtgtgtg ctcaatgtgt gctatgtgt gtgtatgcat 4500
gtgtgcgtac acaatgcatt gtgtgtgtgt ctgtgtgct gtgtgtgtat gcatgtatgc 4560
atacacaaat gtatatgtgt ggtgtgtgaa tgtgtgccta tgtatgtgtg tgcgtgtgtg 4620
gggtgtggtg tgtgtgtgat gtgtggaggg gtgtgtatgt gtggtatgta taggtgatac 4680
tattgggggt taatatgcgt atgtggtttg tgaaatgtag ttcgtgtgtg tgcatgtgtg 4740
cgtgcgtgcg tgcgtgcgtg cgtgtgtgtg tgtgtgtgtg tgtgtgtgtt ggatatagta 4800
tgtgtgaggt gtgtgtactc acctggcct cctcacttg ggggagtga gtcagcagcc 4860
tggaccactc agggacatga gatactcaga cacatcttga tttccacccc tcttttctctg 4920
atcctccttc acgtgtcact tccccaaaca ctggacaaca gtttgggggc atctgattcc 4980
actaatgaca ggacacatcac atgtctccag agggaacacc ttctgtgtca catgtcatc 5040
gagaatgtag cagagtcaca gagaaatgtc acagaaacca aaatgcagag taccaaggta 5100
tagctaggca cagagcagag gggaaagccc tgaatttatt aaaaatgtca gaatcgtaaa 5160
agacagggga cagcgggtggg gacattcagg gtccagtagc acacaggcag tccaaacctg 5220
atcactggaa ggtagttagt aaggaaaggg tgcaacacaga ttattcacac agtttatata 5280
tgtacacaga ttattcacat ggtttgtgta tgtgcacaga ttattcacac agtttatata 5340
tgtgtggctt cgtggttaact ttgagcttac ttccaattta aaaggatctc tctcacaaac 5400
tggggccggg aatggctgca gtcaacactc catcacttag tcacactgtg caaacagcac 5460
ctcctgactc atggtgactt gtagtataat gaagaggcca catttgcatc caagacagct 5520
catcagtacc tagtgaagaa tctgtccctg agtatttgca tgaatggacc cgggtccagg 5580
gcctggctgg gagtctccag gtgttgacgc cagaatgtca ttgtgtttt tcaggatccc 5640
agaagtlcll aaaaacacag ccaagtactc atttgtgtta caaagtatct gactaataga 5700
agtgattagg taacacaaag ccttttaaaa accgagatca ccttgtcat gtccctggcc 5760
tcttagaaca agatccaagc ttttgcctgg tgacaagtgg ggccatccag tgcgtctccg 5820
ttcctgctac ttcatctgga agacctctcc cactaacttg cccctgaccc ctcacacctg 5880
ctgtttctct tccacccgga agtgcttgtc taggctttca tggccatctg actgagcatc 5940
taggcctcag tccagtggtc cctcagctct ctctagtca tgtactaatg gaaacggcca 6000
ctaactacat tttaatatg gaagcctcct cctcaggaa cccaagggc agaagcctcc 6060

agagaaccac tcttgacccc ctggagttct gagtgtctct ggccctctct gtgtctgcag 6120
gactattcac cacttgtgtt gaatggttca gtctcacct cctctggcat gtgtcagtt 6180
ctcatctcat tggggagtc ttcccaggtc actctctct cctgtctttg aagtgtttt 6240
ttccttcatg gtatttctgt ctgggcacac acacagacac acatacacac acatacacac 6300
ccatgcagta tggcagatac atcacctatg ttacagattt ttattctacc atcacccaat 6360
acctgaatcc ccgaaaaagc cttagaaaagc caggaaattg tgtatttttgc tcagcactcc 6420
acccagcac ctgaagccaa gcctgactta atatttttgg tttgtttct aga 6473

```

<210> 4

< 211> 7045

< 212> DNA

< 213> *Cricetulus griseus*

<400> 4

```

caattgatta tagatggatg atagatagat agatagatag atagatagat agatagatga 60
tggatagaca gatgatggat agttagagga tagataatga ctgaataata agtacataaa 120
tagatgatag agcggggcgt tgggtgggca cgtctttaac cccagcacca gagaggcaga 180
ggcagttgga tctctgtgag tttgaggaca gcctgggttac agaatgggtt ccaggacagc 240
caaggctgtc actcagagaa atactgtctc aaataaaaaa agtaagtaa caaataaata 300
aatgataact agttagaaga tagatgattg aatgataagt agataaatag aagatagata 360
gatagatgat tgatagatga tagacagata gacagacaga cagacagaca gacagcagaa 420
agataatgca cggtgaaaca tgggtctgatt tagttagcaa gatcagagaa gccttctttg 480
aaagtgcacat ttgagagcat ttcaaagcgt gtccatgtca ggcatgccaa tggggagaga 540
agggtctgca gaaagcaggc ccggcaagcc atggggagca agctaggagg cagcattcct 600
tgcatttgcc tctgcctcag ctgcttctcg gagttccccg gtttttatca caacagtaga 660
aataaaacca ggacaatggt gtttccatgc atacatctgc aagaacttac tccggttcaa 720
tagacagacc aaggcacctg tgtttgctca agaagcacgg agggaggtgt gtgcacctgc 780
tgggtgctgg tgctctggct gtgccagaca gagagcaaga caggaaagtt cctggtggcc 840
tagagcacac agcccagccc aggaagtcat gtctctctct gtctctgtct ctgcccccacc 900
cccaccccat ttaggccaga gaacagctgt ggcaagcttt ggggttggtt gagtcattcc 960
tcaagagcca agagccgccc acctgtgatg gggtagtttg ttgttgttgt tgttgttatt 1020
atgtgtttgt ttgtttgttt ggtaaaggtt tttcaatagg agttggaatt tggcaattca 1080
gctaggctgg ctgagcagcc agctagcccc gggcactcat ccgtctctac ctccccagtt 1140
ctgggatttc gggtagatgc tgccacatcc gacttttttc ccctgctcca gttcttaaga 1200
ccaagtcttc atgtcaaaca cttcaccacc ttagccatct ttctgggtca gaagtttagat 1260
cttcaggaag acaaggagtgt tatcaggaca tgagcgtgcc ccaactctgc tcagaccttc 1320
tgatagagaa aatgggggga ggggtgtcag aggctgccgg agaaagacaa gtccagggtta 1380
aggaggacga ccctgggctc tgaatecaag ggtgattccc tcaccttgta cacttggcat 1440
tttgggaagg aagcatcaga taaaagcagt gcagacatag tcaggaatat ttacacgtgt 1500
gagtcaacct gggagttagt ctgtgtacaa ctgaacatga agcaagtttt gaagcttcat 1560
ttccagacta ttcccagggt gcaataactt cctgttttcg ttgcagcctt cccagtctct 1620
gccactgcca tctctacttc agtctggaat ggtgggcaca cagaaaaagt ctatggcaat 1680
cctgcgagaa gacaagtggg cgctgactt cgggctcctg ttacaagaga ggaatccagg 1740
agtttatttt gcagctgatt cagtgttgac caagagtcca gctctggggg agtgggaagc 1800
aaccaaagca gagacaggtc ccagcacaat ttttggtttt caagacagca cttctctgtg 1860
gctttgaagg ctatcctaga actgttcttt gtatatcctt ccttgcaact agctcttata 1920
gaccaggctg gtcttgaact cacagagatc catctgcctc tgcctcccaa gtgctgggat 1980
taaaggcgtg cacctcggtc gccaccaccc agctacatac ataatttaca ataataaaaa 2040
taaaatactt taaagtgtta tagcagtttg aatgtaattg gccctgtcat ctcataggga 2100
gtggcactat taggaggtat ggctttgttg aaggaaatat gtcactgtga ggggtggctg 2160
tgaggtttcc tatgctcagg gtaccagcca gtgtctcagc tgaggctcctg ttgctgcaa 2220
gatgtaggac tctcatccct ttctccagca ccatgtctgt ctgcatgcca tcatgttccc 2280
agccatgatg acaatgtact aaaacctctg aaactgccac ccaactaaat gttttccttt 2340
ataagagttg ccatgtctat ggtgtctctt cacagcaata gaaaccctaa ctaagataag 2400
tgtattctcc cctactcccc atgattttaa atttaggaag gcaggtaggc aggcaggcag 2460
gctggtatag tgggttcattc tagcacctga gacctggaat gggaggattg tgagttagtt 2520
ctaggccatt ctgggtgcct gaaaccagag ccggggggtt gcccaatgca gagcacttgc 2580
tctacgtatg gccagacaca ataagtcaat ttctcacct taaaggcttg acaattttaa 2640
aacactggtt tttagttagt ccgtgtctgc tccacagatg gagacagcta atcacagatg 2700
catcaggggc cttcctgagf gctaaacatc aaacagcctt ctccccctct gagcctttgt 2760
gtgcagaatg tgtccatcgc aagaagcaaa cagtcttgct tgcccaccaa cttccttctc 2820

```

