



(12) PATENT

(19) NO

(11) 334836

(13) B1

NORGE

(51) Int Cl.

C12N 15/81 (2006.01)

C12N 9/10 (2006.01)

C12P 23/00 (2006.01)

### Patentstyret

---

(21)	Søknadsnr	20052072	(86)	Int.inng.dag og søknadsnr	2003.09.23 PCT/EP2003/10573
(22)	Inng.dag	2005.04.27	(85)	Videreføringsdag	2005.04.27
(24)	Løpedag	2003.09.23	(30)	Prioritet	2002.09.27, EP, 02021619
(41)	Alm.tilgj	2005.06.27			
(45)	Meddelt	2014.06.16			
(83)	Biol.mat. dep	ATCC 74438			
(73)	Innehaver	DSM IP Assets BV, Het Overloon 1, NL-6411TE HEERLEN, Nederland			
(72)	Oppfinner	Kazuyuki Ojima, 3-22-304, Zengyodanchi, JP-251-0877 FUJISAWA-SHI, KANAGAWA-KEN, Japan Yutaka Setoguchi, 4-27-2, Kataseyama, JP-251-0033 FUJISAWA-SHI, KANAGAWA-KEN, Japan Tatsuo Hoshino, 2-18-14, Fueta, JP-284-0027 KAMAKURA-SHI, KANAGAWA-KEN, Japan			
(74)	Fullmektig	Oslo Patentkontor AS, Postboks 7007 Majorstua, 0306 OSLO, Norge			

---

(54)	Benevnelse	<b>Rekombinant organisme, vektor, fremgangsmåte for å lage organismen derav og fremgangsmåte for å produsere karotenoider.</b>			
(56)	Anførte publikasjoner	Genbank Accession number:X99718, 19960915, Corran A.J. Genbank Accession number:B84401,19991120, Kikuti et al. Shimada H. et al. Increased carotenoid production by the food yeast candida utilis through metabolic engineering of the isoprenic pathway. Applied and environmental microbiology. 1998, Vol. 64, No. 7, side 2676-2680. Robinson G.W.et al. Conservation between human and fungal squalen synthetase: Similarities in structure, function and regulation. Molecular and cellular biology. 1993, Vol. 13. No. 5, side 2706-2717. WO 9609393 A1			
(57)	Sammendrag				

Den foreliggende oppfinnelse vedrører et gen nyttig i en prosess for å øke den mikrobielle produksjon av karotenoider. Karotenoidene astaxanthin er distribuert i et vidt mangfold av organismer som for eksempel dyr, alger og mikroorganismer. Det har en kraftig antioksidativ egenskap overfor reaktive oksygenformer. Astaxanthin anvendes som et fargemiddel, særskilt innen industri for produksjon av røkt fisk, som for eksempel laks, fordi astaxanthin gir utpreget oransje-rød farging i dyrene og bidrar til attraktivitet for forbrukeren i markedet.

Den foreliggende oppfinnelse vedrører et gen nyttig i en prosess for å øke den mikrobielle produksjon av karotenoider. Nærmere bestemt vedrører den foreliggende oppfinnelsen en rekombinant organisme, en vektor, en fremgangsmåte for å lage en rekombinant organisme og en fremgangsmåte for å produsere karotenoider ifølge kravene 1-6.

Karotenoidet astaxanthin er distribuert over et vidt mangfold av organismer som for eksempel dyr, alger og mikroorganismer. Det har en kraftig antioksidativ egenskap overfor reaktive oksygenformer. Astaxanthin anvendes som et fargemiddel, særskilt innen oppdrettsfiskindustri, som for eksempel laks, fordi astaxanthin gir utpreget oransje-rød farging i dyrene og bidrar til attraktivitet for forbrukeren i markedet.

Et av trinnene i den karotenogene reaksjonsvei i, for eksempel *Phaffia rhodozyma*, fra en vanlig metabolitt, acetyl-CoA er isomerisering av isopentenylpyrofosfat (IPP) til dimetylarlypyrofosfat (DMAPP) forårsaket av IPP isomerase. Deretter omdannes IPP og DMAPP til en C10-enhet, geranylpyrofosfat (GPP) ved hode-til-hale kondensering.

I en lignende kondenseringsreaksjon mellom GPP og IPP, omdannes GPP til C15-enhet, farnesyl pyrofosfat (FPP) hvilket er et viktig substrat for kolesterol i dyr og ergosterol i gjær, og for farnesyling av reguleringsprotein som for eksempel RAS protein. Generelt katalyseres biosyntesen av GPP og FPP fra IPP og DMAPP av et enzym kalt FPP syntase. På den annen side, i prokaryoter som for eksempel eubakterier ble isopentenylpyrofosfat fremstilt i en forskjellig reaksjonsvei via 1-deoksyxylulose-5-fosfat fra pyruvat hvilket er fraværende i gjær og dyr. Flesteparten av genene involvert i mevalonatreaksjonsveien og FPP syntasegen ble klonet fra *P. rhodozyma* (EP 955,363).

I et aspekt bringer den foreliggende oppfinnelse tilveie et hittil ukjent DNA fragment omfattende et gen som koder for enzymet squalenesyntase.