```

gcatcagaag agctgggtgc aaactgcaag agtagcctca ccttagagat gggtccatt 2880
gctctacatg ggagcattac cttccaagaa ggcaaaaatg tctcctgggt gagctttttt 2940
tgccacctgt taaaggcaaa tcaacagaga ggctttgtct caccactaa catcttgaa 3000
acaaalacca acgaacgctg gggagatgt ggggaaagca gagcctcat gctctccgag 3060
ggaaaatcac accactgtg gaacagtgt gaaacctcaa agactgggat tacaagcagc 3120
acacaagcca gccacgctac tectggtcac acaccacaaa gacgcttgca catcacgct 3180
tacgtgcca acactagcaa cgttccact gctcctttg agccccgcc cccgccctg 3240
ccccccgcc cgcctctgt gtctatgttc ctctcccta aagtcagctt ccacttctct 3300
gtctcatct tgcgccacc cctccctctc gctacataat tgtctctatt ccatttctct 3360
gctttgaaac agctttttgc aaagcatcaa atctattgtc ctatgcccc aatcaacctc 3420
cagtttcaca agtgatacag gaaatcgttt tcctaattaa aaatcccccc tttgaccatt 3480
tattccact cttggaacat cttcccttg aggaaagtta cagaatqagg tggctctcct 3540
cttctattc gaggtgtttc cttcagactt tgtcgtgtc taatcttttt aactgttggc 3600
caggcctcca ccacggcaca gatgaactgt ggggttcatt tacctgaaac tctatggag 3660
gatgtttatt tctccttcac tttagcaaat gataaagggc accattcact ctgtctattc 3720
tgcaggggcc attcctttct ctaggccaga tactgagaat tgctccaga atcaatgtgg 3780
tatacatatt tccccctcaa cattgatagg cattgatcac acacacacac acacacacac 3840
acacacacac acacagtagc acaaatgtat tccctagcc cgcttccatc ttgccacagg 3900
actccagagt ggccctggat agcaagcttc ctgttttgt tctctgttcc tgctgctttt 3960
ccaccctcca gtctatcttt totaagtcct tctgccattg tctcttccc aactgtcctg 4020
agatgcagtc attgtctggg attcagacct tctctctctg cccaagttag tatattgacc 4080
ccacgggtt gtacaacct aacttcaggg agcccgacaa aaactgtttt alqagccaag 4140
tagtccagg acttgagagg tagaggcggg aagatcagca gtttgaggcc agcctggaga 4200
gcataagagc cggctcctaaa acaacaatgg aaactagata ctaagtaaaa atcctgggg 4260
gtttcatcat gaatgtctgt tcttctagta ccacgctgaa ctccgtacac agctccagct 4320
gttcaggctt tcttagaatc cataactctt tttttttttt tttttttttt ttttttttgg 4380
tttttcgaga cagggtttct ctgtggcttt ggaggtgtc ctggaactag ctctataga 4440
ccaggctggt ctgaaactca cagagatcca cctgcctctg cctccagagt gctgggatta 4500
aaggcgtgct ccaccaacac ccggcagaat ccataactct tttaaaaaaa gatttatcaa 4560
ttactatgt atacagcttt ctgcctgcac gtatccatgc atgtcagaag atggcaccag 4620
gtcgcattac agatgggtgt gagccacct gtggttgctg ggaattgaac tcagaatgtc 4680
tagaagagca accagttctc ttaacctctg agccatctct ccggccccc aaaaatccata 4740
ttcttgagga ttttttacac cccccccacc aaaagacgta tatctaaatt ttaatgtgag 4800
aattcacatt ttcttaagag ttgaacatag atttagagga aaatcagatc ccacatgatt 4860
aacaaagcat gcttgtgggc aggtctgcta ccaagaggtg ggccgtagct tctagctcag 4920
acaaactcac tcccttctct gtggctctt cgcctcaag tcagaaactc accctgtgat 4980
tctgccccag aagtgtctct agagcacagt gcatccttcc gtcttactc tgtggcttga 5040
attgtgtcca tcgcttatga ttacaacccc tcacagagca tcttaactgg tttctttgca 5100
tgctatggg cactcctcca ttctagaaca ccttgccat caatactatg aaaggagggg 5160
tgaggaggga agagcaggaa gagcgaggga agcgaggga acggatggca 5220
atgaggaggg gggagcacc aagtcctccc tggatgagag tctcactggg agacttaata 5280
ttaattataa atgcttggtc agcagctggg caggataagg ttaggcagga gaaccagact 5340
aaggactctg ggaagcagaa gggcagagtc agacaaggag aggaacagg aagtacaagg 5400
taaagtcacy tggcagaatg tagataalag aaatgggttc attlaagttg gaagagttag 5460
ctagtaacaa gcttgagcta tcagccgagc atttataatt aatattgagc ctccatattg 5520
gttatctggg aatttgccgg cagaaaaaaa aaagtctgcc tacaagtcaa tgtcatgtag 5580
ctcccaaagc caaggtacct ttgttcagt cttgactgag ccagcattat aaattttctc 5640
cagatgtacc gaatcacatt tcatagcaac atgcagacat caagttttcc ctgaagctct 5700
aaccagctgg ttgcatgctg tccggagtct cagctataac ccagaagtga cctgggtcgg 5760
ggaagaggtg gtactttgcc ttctttgcac tctctgtgt gcctcaccac ttcagcttca 5820
agcaatgtga ctgctgacc ctgagggcgt ttacaacgcc tgaccacag accacaagtc 5880
aaccagctgg tgtgtcacg atacctagtc tgaacatag cctgtctcc accctgctc 5940
catctccacc ctttcttca tgetcatcac agctggctag caaagactgc ctacagacctg 6000
agcacaggct ccactccaca gccgtgactg ttcgagccac ttaaatacaa gagcgcttgt 6060
cttccgctca gtaaatctct cctcagctca ctgatgactg tgactttctc tagacagcac 6120
atttgggttt aagacactgc tacttgagct ctctcactc ttctcagaa tacctcattt 6180
gggtcagatt cccaaagagg aagatagggt tectggcaga cagacatgtc tcttctttt 6240
gaaatccttc agagaaatgc agtgactatg gcaacctctt aaaaagcaca cacacaaata 6300
acacacacac acacacacac acacacacac acacacacac atatccccct cactgtcatc 6360
cttgatatgt atatgatata tataaaatca ttgttttata ctgtgataat tgattatgaa 6420
taaaatttac taaaatgaac aattaaaatt atgggggggg ctggagagat ggctcatcag 6480

```

ttaagagaac	agttgctgct	cttgcagaac	acgagagttc	agttcccagc	acccacatca	6540
ggcagctcat	aaccatgtgt	ggtgtcagtt	ccaggagatc	tggtgcoctc	ttctggcoctc	6600
ctccagcacc	tgtacatgt	ggttcacaca	cacacacaca	cacacacaca	cacacacaca	6660
cacacacaca	caaataaata	taaagattat	ttttttcaaa	actgagttaa	aaatagggttc	6720
tatctgattc	atactaaggc	ttttcacagt	ggttaagtct	attagatatg	tctagccata	6780
tcctttctcc	cttctttctt	gaggagaggc	ttttaaagct	acaagttaca	gccttctttg	6840
caaataagag	taccatttaa	caggcctctg	accaatgaga	tgccagaatc	ggttgcccag	6900
gagcttccca	aacagtccat	tatagggaaa	ggtggtacaa	accagtagat	taggcatggt	6960
ccacttccca	agtgcctgct	caaataagga	aatggcctca	aatgtttgcc	ttttatcttc	7020
acccacctct	gaattgcacg	ctagt				7045

<210> 5

< 211> 13515

< 212> DNA

< 213> *Cricetulus griseus*

<400> 5

tctagaaaca	aaacaaaaaa	tattaagtca	ggcttggttt	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggtctttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaaatctg	aaacataggt	gatgtatctg	ccatactgca	tgggtgtgta	180
tgtgtgtgta	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaaca	240
cttcaaagac	aggagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtgag	gactgaacca	ttcaacacaa	gtggtgaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggggtca	ggagtgggtc	tctggaggct	420
tctgcccttg	gaggttctctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgacta	gagagagctg	agggaccact	ggactgaggc	ctagatgctc	540
agtcagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcaggtgt	600
gaggggtcag	gggcaagtta	gtgggagagg	tcttccagat	gaagttagcag	gaacggagac	660
gcactggatg	gccccacttg	tcaaccagca	aaagcttgga	tcttgttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggtttttaaa	aggctttgtg	ttacctaatc	acttctatta	780
gtcagatact	ttgtaacaca	aatgagtact	tggcctgtat	tttagaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggcctggac	900
ccgggtccat	tcatgcaaat	actcagggac	agattcttca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcattttac	tacaagtcac	catgagtcag	gaggtgctgt	1020
ttgcacagtg	tgactaagtg	atggagtgtt	gactgcagcc	attcccggcc	ccagcttgtg	1080
agagagatcc	ttttaaatg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacacaa	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagccttct	cttacctact	accttccagt	gatcaggttt	1260
ggactgcttg	tgtgctactg	gacctggaat	gtccccaccg	ctgtcccctg	tcttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	ccccctctgt	ctgtgcctag	ctataccttg	1380
gtactctgca	ttttgggttc	tgtgacattt	ctctgtgact	ctgctacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatgccccca	aactgttgct	cagtgtttgg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggttg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtggg	ccaggctgct	1620
gacttcactc	cccccaagtga	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacat	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatat	cacacataca	cacctctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacat	acataggcac	acattcacac	accacacata	tacatttgtg	tatgcataca	1920
tgcatacaca	cacaggcaca	cagacaccac	acacatgcat	tgtgtacgca	cacatgcata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatg	cacacataca	taggcacaca	tagagcacac	2100
acatacatat	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggttcttc	2160
actaagggttc	acatgaactt	agcagttcct	ggttatctcg	tgaaacttgg	aagattgctg	2220
tggagaagag	gaagcgttgg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggtttaaaaa	tgagagggtc	tcaatgtgga	acccgcaggg	cgcagttca	gagaagagac	2340
ctaccaagc	caactgagag	caaaggcaga	gggatgaacc	tgggatgtag	tttgaacctc	2400
tgtaccagct	gggcttcatg	ctattttgtt	atatctttat	taaataattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaatgt	atgggcactg	tatgtgttct	tgggtgccgg	2520
ggaggccagg	agagggcatg	gatcctccgg	agctggcgtt	tgagacagtt	gtgaccacca	2580
gtgtggggtc	tgggaactgg	gtcttagtgt	tccgcaagtg	cagctggggc	tcttaacctc	2640

```

tgagccatcc ctccagcttc aagaaactta ttttcttagg acatggggga agggatccag 2700
ggcttttaggc ttgtttgttc agcaaatact cttttcgtgt attttgaatt ttattttatt 2760
ttactttttt gggatagaat cacattctgc agctcaggct gggcctgaac tcatcaaaa 2820
ctcctgtctc cagtctacca ggtgataaga ttactgatgt gagcctggct ttgacaagca 2880
cttttagagtc cccagccctt ctggacactt gttccaagta taalatalat atalatatat 2940
atatatatat atatatatat atatatattgtg tglgtgtgtt tgtgtgtgta tgagacactt 3000
gctctaeggg tatcatatat atccttgatt tgcttttaat ttatttttta attaaaaatg 3060
atlagctaca tgtcacctgt atgcgtctgt atcatctata tatccttcc tcttctctc 3120
tctttctctc ttcttctctc caccaccaag catctatttt caaatccttg tgcggaggag 3180
atgccaaagag tctcgttggg ggagatggtg agggggcgat acaggggaag agcaggagga 3240
aagggggaca gactggtgtg ggtctttgga gagctcagga gaatagcagc gatcttccct 3300
gtccctggtg tccctcttta cagccaacac cattttgtgg cctggcagaa gatttgtcaa 3360
gctggtcgca ggtctgccac acaaccccaa tctggcccca agaaaaggca cctgtgtgtg 3420
actctggggt taaaggcgct gcctggtcgt ctccagctgg acttgaact cccgtttaat 3480
aaagagttct gcaaaaataat acccgagag tcacagtgc aggttccctg gcttctctga 3540
agcgccaggc acgggttccc taggaaatgg ggccttgctt gccaaactcc caccgcttgc 3600
cctgcaaacg gcctgaatga tctggcactc tgcggtgcca ctgggatgaa atggaaaaa 3660
gaaaaagaag aagtgtctct ggaagcgggc gcgtcacac aaaccgcaa cgatttgtta 3720
aacactctcc attgagaatc tggagtgcgg ttgcccctca ctggggagct gaagacagct 3780
agtggggcg gggggaggac cgtgctagca tcttcccacg gtgctcgctg gctgtggtgc 3840
atgccggcaa ccgaaacgcg gaactaaagt caagtcttgc ttgggtgga ctgacaalca 3900
acgaaatcac ttcgtattgt tctctctttt tactggaatt cttggatttg atagatggg 3960
gaggatcaga gggggagggg aggggcgggg agacggaggg agggaggagg gagggaggga 4020
ggggaggagg gggaggaggg aagggatgga ggaataact aacttttcta attcaacatg 4080
acaaagattc ggagaaagt caccgctagt gaccgggagg aggaatgcc tattgggcat 4140
latattccct aatgctcaat ggaatcaaac tcttgggtcc agcacaagg attctgagcc 4200
tattctattc aagacagtaa ctacagccca caggaagag gctatcaaac tgaagaaata 4260
aaattttcac ttatttctat ttctgtgact gcattgtcac atgtagagag ccacctgtgt 4320
ctaggggctg atgtgctggg cagttagagt ctgagccctg taactggaac aaccagaac 4380
tcccaccaca gtttagagct gctgagagag ggaggccctt ggtgagattt ctttgtgtat 4440
ttatttagag acagggctc atactgtagt ccaagctagc ctccagctca cagaaattct 4500
cctgttcccg ttcccaagt actggagtta tgagtgtgtg ttaatgaa cctaagaatt 4560
tgctgattga agaaaacctc aagtggtttt ggctaattcc caccaccca gaggtgagg 4620
caggaggaat gagagaattc aaggtttgcc agagccacag ggtgagctca atgtggagac 4680
tgtgaggctg agctcaatgt ggagactgtg aggtgagct caatgtggag actgtgagg 4740
tgagctcaat gtggagactg tgagggtgag ctcaatgtgg agactgtgag ggtgagctca 4800
atgtggagac ctgtatcaag ataataatag tagtagtaac aatgcaggcg aggggtgtgt 4860
tgagtggtag agcgttagt gattttgaca tgcttgaggt ctcccgctcc atctgtggcc 4920
ctgcaacagg aagggggga ggaaggggg gaacgagaga gaggaaagag agacagaagc 4980
taagatagg aatgagagag acgggaagaa attcagactc cttcctgagt 5040
tccgccaacg cctagtgaac tctgtgcac accctaaggt ggcctttgtg tggcactggc 5100
ttgggtgctc gggaaggca ttttcagctt gttgcagaa gtctctacta cttgtgact caccagtga 5220
aatttcttta attgtctcct ggtgttctgg gttttgcat tttgtttcta aggatacat 5280
cctgggtgat gtcatgaagt ccccaagac acagtggggc tctgttgat tttgttgat tgggaaagat 5340
gattttatct ggtgtcact aggaaaagaa gggaaacagg caattgggaa aatgtcctcc 5400
cgcaccccg aattttggtc tggcaacctt ggtggaggag caatgaaac gtggacgttt 5460
gaggaggcat ggggtcctag gaggacagga agcagaagaa gaaagtggg ctgacagcct 5520
gcaggcattg cacagtttca gaaggagatt acagcatgac taagttttta gggatccaac 5580
agggacctgg gttagagattc tgtgggctct gaggcaactt gacctcagcc agatggtatt 5640
tgaataacct gctcttagag ggaaaacaga catagcaaac agagccactg ttagtgatga 5700
aactctcact ttgcctgagt catgtgcggc catgccaggy ggtcaggtg acactcaact 5760
caaaaacaag tgagaattg aagacaatcc gtggtggcag ctactggaag ggcaccaca 5820
tcccagaaa gagtggagct gctaaaaage catttgtgat aggcacagt atcttgaatg 5880
catggagcag agattacgga aaaatcgaga atgttaatga ggcacattc gaggtagtc 5940
attcagtggt ggaacccag acgcttccat cccctaaaag gaacatctt gctctcagtc 6000
aaatggaaat aaaaattgg gcttgaattt ggcaaatgat tcagactct gtgtaggtat 6060
tttcacacgc acagtggata attttcatgt tggagtttat ttgtgctaaa aggcagaaaa 6120
gggtaaaaag cacatcttaa gagttatgag gttctacgaa taaaaataat gttacttaca 6180
gctattcctt aattagtagc ccttccacc tgtggttaatt tctgagata gtcagtgggg 6240
aaaagatctc tcttctctt ctttctcccc ctccctcct cctcctcct ccctcctcc 6300

```

```

ctccctcctc  tccctccctc  cccctttcct  tctttctttg  ctccctctcc  tctgctcctc  6360
tctccctttc  ttcttcattt  attctaagta  gcttttaaca  gcacaccaat  taccctgtga  6420
taacgggaaa  acacaggttc  aagcagctta  gagaagattg  atctgtgttc  actagcgtgc  6480
aattcagagg  tgggtgaaga  laaaaggcaa  acatttgagg  ccatttcctt  atttggcacg  6540
gcacttagga  agtggaaacat  gcctaatact  ctgggttgta  ccacctttcc  ctataatgga  6600
ctgtttggga  agctcctggg  caaccgatcc  tggcatctca  ttggctcagag  gcctgttaaa  6660
tggtactctt  atttgcaaag  aaggctgtaa  cttgtagctt  taaaagcctc  tccctcaagaa  6720
agaagggaga  aaggatatgg  ctagacatat  ctaatagact  taaccactgt  gaaaagcctt  6780
caactgttct  cttaaactgat  gatccatctc  tccagccccc  cccataattt  ctttatattt  6840
atltgtgtgt  gtgtgtgtgt  gtgtgtgtgt  gtgtgtgtgt  gtgtgtgtgt  gaaccacatg  6900
tagcaggtgc  tggaggaggc  cagaagaggg  caccagalct  cctggaactg  acaccacaca  6960
tggttatgag  ctgctgatg  tgggtgctgg  gaactgaact  ctctgtttct  gcaagagcag  7020
caactgttct  cttaaactgat  gatccatctc  tccagccccc  cccataattt  taattgttca  7080
ttttagtaaa  ttttattcat  aatcaattat  cacagtataa  aacaatgatt  ttataatata  7140
catatacata  tcaaggatga  cagtgaaggg  gatatgtgtg  tgtgtgtgtg  tgcgtgtgtg  7200
tgtgtgtgtg  tgtgttattt  gtgtgtgtgc  tttttaagaa  ggtgccatag  tcaactgcatt  7260
tctctgaagg  atttcaagg  aatgagacat  gtctgtctgc  caggaaccct  atcttctctc  7320
ttgggaatct  gacccaaatg  aggtattctg  aggaactgaa  tgaagagctc  aagtagcagt  7380
gtcttaaac  caaatgtgct  gtctaagaaa  agtcaacgtc  atcaglgagc  tgaggagaga  7440
tttactgagc  ggaagacaag  cgcctcttga  ttttaagttg  tccgaacagtc  acggctgtgg  7500
aglgagagct  gtgtcaggt  ctgaggcagt  ctttgcctag  cagctgtgat  gagcagtgaa  7560
gaaaggggtg  agatggaggc  aggggtggag  cagggtatg  gttcagacta  ggtatcgtga  7620
gcacaccagc  tgggtgactt  gtggctctgt  ggtcaggcgt  tgtaaaagcc  ctcagggtca  7680
ggcagtcaca  ttgcttgaag  ctgaatgggt  gaggcaacac  agagagtga  aagaaggcaa  7740
agtaccacct  cttcccccag  ccaggtcact  tctgggttat  agctgagact  ccggacagca  7800
tgcaaccagc  ttggttagagc  ttcaggaaa  acttgatgtc  tgcattgttc  tatgaaatgt  7860
gattcgggtac  atctggagaa  aatttataat  gctggctcag  tcaagcactg  aacaaaggta  7920
ccttggtctt  gggagctaca  tgacattgac  ttglaggcag  actttttttt  tctgccccgc  7980
caattcccg  ataaccaata  tggaggctca  atattaatta  taaatgctcg  gctgatagct  8040
caggcttgtt  actagctaac  tcttccaact  taaatgaacc  catttctatt  atctacattc  8100
tgccacgtga  ctttaccctg  tacttctgt  ttctctctct  tgtctgactc  tgccctctct  8160
cttcccagag  tcttagtct  ggttctctct  cctaacctta  tctgcccag  ctgctgacca  8220
agcatttata  attaatatta  agtctcccag  tgagactctc  atccaggag  gacttgggtg  8280
ctccccctc  ctcatgcca  tccgtgtctt  cctcttccct  cgttccccc  tctcttctct  8340
gctcttctc  ctccacccct  cctttcatag  tattgatggc  aagggtgttc  tagaatggag  8400
gagtgcctat  aggcagcaa  agaaaccagt  taggatgtct  tgtgaggggt  tgtaatcata  8460
agcgtatgac  acaattcaag  ccacagagt  aagacggaag  gatgcactgt  gctclagagc  8520
aacttctggg  gcagaatcac  aggggtgagt  tctgacttga  gggcgaagag  gccacgagga  8580
agggagtgag  tttgtctgag  ctagaagcta  cggcccacct  cttggtagca  gacctgccc  8640
caagcatgct  ttgttaatca  tgtgggact  gatcttctc  taaatctatg  ttcaactctt  8700
aagaaaatgt  gaattctcac  attaaaattt  agatatacgt  cttttgggtg  ggggggtgta  8760
aaaaatctc  aagaatatgg  atttctgggg  gccggagaga  tggctcagag  gttaaagaaa  8820
ctggttgctc  ttctagacat  tctgagttca  attcccagca  accacatggt  ggcacacaac  8880
catctgtaat  ggcacctgg  gccatcttct  gacatgcatg  gatacatgca  ggcagaaagc  8940
tgtatacata  gtaaatgat  aaatctttt  ttaaaaagag  tatggattct  gccgggtgtt  9000
ggtggcgac  gcctttaatc  ccagcactct  ggaggcagag  gcagglggat  ctctgtgagt  9060
tcgagaccag  cctggtctat  aagagclagt  tccaggacag  cctccaaagc  cacagagaaa  9120
cctgtctctg  aaaaaacaaa  aaaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  gtaaggattc  9180
taagaaagcc  gtaacagctg  gagctgtgta  cggagtctag  cgtggtacta  gaagaacaga  9240
cattcatgat  gaaacacccc  aggttttta  cttagtatct  agtttccatt  gttgttttga  9300
gaccggctct  tatgctctcc  aggtggcct  caaactgctg  atcttcccgc  ctctacctct  9360
caagtccctg  gactacttgg  ctataaaaac  agtttttgtc  gggctccctg  aagttatggt  9420
tgtacaaacc  gtgggggtca  ttgggcagag  agagaaggtc  tgaatcccag  tgaaatcccag  9480
acaatgactg  catctcagga  cagttgggaa  gaggacaatg  gcagaaggac  ttagaaaaga  9540
tagactggag  ggtgaaaaag  cagcaggaac  agagaaacaa  aacagggaagc  ttgctatcca  9600
gggccactct  ggaqtcctgt  ggcaagatgg  aagcgggcta  ggggaataca  tttgtgctac  9660
tgtgtgtgtg  lgtgtgtgtg  tgtgtgtgtg  tgtgtgtgat  caatgcctat  caatgtgaa  9720
ggggaaatat  gtataaccaca  ttgattctgg  gagcaattct  cagtatctgg  cctagagaaa  9780
ggaatggccc  ctgcagaata  gacagagtga  atggtgccct  ttatcatttg  ctaaagtga  9840
ggagaaataa  acatccttcc  atagagtttc  aggtaaatga  accccacagt  tcatctgtgc  9900
cgtggtggag  gcctggccaa  cagttaaaaa  gattagacac  ggacaaagtc  tgaaggaaac  9960

```

```

acctcgaata ggaagaggag agccacctca ttctgttaact ttctcaagg ggaagatggt 10020
ccaagagtgg gaataaatgg tcaaaggggg gatTTTTtaat taggaaaaag atttcctgta 10080
tcacttgatga aactggaggl lgaattgggg cataggacaa tagatttgat gctttgcaa 10140
aagctgtttc aaagcagaga aatggaatag agacaattat gtagecagga gggagggtgg 10200
ggcgaagatg gagacagaga agtggaagct gactttaggg aagaggaaca tagaccacag 10260
gggcggggcg gggggcaggg gcggggggcg gggctcaaag gaggcagtgg gaacgttgct 10320
agtgttcgca gcgtaaagcgt gaatgtgcaa gcgtctttgt ggtgtgtgac caggagtaac 10380
gtggctggct tgtgtgctgc ttgtaatccc agtctttgag gttccacac tgtccacag 10440
tgggtgtgat ttccctcgg agagcatgag ggctctgctt tccccacatc cteccacgg 10500
ttcgttggtt tttgtttcca agatgttagt ggggtgagaca aagcctctct gttgatttgc 10560
ctttaacagg tgacaaaaaa agctcaacca ggagacattt ttgccttctt ggaaggtaat 10620
gtcccatgt agagcaatgg gacccatctc taaggtgagg ctactcttgc agtttgacc 10680
cagctcttct gatgcaggaa ggaagttggt gggcaagcaa gactgtttgc ttcttgcat 10740
ggacacattc tgcacacaaa ggctcaggag gggagaaggc tgtttgatgt ttagcactca 10800
ggaaggcccc tgatgcatct gtgattagct gtctccatct gtggagcaga caggactaa 10860
ctaaaaacca gtgtttttaa attgtcaagc ctttaagggtg aggaaattga cttattgtgc 10920
tgggccatag gtagagcaag tgctctgcat tgggccaacc cccggctctg gtttctaggg 10980
accagaatgg ccagaaacta actcacaate cteccattcc aggtctcagg tgctagaatg 11040
aaccactata ccagctgccc tgctgctcct cctgccttcc taaattttta atcatgggga 11100
gtaggggaga atacacttat cttagttagg gtttctattg ctgtgaagag acaccatgag 11160
catggcaact cttataaagg aaaacattta gttgggtggc agtttcagag gtttttagtac 11220
attgtcatca ttggtgggaa catgatggca tgcagacaga catggtgctg gagaaaggga 11280
tgagagtcct tggctctataa gagctagtgt caaggaaagg tatacaaaaga acagttctag 11340
gcataggaaa cctcacagcc caccctcaca gtgacatatt tccctcaaca aagccatacc 11400
tctaataagt gccactccct atgagatgac agggccaatt acattcaaac tgctataaca 11460
cllLaagta ttttattttt agcacttggg aggcagaggg agatggatct ctgtgagttc 11580
aagaccagcc tggctctataa gagctagtgt caaggaaagg tatacaaaaga acagttctag 11640
gatagccttc aaagccacag agaagtgtgt tcttgaaaac caaaaattgt gctgggacct 11700
gtctctgctt tgggtgcttc cactccccc agagctggac tcttggtcaa cactgaatca 11760
gctgcaaaat aaactcctgg attcctctct tgtaacagga gcccgagtc aggcgccac 11820
ttgtcttctc gcaggattgc catagacttt ttctgtgtgc ccaccattcc agactgaagt 11880
agagatggca gtggcagaga ctgggaaggc tgcacgaaa acaggaagtt attgcacct 11940
gggaatagtc tggaaatgaa gcttcaaaac ttgcttcag ttcagttgta cacagactca 12000
ctcccaggtt gactcacacg tgtaaatat cctgactatg tctgcaactgc ttttatctga 12060
tgcttctctc ccaaaatgcc aagtgtacaa ggtgagggaa tcaaccttgg attcagagcc 12120
caggtctgtc cctcttaacc ttgaacttgt tttctcggc agcctctgac accctcccc 12180
ccattttctc taticagaagg tctgagcaga gttggggcac gctcatgtcc tgatacactc 12240
cttgtcttcc tgaagatcta acttctgacc cagaaagatg gctaagggtg tgaagtgtt 12300
gacatgaaga cttgtgttta agaactggag caggggaaaa aagtoggatg tggcagcatg 12360
taccgaaat cccagaactg gggaggtaga gacggatgag tgccccgggc tagctggctg 12420
ctcagccagc ctatgtgaat tgccaaattc caactcctat tgaaaaacct ttaccaaaca 12480
aacaacaaaa caaataataa caacaacaac aacaaacaac taccatacag aaggtgggg 12540
gctcttggtc cttgaggaat gactcaccca aacccaaagg ttgccacagc tgttctctgg 12600
cctaaatggg gtgggggtgg ggcagagaca gagacagaga gagacatgac ttctgggct 12660
gggctgtgtg ccttaggcca ccaggaactt tctgtctgt ctctctgtct ggcacagcca 12720
gagcaccagc ccccagcagg tgcacacacc tccctccgtg cttcttgagc aaacacaggt 12780
gccttggtct gtctattgaa cggagtaag ttcttgaga tgtatgcatg gaaacaacat 12840
tgtcctggtt ttatttctac tgttgtgata aaacccggg aactccagga agcagctgag 12900
gcagaggcaa atgcaaggaa tgctgctctc tagcttgtct cccatggctt gccgggctg 12960
ctttctgcaa gcccttctct cccattggc atgctgaca tgaacagct ttgaaatgct 13020
ctcaaatgtc actttcaaag aaggtctctc tgatcttgt ctgtctatct gctatcatc 13080
caccgtgcat tatctttctg ctgtctgtct gtctgtctgt ctgtctatct gctatcatc 13140
tatcaatcat ctatctatct atcttctatt tatctacct tcatccaatc atctatctc 13200
taactagtta tcatttattt atttgtttac ttactttttt tatttgagac agtatttctc 13260
tgagtgcagc ccttggtgtt cctggaaccc attctgtaac caggctgtcc tcaaaactcac 13320
agagatccaa ctgctctgct ctctctgggt ctggggttaa agacgtgcac caccacgccc 13380
cgtctctatc atctatttat gtacttatta ttcagtcatt atctatctc taactatcca 13440
tcactgtct atccatcatc tatctatcta tctatctatc tctatcatc 13500
atctataatc aattg 13515

```

<210> 6

<211> 14553

<212> DNA

<213> Mus musculus

<400> 6

```

cttgaagaac acatgttttc caagagggag caccatgtt ggaatgacaa tgtagttagt 60
gctcctctcc tgtaggtagt tgctcctttg ctataggtaa gtgctcctct cctataggtc 120
agtgtcctcc tcctataggt tagtgctcct ctctataggt ttagtgctcc tctcctacag 180
gttagtgctc ctctgctcta ggtagtccct gctctcctat agtacctaga gagctagggc 240
aaatgggcta ggccgaagt gcagagacaa acagctatgg aagactgggt aagcacttcc 300
aagctacgaa agagcagtg gttagtgca ggttggtgca gtagtaggg gagatcttcc 360
agttgaagaa acagaagaac tgagagccac tgggtatcat cctcctgcgc catgccttcc 420
tggtactgct catgctccca ccttgatgat aatggaatga acctctgaac ctgtaagcca 480
gccccaatga aatatgttt ttatgagagt tgcttggtc atgctgtctg ttcacagcag 540
taaaacccta aataaggcag aagttggtag cagtattgct gtgatagacc tgaccatgct 600
ttcctttgaa agaattgtga tttggtgact ttggatttgc aacacagtgg aatgctttaa 660
atggagatta atgggtcatc aattcctagt aggaatatgg aagactttgt tgcctggagt 720
atttgaactg tgttgacctg gcctaagaga tttcaaagga gaagaatttc agaattgggc 780
ataaagacag tttttgtggt attttggtga agaattgggc taacttttgc ccttgcttga 840
aaagtctgcc tgagactaaa gtgaagagaa tcagattaat tgcattgaca aggggaagtt 900
gtggtcgcgc tatctggaaa cttacagcca gctccttga cctcgggtga cttacgcaa 960
tactcagggg cagagatgct tgactctgta ctgatgagt gtcttggatg caaatatggg 1020
ctcttcattt gactacatgt cagcatgagt caggagctgc tctctccaga gtgtgacaaa 1080
gcgaggggat gctgacggta gctgttctag ctttgaaggt aagcctgcac ttatgctaaa 1140
gtcacacata cacgagccgg tgggagaacc tgtctgtgtg gagacacctt tcattacctg 1200
tggcatccag cctctcaagc ttggactgcc tgtgtgctcc tggactctgg aggtccact 1260
gctctgtcct ctgctgctta tgactctgac atttttaaag aatccagtgg ttccccctg 1320
tactcgggtg ctacttctac ctggatgttc ctcatttatg ttctgtgaca cttctctgtg 1380
actctgctgc attcctgggt gacatgtgga caccctgtcc ctttgagac catgatgtca 1440
ctgtcactag tggaaatcaga tgccccaaagt gttgtcctgt gtttgggaac gtgacaggca 1500
gtacagaagc agaagaggaa ggggtgaaaac ggaaatgtca cagcagcatc tgatgtgtgc 1560
ctcagtcacg catgctgctg attggaacta ctcagcatga gagaggcca tgggtaatac 1620
acaaccttat acacactgtg tccatttctc tctctctctt acacagagag agagggagga 1680
gggggagggg gaggcggagg gggaggggga gggagaggga gtgggagagg gagagggaga 1740
gggagaggga gagggagagg gagagggaga gggagagttt aatgtctgtg aagagatacc 1800
atgaccaaag caactcttat aaaggacaac atttaatttg ggctggctta caggttcaga 1860
aattcagttc attctcacca tgggtgggaag catgcaggtg gatgtggtgc tggaggaacc 1920
aagagttcta tatcctgac tgaaggcagc caggagaaga ctgcctcttc tgcacagggc 1980
agagcttgag catagaacat caaagccctt cccacactt cctccaacaa ggtcatacat 2040
acttcaacaa agacacacct cctaaccgtg ccaactccctg tggaccaacc atttaaagc 2100
atgagtctat gaggttcaaa gctcttcaaa ccaccacact catgtacaca cacacacaca 2160
cacacacaca ctctcataca cacacacaca cacactcaca cacacacaca cacacacaca 2220
cacacacaca ccacacacac acacacacac agagttctat tttgcaactg ttcactgtca 2280
caaggttcta ctatctcag acacactgcc aggaattgtg tgggaagact ttcagtttct 2340
ttgggttcac atggacttag cagttcttgg tgatcctgaa agatttctgc agaaagaagc 2400
caaagtgttg agcccaaggc ctggccacac attagtctct tctagatgaa caggggttta 2460
aaaataaggg ggcataaggt tgaagccagc aggggctgac ttagagagga gaccaccca 2520
agccaactgc tcgaagtcaa aagcgatgaa tcccatact cagctgtgcc cggtgctgtc 2580
ttgctacatc tttagtaaat gttcttttag ttgtatgcgt atgaatattt' tgcctgcata 2640
tatttgtgta caccataggt gttcctaggg cctatggagg ccagaagagg gcatcagatc 2700
ctttggaact ggaattatag acacttgtaa cccatagagt agattgtggg aaatgagcct 2760
ttagtcttcg agagcggcca gtgctcttaa cctttggctg tttctccagg tctttgagac 2820
tttattttct tggacatcag gacaggatcc agggcttga gcttgtttct tcagccagct 2880
ttcttttcat gtatatataa ttttatgtta ttttgcttct tttttcccca agacagaatc 2940
acactctata tagctcaggc tgggtttgaa ttcagtttcc ctgtctcagt ctaccgggta 3000
atatgattac agatgtgagt ctgactttgg tatcaaagtc cccagccctt ctggatatgt 3060
gttttaagga tatcagatat atccttgatt tgctttgaat tttcttttta gttacaacat 3120
aattagttcc gtgtcacctg aatatgtgta tgtcacctac atagtcttcc ttcttctctt 3180
cttccctctc ccaccttccc aggtacctgt ctgtcttcat atccttgtgc tgagagtctt 3240
gttgagggag atgatgaccg agacagagcc actggggaag ggagatgggc tagtgcaggt 3300

```

```

cttcagagag gagctcgtga atattgtagc cccttttagtc cctggcatgt cctcttgtat 3360
agccaccgcc atgctgtggc ctggcagaag tgaataagtt gtccagctgt tgacaggcct 3420
qccclccaga cccagctctga tcccaagaaa gggcatctgt gtclgtctct gaggccgtaa 3480
gtgctgccty gttgtctcca gcttgacttg aactccctc ctttaataaga gtaccacaga 3540
acagggtctg cagagtcctt gggccagggt cctgtgctgt cctggaatgc caggcgtgaa 3600
tttctctgtg agtaggactt tgetcgccta gctcccacgg cttgcccttc agatagccag 3660
aattatctgg taccttgcac tgccttcaa tacgcagagt atcactggaa gcgcgcgcgc 3720
gcacacacac acacacacac acacacacac acacacacac acacgcccac tccatcttta 3780
aaccaccacc cccagcaacg gcggtgtaaa cactctccat caggaaagctg aaacgcagtt 3840
gccctctgct ggggagatga aggcagcttg ctgggggcga ggaccgtgct agcaaccttc 3900
cctgglgcac acgggctctg glgcatgacg ggaacggaaa cgcggaaacta aagtcagtc 3960
tgcttttttt tttttttttt tttttttttt tttttttttt tttttttttt ggcgttggtg 4020
gtggactgag tgacaatcag tgaaatcact taggttggtt ttctctctct cgttggtgtt 4080
gatagacggt gggagagggg cagaggagaa ggggagggat ggggagagag ggaggaggga 4140
ggggcgggag gcggggggcg aggaaaacgt gctaacttct ccaatcctac aagacaaagg 4200
tttgagagaa gccgcactga gtgacccagc agaaggaatc caggaaatgt cgctggaatc 4260
tgactgttga ttccagcgcc atgcagagaa tctaggctgg taggaacatt ctttgtccta 4320
tccgacataa taactccaac caacacggaa aagaaaaggct atacaagtga agaaatgca 4380
ttttcacttt catgactata caatcacttc caggtagtaa cacglgtcta gcacagcgg 4440
tctcaacctg ggggtcacga tccccactt ttctgcatac cagacatttt tacgttgtta 4500
ttcataacag tagcaaaatt gcagctatga agtaacaatg aaatgcattt atggtgcgtg 4560
tgtgtgtgtg ttgggggggt taaccttaac atttactgta agaaggttga gaatactgct 4620
ccagcagcta gtgtgttga cttaggttct gggatatatta ttagcaatag ccaaccagaa 4680
tccccacca ccacagcatt gaggcccat gcagggtctg ctgggagagg cactgataag 4740
acttctttat gtatttattt agagacgaat actcattagg taggccaagc tagcgtcaaa 4800
ctcatggcaa ttctctctc cagtttctc aagtaactga ctcaggagtg tgttgccatc 4860
atatacagta aggatttatt gactgaagaa aatctcaagt ggctttggtt aatccctact 4920
acgccagagg ctgagycagg aggcgcgcaa ggtcaaggct tgcttggcct acatatagag 4980
tgagctcaat tttagacact tttagacact tttagacact tttagacact agtaaaagt 5040
tcaggtgggg ccggtgattg gacacacttg ggtctctctg gtccatctgc agctgtgcaa 5100
caggaaagag gggaatgag aggaagagag gaaaagacag aatgagagag agggaggaa 5160
agagaaaaag gaaaagagag aggaagagaa aaaggaaaat gaggaaagcg agaaagaaga 5220
aatgagaaa ggaagaggga gaaagaaatg agagagagaa aagaaaagac agaatgcgag 5280
agagggagga agagagaaaa aggaaaagag agaggaaggg aaaaaggaaa atgaggaaa 5340
cgagaaaaga gaaatgagaa agaggaaggg gagaaagaaa tgagagagag aaaagaaaag 5400
acagaatgag agagaggag gaagagagaa aaaggaaaag agagaggaa ggaagaaagg 5460
aatgaggaa agcagagaa agaaatgag aaagaggaaa gggagaaaga aatgagagag 5520
agaaaagaaa agacagaatg cgagagaggg aggaagagag aaaaaggaaa agagagagga 5580
agggaaaaag gaaaatgagg aaagcgagaa agaanaatg agaaaggga aagggagaaa 5640
gaaatgagag agagaaaaaga aaagacagaa tgcgagagag ggaggaaag agaaaaagg 5700
aaagagagag gaagggaaaa tggaaaatga ggaaagcgag aaagaagaaa tgagaaagag 5760
gaaagggaga aagaaatgag cgagataaaa gacagaattt gagagaggga ggaagaaata 5820
ggaaaagaga ggaagggatg gagaaaagag agaaagaaag agagatgaaa gagagaaagg 5880
agaaatgaaa tgagagagag agagagacac aaagagccag agagagaaga aaaaagggga 5940
aagagaaaga gaaagaggaa ggctctctct ggacacatct tcttttatct ttccctggg 6000
accgcaaaag cctggtggca tactgtacat lctgtacact gttcattcaa aacaggctct 6060
gtcttaaaag tggctgagc ggtcagaaaa gggattgttt aacttgtttg caaaactgcc 6120
tcaggagagt gctgagtgc tgaaagtgc tgccggttaa ggagaagtct ctactacttg 6180
tgatctcacc atcgaaaatt tctttaattg tctcctggtg ttctgggttt tgcagttttg 6240
tttctaagga tacattcttg ggtgatgca caaagtcccc aaagacacgg tggagctgtg 6300
ttagatggg aaagacagtc tgetgaggat ttatctggaa ctgtcagaag gaaaagaagg 6360
taaatgggg acttgggaaa gtggcctcta gtttgacttc tggttagca aaggttgtgg 6420
ggagataagg catcacagt agttagcagg aggcaacagg gtccctgggag gacgcaggc 6480
agaaggagag gctgggctga cagcatgcaa tcattgcata gtctccaaag gagattgcaa 6540
catggctgag ttttcaagg tcctacagag cccgtggtag agattctgtg ggttctgaga 6600
caacttgact ttagccagat ggtatttgag taatctggga gagagaaaac agctacagca 6660
aacagggcca catttagtga actttgactg actttgactg ttgagtcatt tgcagtggc 6720
cctgaggtca ggctggccct cagctcaaaa acaagcgagg aactgaagca attactcaga 6780
taatccacag ccacagccac tggaaagggc cacatcccca gagacagcac agcaggggtg 6840
ggggtggggc tatgagaaa ttagtgattg tagcagttat ctagaatgtg cggagcagag 6900
gaggttacac aaaaacctag aatgtcattc aatgtgggaa accgagaggc tcccaagccc 6960

```

taaaaggaac	agtttgcttt	cagccaaaat	ggaaataaaa	tttggggctt	aaatctggca	7020
aatgattcag	accttctgtg	taggtgtctt	taaatgcaca	gcagattgat	tttcatgttg	7080
gagtttattt	gaactaaaag	acagaaatgg	tgaaaagcac	acctgaagaa	attgagatgc	7140
talgaataaa	alcatttact	lacagctatc	acttaattag	tacctccttc	caccttgctg	7200
atttattggg	ctagtcaagg	aagaaaagat	cttccctcct	ccttctctcc	tcctccccc	7260
cctctcctcc	tccctccccc	tccttgacct	tcctctcctc	cttttccctc	ctccccctct	7320
tcttctcttc	acccctcctt	ccctccctct	cctctgtact	cctccctttt	cctcccaatc	7380
tcttttttct	cccccttctt	ctctttctcc	ccctcctctt	tcctcctctt	tcctccctcc	7440
ctccctcctc	ctcctcctcc	tcctcttctt	cttcctcctc	ttctccttcc	tcctctcctt	7500
cctcctcctt	ttccagccct	acctaccttc	cctttcttct	tcatttatct	aaagtagctt	7560
tgaacagcac	tactcgggtt	agtgtgtgat	aaaaggaaaa	tgcagggtcc	agcagcttgg	7620
ggaagattgc	tttttgcctt	ctggaggcag	atgatgacag	ttcaagatca	ttccttttgc	7680
tcctatglc	aggaaggggg	acatgccgaa	tctaccagtt	tgcagccacc	tacacaggat	7740
ccaccttcac	ttctaaggaa	atgtttggga	agctacctac	caaccacttc	tggcatctca	7800
tgggctagag	gactcttaaa	tggcactcct	atttgtttta	taaaggaggt	tgtgacgtgt	7860
agttttaaat	cccttcacac	caacaattgc	tactctctga	ccaaaaaaga	agggagacag	7920
gatacggcta	gggtcttagt	agactttacc	actttgaaaa	gccttaatat	aatcaggta	7980
gatacatctt	tttaacttat	tctgttaaag	acaaaaacaa	aactttatct	ttatttctgt	8040
gtatgcttgt	gtgtgtgtgc	ctgtgtgtat	accacatgtc	gctggtgccg	gagaacacca	8100
gaagagggga	cctgatctcc	tggagctaaa	gctatccatg	gttctgagct	gctgatgtgt	8160
gggtgctggg	acagaactct	ggctctctgc	aagagcaaca	agcctcctct	taactacgaa	8220
tctcctcccc	atccccccaa	ttattcattt	ttattcattt	tagcagcttt	atttcgtaac	8280
tacttalcac	agcataaaac	aaggatttta	tatatattac	atgcaatcga	ggataagagt	8340
tgaagggaga	tgcgtgtgct	ccttctgggt	gtctgtgctt	ttgaagaatg	taagcagtgc	8400
acaagggacc	gaggcgtgcc	tgtctgccag	gagctgtctt	cttcccttgg	actctgagct	8460
gagtgcagt	gtccgaagaa	gtaaaagacg	acctcatgaa	gcaatgtctt	caacccaaac	8520
atgctgtcca	gacaaagtcc	agcttcatta	gtgctctgag	gagagactta	ctgagcctca	8580
ggaaagcccc	cctcagcatg	gcgaaagtcc	actttgattg	aagtgactcg	aaagccatgg	8640
cagtgcggcg	gcggcgcgct	ggagcttgtg	ctcgagtcgg	aagcggcatc	tttgtcaggc	8700
ggctgtgatt	agcacgggga	ggcaggactg	gagtgaagga	agagtggggg	gcggggctta	8760
gcgctctggg	ctcctaagct	gtagtcagcg	cctcaagatt	tgtaacctgc	cttctgcctt	8820
cccagccagg	cagtcaagtg	gctccaagct	gaagactgca	aagtgcctct	aaccttttgg	8880
ttatagcgag	gctgaagaca	ccglgctctt	tcatgaagac	cggatgtctg	aaatccgat	8940
tgataaatat	ggataaaaacg	tataacgctc	gatcaatcga	atcgaaggag	ctcagcattg	9000
gcaccacggc	tttggggaca	acagagtact	gactcgttgg	gaggacttgg	atacttcccc	9060
tcctcttcca	tctcttcccc	tttctcactt	tcctcctcct	tccttctcca	ttttctccct	9120
cttcaactgt	tcttactatt	tttacaagag	attttattta	tttatttatt	tatttattta	9180
tttatttatt	tatttattta	tttatttaat	gtatgcgagt	acactgtagc	tgtcttcaga	9240
cacaccagaa	gagggcgtca	agttccatta	gagatggttt	cgagccacca	tgtggttgc	9300
ggggcctctg	gaaggacgcg	cagtgctctt	aacctctgag	ccatttctcc	agtaaccttc	9360
tcaccgcttc	tcttcaatct	tcttctctct	ccttctccac	tttctctgtc	ttcttggttt	9420
catttatctt	ctccctttct	tcctcttctc	cccttcttcc	tcctccactg	tagttttcct	9480
tccttactct	tttctgctt	ccctcctcct	ccctctctat	tcctcctcct	ctttctcctt	9540
tctcctcctt	cctcttctct	tctcctctct	ccctctcccc	tctcctctct	cccttctccc	9600
cctcctcttc	ctctttctct	ttctccaccc	ctcctgtcac	agtatcaatg	gcaaggggtg	9660
tctagaatgg	aggagtgtcc	cctaggcact	aacgaaagcc	agttaggatg	ctctgagacg	9720
ggtaacaatt	agggagggcc	gtggggatgg	aagggttgtg	ctgcnattca	ttctggagca	9780
acccccaggc	agaatcatga	ggttggttcc	ggattcgcag	ggcacaattc	agaagaggaa	9840
ggtttcagga	aggacgagtt	tgtctgagat	aggagttaca	tctgatgtct	tggcagcaga	9900
gccactgtac	aagcgtgctt	tattaaccac	gtgggattaa	atcttctttt	aaatttattt	9960
tcaactctta	aggaaacgtg	aactttcaca	ttcaaattta	gaattgcagc	tcttatgggg	10020
aaaaaaaggg	gatcttaaga	atattaagca	tagggcgctg	gagagatggc	tcagcgggta	10080
agagcactct	ctgctctccc	agaggtcctg	agttcaattc	ctagcaacca	cataatagtt	10140
aacaacagtc	tttaatgaat	tctaattccc	tcttctgggt	tgtctgaaga	cagttacagt	10200
gtactcata	aaataaaaata	aagaaattta	aaaaaatgaa	tattaggcat	agattcctgg	10260
atcctaagaa	agccatcaga	gctggagcca	tgtgtgggat	cctgcttggg	gctggagggg	10320
cagagttcat	gccccggggg	tttttactta	ttatcacatt	ttcatcgttg	ttttgaaaaca	10380
gggtcttgtg	tgggtccaggc	tggccttgaa	ctcatcttcc	agcctctacc	tcacagggtc	10440
tgggattact	tgggtccctaa	aagtatctcc	gtcaagctcc	ctgggtgtat	ggctgtgcca	10500
accaggaggg	tctatacaact	cgtcaggtga	gagggagaa	atccgaatct	ctgacaggga	10560
ctgctgcctc	tccggggcaaa	tggagtgaag	gacagcggca	gaaggattta	ggaaagatgg	10620

acgggagagt ggaaatgctg cagaagccag aaaacaaagc aggaagcctg ctgtccagtg 10680
 gggctcaaga gcggagggat gcgagggggc tgcgcaggaa catttagcgt ctgcgtctat 10740
 gggggtaggg gcgggggtgcc agcacctaqt cacctgaagg ggaaatgctt gcccaggagg 10800
 caggctctcag tagctgacct agagaaagga gcggccctta cagaggagac acgggtcact 10860
 gtttgctaaa gtgaaggaga aataaatatt ctttcaaaga atcttaggtg agcccgattc 10920
 atctgcgctg tggagggcctg gggaacagtt aaaaagaccc tgacacacac ccaaggcaaa 10980
 caagcaacac agggctcctt ccgtaagggg ccatgattct ctgaagaatc agccccggaa 11040
 tcagccccgg aatcaggtag tccgtaaaaca caatgagtggt tttactctgc agaagtcag 11100
 cctgctggcg tctcccatga ccaaaataga gggatagtc cgtgagctca ccggctcgat 11160
 ttaaggcacg tgggttttcca gggtagatga gctttggctt ctggaacccat tatggggcac 11220
 gaaggatgga gccaggattt tttttttttt tttttttttc tattagcaat tgatttgctt 11280
 gggcttggct ggacttgccc agttcttagg cccagtcttc llaactgccg atctgaagtc 11340
 tgtcatggag tcagcctagc cttctcactt ccttcagct cgaataggaa gaggaggtgc 11400
 acaccagatg gtctgagagc agggataaat ggtgtgcctt tgtctttcag tatctcgtta 11460
 ttttaagtag gaagatgctt ttctgtatta cattgcttgt. gaaacccgaa gttgattcgg 11520
 ggcacaggac aatggatttg gtgttttgca aggactgttt cagaagagag aggagtgga 11580
 ggggtggttag agtgaggagt ggggtgggac gggatggggg aagagaagga agggccagac 11640
 aggctaggtta gggctgagag gaggcgggtg gaacttcttg agttagcgca gcagtaaaact 11700
 tggatgtgctg tgtatctttg tgatatatga cccggagccg tgtagctggc tccgatagta 11760
 ctgctaagtgt cagtgtcggg gggggggggg cccataactgt tccacagggg ctgcacattc 11820
 ccatcgagag caggaggggt cctctctcca tacatcctcg ccagcattcc ttgttgtttc 11880
 tgtgatgaca ggggtggga tgaactctct ctgttggttt gagagaccgt gaagaagctc 11940
 aaccccaqga catlllgcag tcttggaagg cagtgcctcc atgtggagcc gtggagccca 12000
 tctctgagtc cagtcactc ttgcagttcg cactcagctc ttcagatgca ggagagacgt 12060
 tgggtggaaa gcaagattgt ttgcttgttg agatagacac attctccaca caaaggctca 12120
 cgtggggcaa aggcgtgatt acgtacagcg ttcaaggacg cctgtggtag agctatgatt 12180
 agctgtctcc atctatgaag cagacaaaga gttataaaaa aaatcaatgt ttccaaattg 12240
 tcaaaactttt aacccgacag caagcgtctc gtccctgggc taatccctag ccttggtttc 12300
 ttgagatggg gtcttttggt cactagactg gccatagaact caccgatctta gtgttccagc 12360
 ctcccagctg ctgggatgag ccgctataac cagtctgcct gccttcctaa attttaagtg 12420
 atgggaagtg ggggagaata tatgcagatc tatgcagatc tgagagcagg aacctggcaa 12480
 agccaagggg ccggagttac agggcggtta catggtgtct gggaaactgac ccaggctcctt 12540
 gagaggagca gtgtgtactc ttgaccaaac aggtccgtct ctccagtcct cgtagtatta 12600
 aaaataggla clacgggcatt ggtggtgcac acctttaatc ccagcactag ggaggcagag 12660
 gcagggtgat ttctgagttt gaggccagcc tggctacaaa aatgagttcc aggcagcca 12720
 cggctatata gagaaacctt gtcttgaaaa caaaacaaca acaaaatagg tactacaaag 12780
 cgatgtaatt gtgtcacaac atgcaaaccg aggggactgt atgcataaga aagagaaaga 12840
 cggccacact ggttctatct ggggtgacagg aaatcagtat ttttattttt cacattcatt 12900
 tttttgtgtg tgtgttgac acagtgattt ttctatcaaa aacattattt cttttatagt 12960
 tccccagagg agctgttttt tttgaaaaaac cattgaagga gcagaggcag 13020
 ggagactcct gtgtggcagt cgggtgaagca ggcctctctg aggcaggtg gccctggact 13080
 tgggagttct tttccctccc tectgtgtct aatatgcaaa tgtcaggctt caatgtagct 13140
 agaaggttct agaattgatta agtttccaag gctgaagagc ttcctgttll gcclttcaat 13200
 tcctggaga tgtcgtttgt tgttccggag tclgcaaggt gcctttggtg atgcgggtgg 13260
 ttcatctcgg gagattccgc ctggaggacc caagttcaag ccttgcctga gctacagagt 13320
 gactttcagg tcttctgcgc aattcagtga gacccagctc. acaaaataaaa agtaaaaaaga 13380
 aggctgtgga tggaaactcg aggaagaaag aagaagagaa gggaagagga gaaggaaagg 13440
 gagggagagg gaggaaggaa gagagaagag ggagggaagg gagggaaagg aaggggaaag 13500
 agggaaagggt ctgacaagaa caagaagaga gaaggggag ggaggggagg gaaaggaagg 13560
 gaagggaagg aaggggctga aagaaactgt tccaatggtc tggggccacag agtgatggcc 13620
 ggaaagaaga gaagggtgta atccttgatt tgacacaacc tagaatctgg gaagcgagtt 13680
 ttttgggtg atcagctgta ctggctggcc tgtggcgctg catgtgggag actgtcataa 13740
 tctgtgaagg agcattcaca agtcccagcc cactacaaat ggcttcgttc catacccaag 13800
 ttaggttcat taatacagga tggagaagc aagcaagctg tggatacccc acgctctttc 13860
 agatgctaac tglagcgggt ggtgactgt gtctcttggt attttaaagt cctgcttga 13920
 acctcggctc ctgggggggt gtaactggaa ttgtgagctt tagtccttta gttttctacg 13980
 cgtccctgct gtgacagact ttatcgcagt aacagaaaca agaccaggac acttgatctc 14040
 ttgggttttc tcaggatatt ttacaaaaca ggctgaggaa acaaaacttc ttctccctct 14100
 ctctgatcaa cactgaagag tectctctgc tccctccctt gccctctctc tccctgtctc 14160
 cccctctctg tccctccctc tctctgtctc tgtctctgct tctccctcc cctccctcc 14220
 tgtctctgct tctgtctctg tctctgtctc tgtctctgct tctccctcc cctccctcc 14280

ctctgtctct gtctctgtct ctgtctctgt ctctgtctct gtctctgtcc ctttctctcc 14340
 tatctcctaa atggctggag gccatgctag ctcaatggtt aactttgaac acgtatttag 14400
 gaaatctttt ttcttaacag ttctgaagtg ctgaagtggg ggttttagtct ctccggcctga 14460
 caagctcact tctctcact ctgtcttaat gaccaaactc gccatttccc taaaacagca 14520
 caggctccag ctccaggttg ctccggagcg gag 14553

<210> 7
< 211> 34
< 212> DNA
< 213> Kunstig sekvens

<220>
< 223> Syntetisk

<400> 7
ataacttcgt ataattgtatg ctatacgaag ttgt 34

<210> 8
< 211> 34
< 212> DNA
< 213> Kunstig sekvens

<220>
< 223> Syntetisk

<400> 8
ataacttcgt ataattgtata ctatacgaag ttag 34

Patentkrav

1. Isolert celle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5, og minst ett rekombinasegjenkjenningssete som befinner seg i en nukleotidsekvens i SEKV ID NR: 1, 3 eller 5.
2. Isolert celle ifølge krav 1, der det minst ene rekombinasegjenkjenningssetet er uavhengig valgt fra gruppen som består av et loxP-sete, et lox 511-sete, et lox 2272-sete, og et frt-sete.
3. Isolert celle ifølge krav 1, som videre omfatter et første gen av interesse (GOI) som koder et protein, der det første GOI-et er integrert ved det minst ene rekombinasegjenkjenningssetet; der GOI-et opsjonelt er funksjonelt knyttet til en promotor.
4. Isolert celle ifølge krav 3, som videre omfatter et andre GOI som koder et protein, der et rekombinasegjenkjenningssete er til stede mellom det første GOI-et og det andre GOI-et; som opsjonelt videre omfatter et rekombinasegjenkjenningssete 5' til det første GOI-et og et rekombinasegjenkjenningssete 3' i forhold til det andre GOI-et.
5. Isolert celle ifølge krav 1, der cellen er en CHO-celle, og der den ekspresjonsforsterkende sekvensen er SEKV ID NR: 5.
6. Isolert celle ifølge krav 5, der rekombinasegjenkjenningssetet befinner seg ved en posisjon mellom nukleotidene 5.000-7.400 i SEKV ID NR: 5; nukleotidene 5.000-6.500 i SEKV ID NR: 5; nukleotidene 6.400-7.400 i SEKV ID NR: 5; eller nukleotidene 6.400-6.500 i SEKV ID NR: 5.
7. CHO-celle som omfatter en ekspresjonsforsterkende sekvens som er nukleotidene 6.400–6.500 i SEKV ID NR: 5, videre omfatter minst ett rekombinasegjenkjenningssete flankert 5' eller 3' av et humant gen av interesse (GOI), der det humane GOI-et opsjonelt er flankert 5' av det første rekombinasegjenkjenningssetet, og CHO-cellen videre omfatter et andre rekombinasegjenkjenningssete som flankerer 3' det humane GOI-et.
8. CHO-celle ifølge krav 7, som videre omfatter et andre humant GOI som flankerer 3' det andre rekombinasegjenkjenningssetet; som opsjonelt videre omfatter et tredje rekombinasegjenkjenningssete som flankerer 3' det andre humane GOI-et.

9. CHO-celle ifølge krav 8, som videre omfatter minst ett markørgen mellom det andre rekombinasegjenkjenningssettet og det andre genet av interesse.

10. CHO-celle ifølge krav 9, som videre omfatter en promotor som er funksjonelt knyttet til det første humane GOI-et, og en promotor som er funksjonelt knyttet til det andre GOI-et; der promotorene opsjonelt er eukaryote promotorer, og de eukaryote promotorene er funksjonelt knyttet til en prokaryot operator.

11. CHO-celle ifølge krav 10, der det andre og det tredje rekombinasegjenkjenningssettet er i en orientering motsatt av det første tredje rekombinasegjenkjenningssettet.

12. Isolert celle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5 og minst ett eksogent tillagt GOI i den ekspresjonsforsterkende sekvensen.

13. Isolert celle ifølge krav 12, der cellen er en CHO-celle, den ekspresjonsforsterkende sekvensen omfatter SEKV ID NR: 5, det minst ene eksogent tilsatte GOI-et er et humant gen, og det humane genet er funksjonelt knyttet til en eksogent tilsatt promotor.

14. Framgangsmåte for å lage et protein av interesse, som omfatter:

(a) å tilveiebringe en vertcelle som omfatter en ekspresjonsforsterkende sekvens valgt blant SEKV ID NR: 1, 3 eller 5, eller nukleotidene 6.400–6.500 i SEKV ID NR: 5,

(b) å sette inn i vertcellen, innenfor den ekspresjonsforsterkende sekvensen, et gen av interesse (GOI) som er funksjonelt knyttet til en promotor;

(c) å opprettholde vertcellen i (a) under betingelser som gjør at GOI-et kan eksprimere et protein av interesse; og

(d) å hente ut proteinet av interesse.

15. Isolert celle, CHO-celle eller framgangsmåte ifølge et av de foregående kravene, der det første og/eller andre GOI-et koder et protein valgt blant en antistoff-lettkjede eller antigenspesifikt fragment av dette, en antistoff-tungkjede eller antigenspesifikt fragment av dette, en ligand, og en reseptor eller ligandspesifikt fragment av dette.

16. Bruk av isolert celle ifølge krav 1 eller CHO-celle ifølge krav 7 for å lage et antistoff eller fragment av dette.

EP 2 150 617 B1

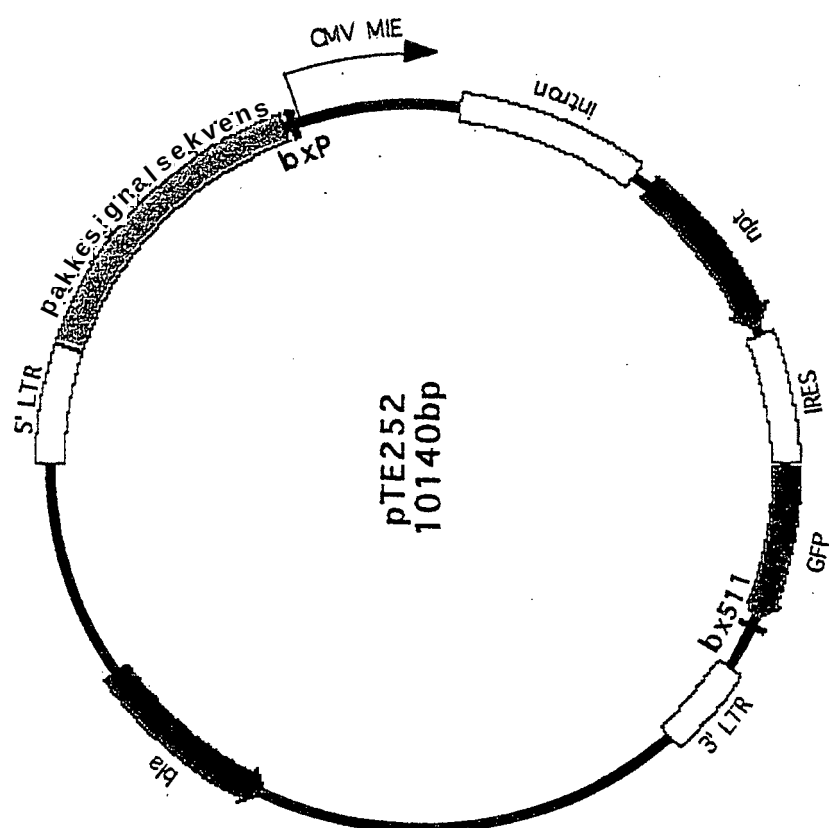
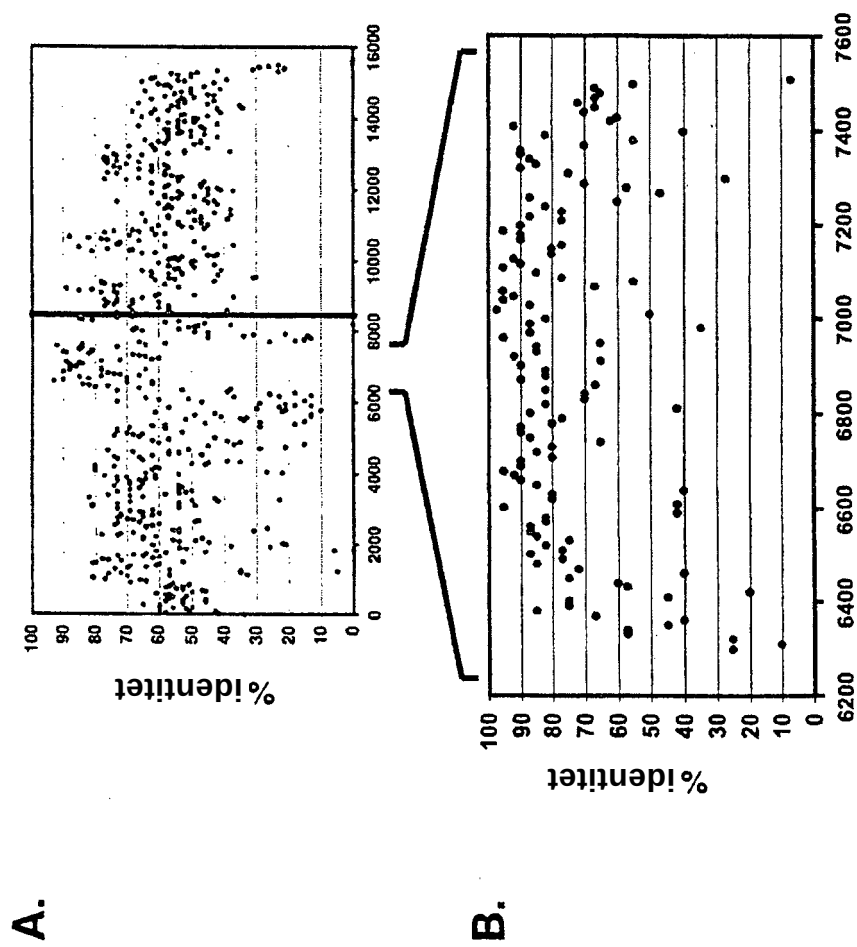


Fig. 1

EP 2 150 617 B1

**Fig. 4**

EP 2 150 617 B1

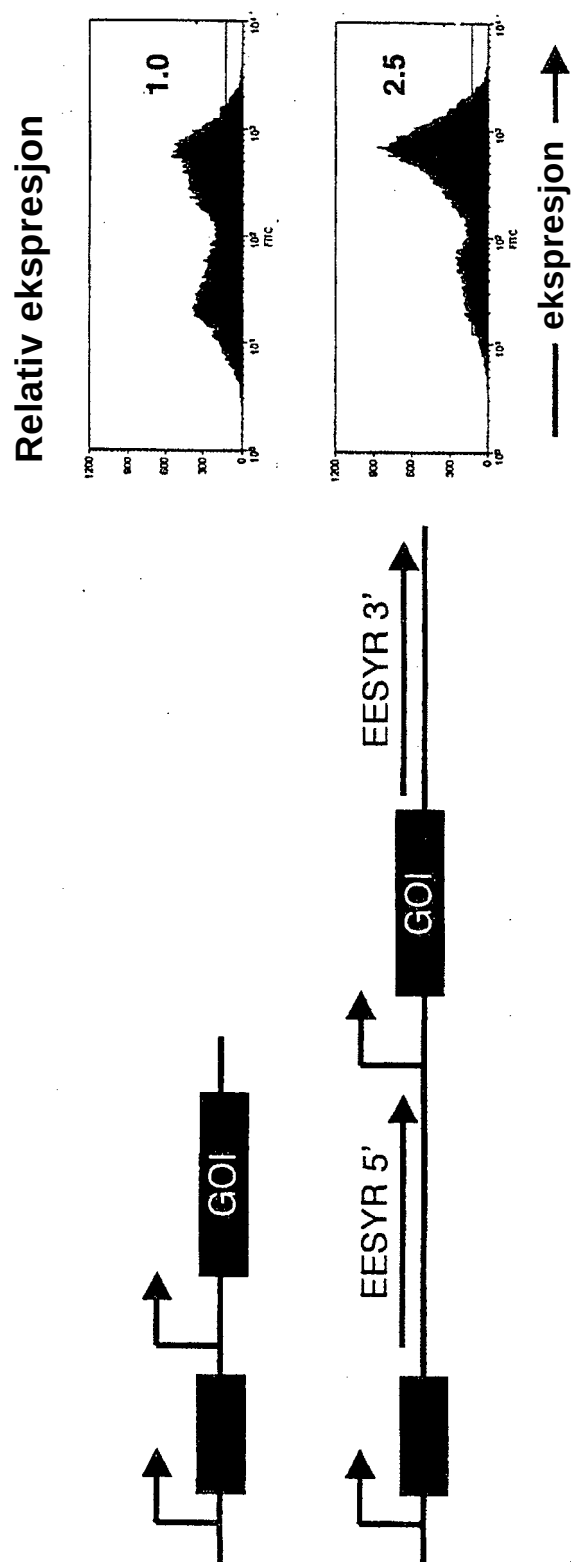


Fig. 5

EP 2 150 617 B1

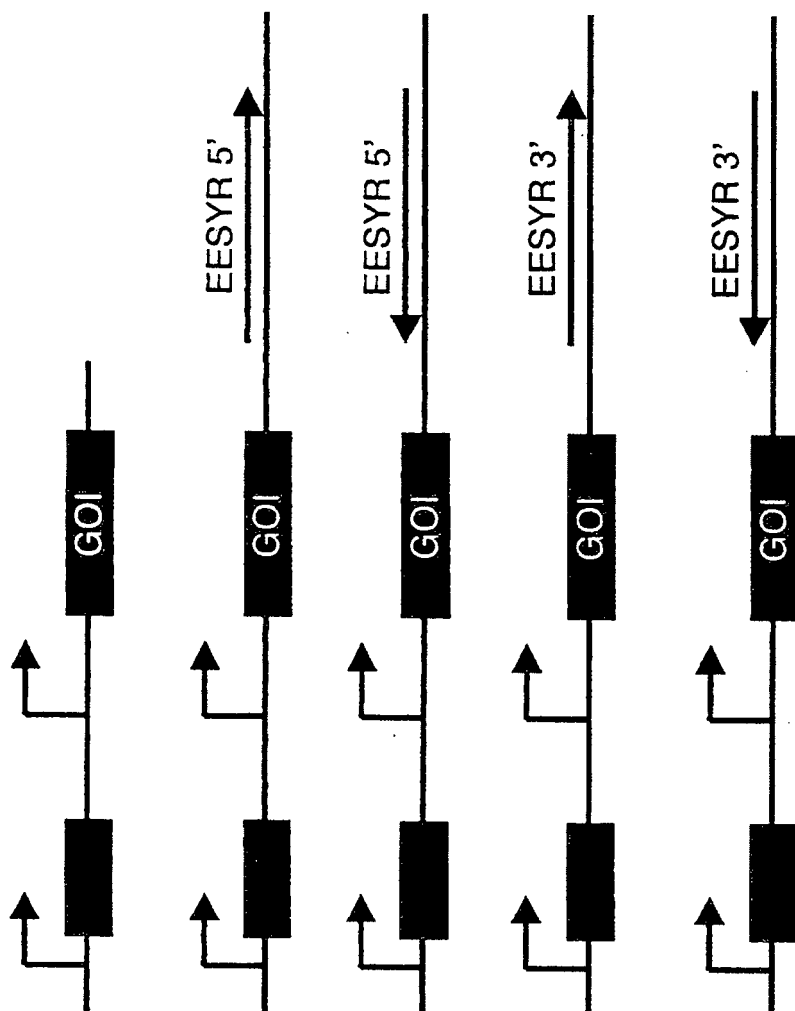
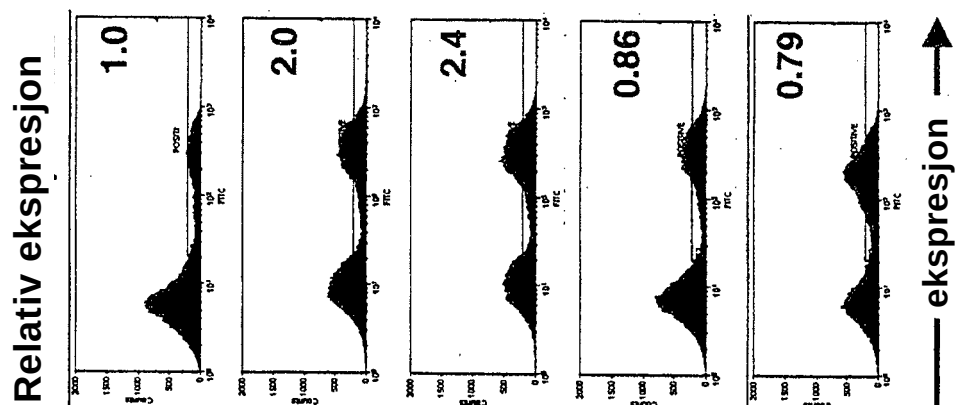
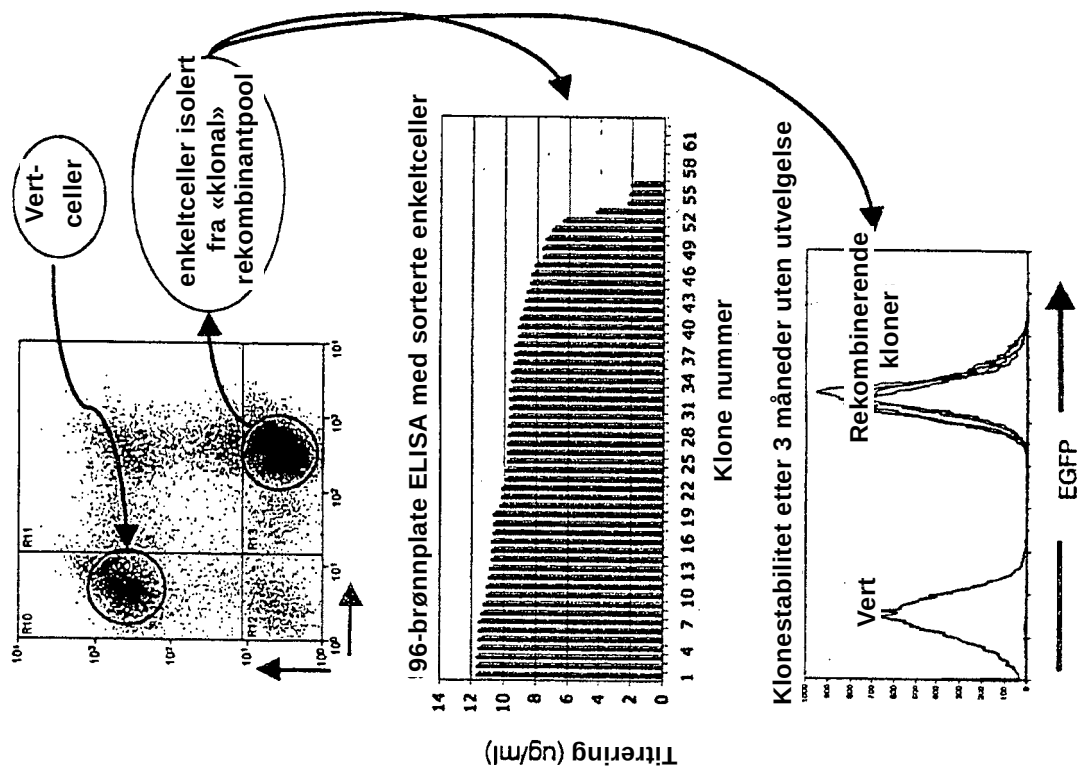


Fig. 6

EP 2 150 617 B1



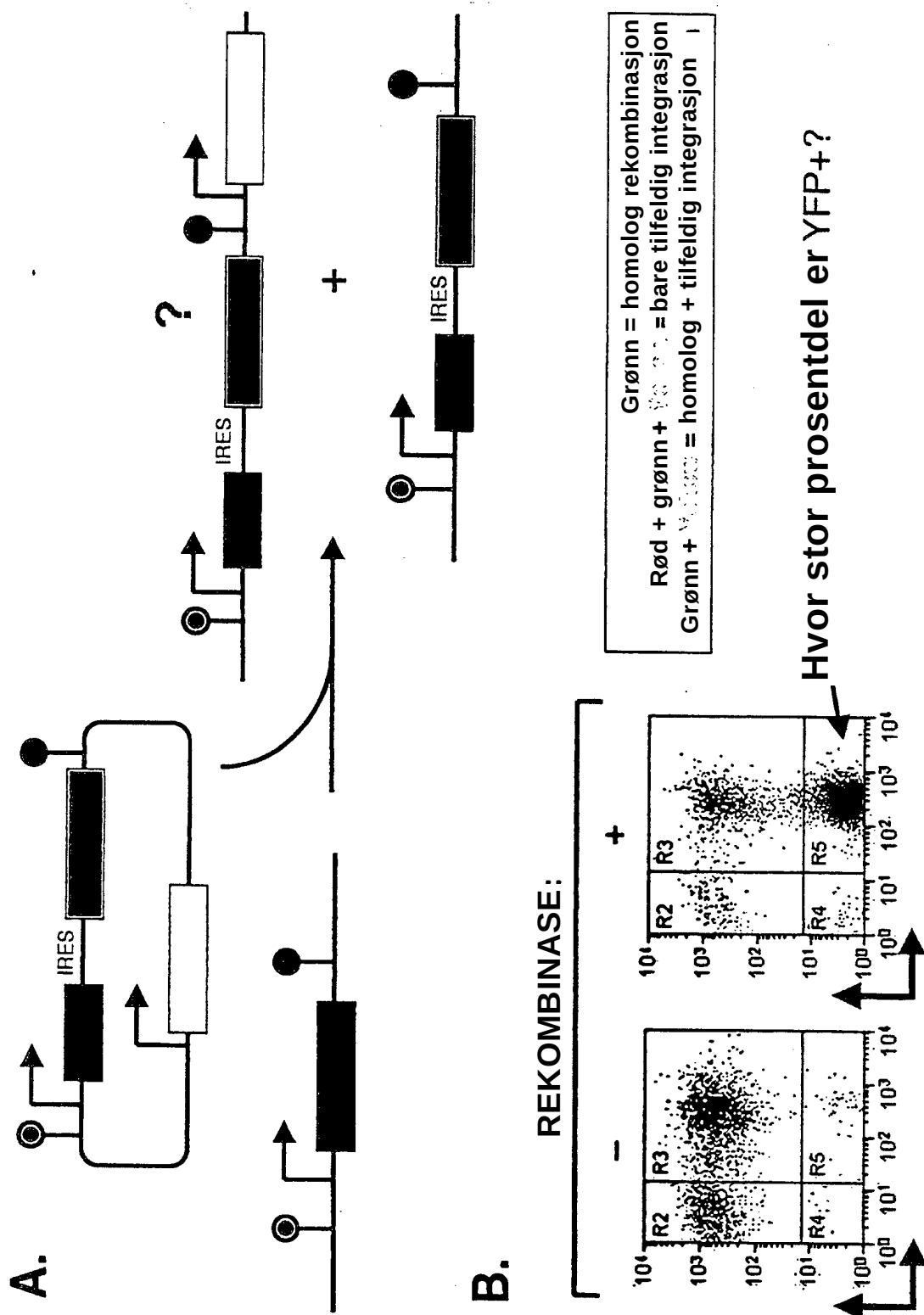


Fig. 8

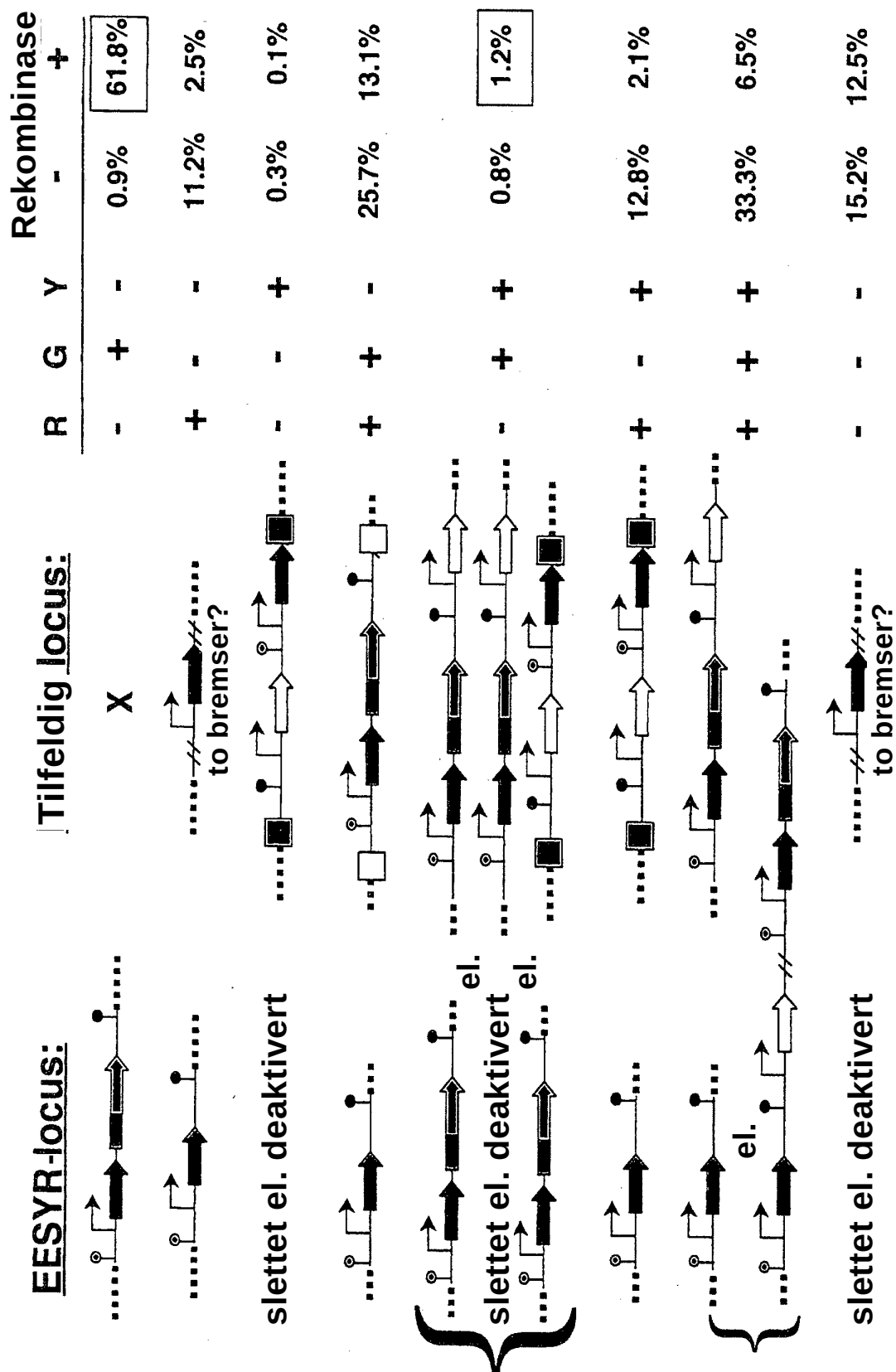


Fig. 9

EP 2 150 617 B1

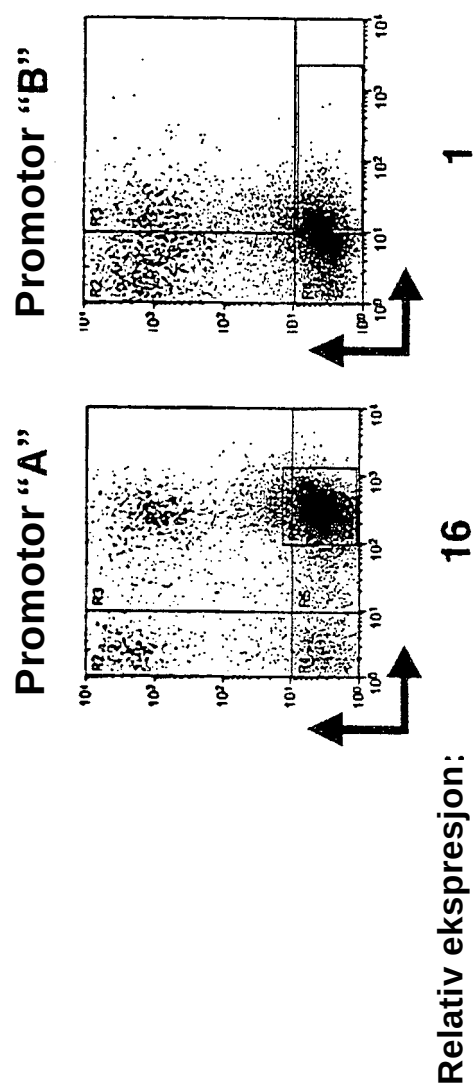


Fig. 10

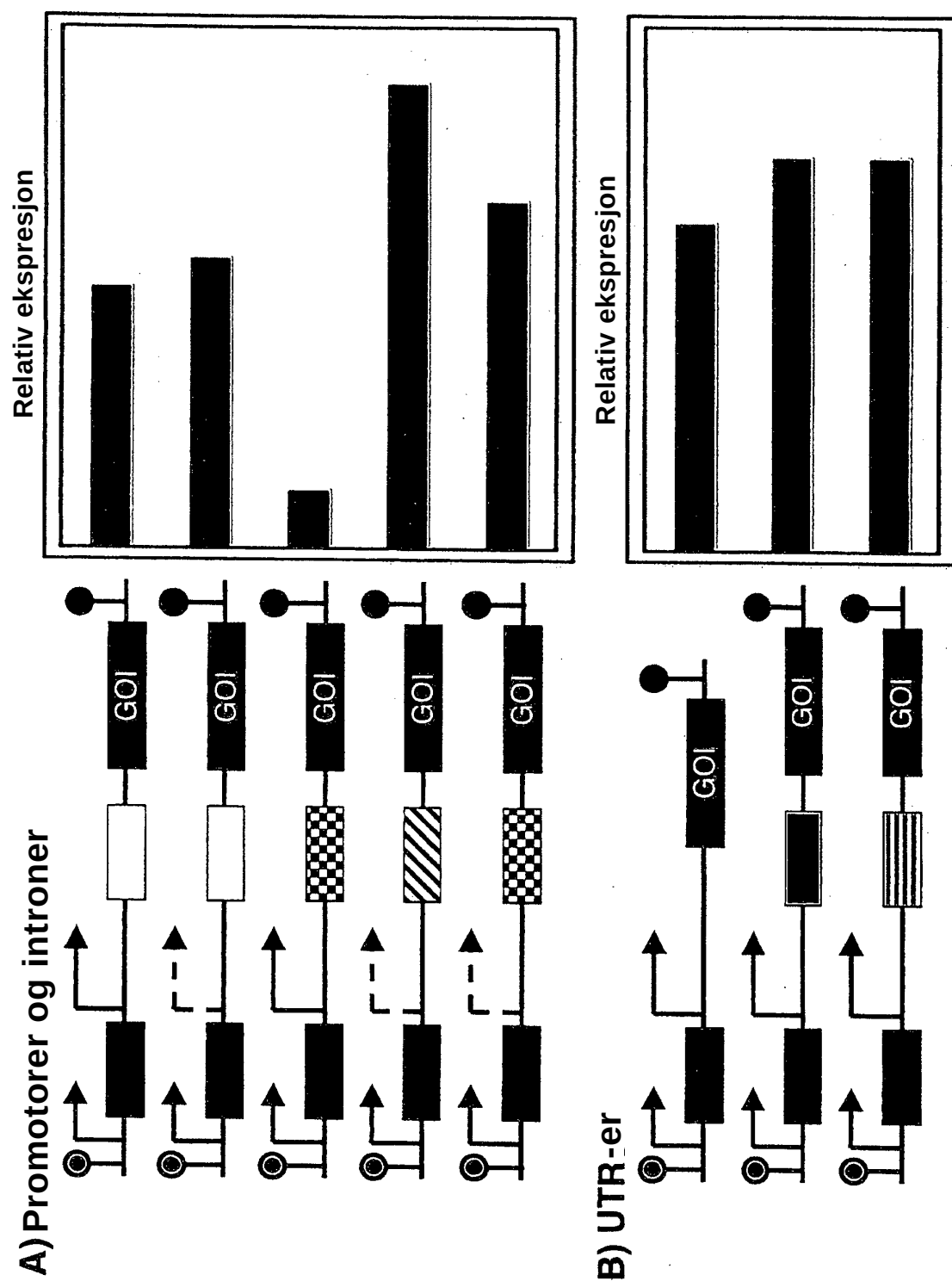
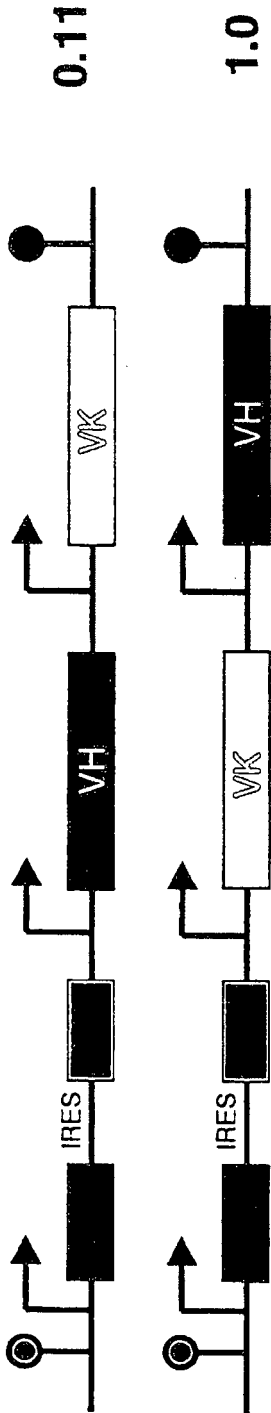


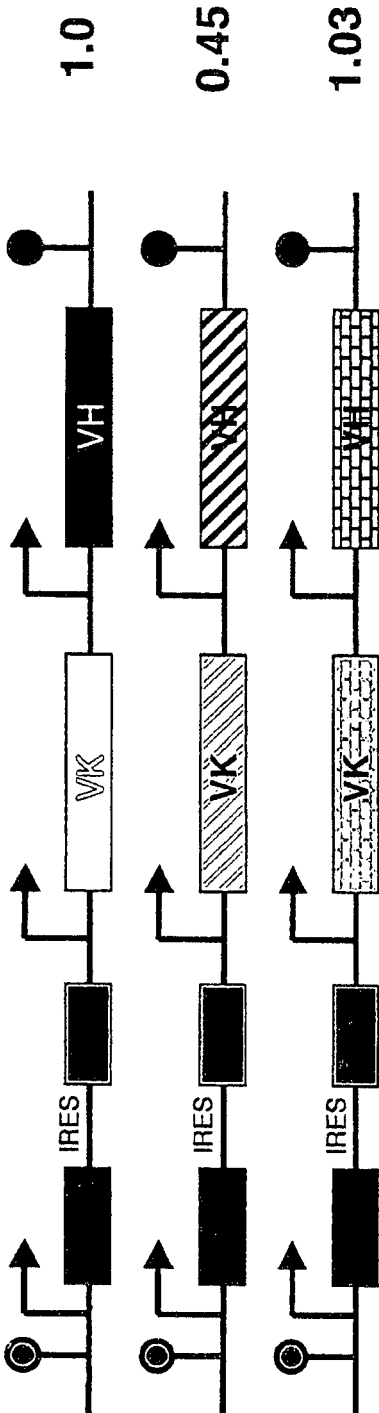
Fig. 11

Antistoff: Tung- og lettjedeorienteringntation

Relativ ekspresjon



Kodonoptimalisering



Fia. 12