Mer spesielt bringer den foreliggende oppfinnelse tilveie et DNA inneholdende regulatoriske regioner, som for eksempel promoter og terminator, så vel som den åpne leseramme av squalenesyntasegen.

Den foreliggende oppfinnelse tilveiebringer et DNA fragment som koder for squalenesyntase i *Phaffia rhodozyma*. Det nevnte DNA betyr et cDNA hvilket

inneholder kun åpen leseramme omgitt av de korte fragmenter i dets 5'- og 3'-ikketranslaterte region, og et genomisk DNA hvilket også inneholder dets regulatoriske sekvenser som for eksempel dets promoter og terminator hvilket er nødvendig for ekspresjon av squalenesyntasegenet i *P. rhodozyma*.

5 Som det følger av dette vedrører den foreliggende oppfinnelse en rekombinant organisme hvis genekspresjon av squalene-syntase er redusert sammenlignet med vertsorganismen, som derved er i stand til å fremstille karotenoider i et økt nivå i forhold til vertsorganismen, som er kjennetegnet ved at den rekombinante organismen er en soppcelle som inneholder et antisenspolynukleotid mot et  
10 nukleinsyremolekyl valgt fra gruppen bestående av

(a) nukleinsyremolekyler som koder minst den modne formen av polypeptidet vist i SEKV. ID. NR.: 3;

(b) nukleinsyremolekyler omfattende kodesekvensen som vist i SEKV. ID. NR.: 2;

(c) nukleinsyremolekyler hvis nukleotidsekvens er degenerert som et resultat av  
15 den genetiske koden til en nukleotidsekvens av (a) eller (b);

(e) nukleinsyremolekyler som koder et polypeptid avledet fra polypeptidet hvis sekvens har en identitet på 51.3 % eller mer til aminosyresekvensen av polypeptidet kodet av et nukleinsyremolekyl av (a) eller (b);

(g) nukleinsyremolekyler omfattende et polynukleotid som har en sekvens av et  
20 nukleinsyremolekyl amplifisert fra et *Phaffia* nukleinsyrebibliotek som anvender primerene vist i SEKV. ID. NR.: 4, 5 og 6;

(j) nukleinsyremolekyler som koder et polypeptid som har squalenesyntaseaktivitet, hvor nevnte polypeptid er gjenkjent av antistoffer som har blitt dyrket mot et polypeptid kodet av et nukleinsyremolekyl ifølge ethvert av (a), (b), (c) og (g);

25 (k) nukleinsyremolekyler oppnåelige ved screening av et passende bibliotek under stringente betingelser med en sonde som har sekvensen av nukleinsyremolekylet ifølge ethvert av (a), (b), (c), (e), (g) og (j), og som koder et polypeptid som har squalenesyntaseaktivitet.

Den foreliggende oppfinnelse vedrører også en rekombinant organisme hvis  
genekspresjon av squalene-syntase er redusert sammenlignet med  
vertsorganismen, som derved er i stand til å fremstille karotenoider i et økt nivå i  
forhold til vertsorganismen, kjennetegnet ved at den rekombinante organismen er  
5 en soppcelle som inneholder et antisenspolynukleotid mot et nukleinsyremolekyl  
valgt fra gruppen bestående av:

(m) nukleinsyremolekyler omfattende nukleotidsekvensen som vist i SEKV. ID. NR.:  
1;

(n) nukleinsyremolekyler hvis nukleotidsekvens er  
10 degenerert som et resultat av den genetiske koden til en nukleotidsekvens av (m);

(p) nukleinsyremolekyler som koder et polypeptid avledet fra polypeptidet hvis  
sekvens har en identitet på 51.3 % eller mer til aminosyresekvensen av  
polypeptidet kodet av et nukleinsyremolekyl av (m);

(q) nukleinsyremolekyler omfattende et fragment kodet av et nukleinsyremolekyl  
15 ifølge ethvert av (m), (n) eller (p) og som har squalenesyntaseaktivitet;

(r) nukleinsyremolekyler omfattende et polynukleotid som har en sekvens av et  
nukleinsyremolekyl amplifisert fra et Phaffia nukleinsyrebibliotek som anvender  
primerene vist i SEKV. ID. NR.: 4, 5 og 6;

(s) nukleinsyremolekyler som koder et polypeptid som har  
20 squalenesyntaseaktivitet, hvor nevnte polypeptid er et fragment av et polypeptid  
kodet ifølge ethvert av (m), (n), (p), (q) og (r);

(t) nukleinsyremolekyler omfattende minst 15 nukleotider av et polynukleotid av  
(m) eller (a);

(u) nukleinsyremolekyler som koder et polypeptid som har  
25 squalenesyntaseaktivitet, hvor nevnte polypeptid er gjenkjent av antistoffer som  
har blitt dyrket mot et polypeptid kodet av et nukleinsyremolekyl ifølge ethvert av  
(m), (n), (p), (q), (r) og (s);

(v) nukleinsyremolekyler oppnåelige ved screening av et passende bibliotek under stringente betingelser med en sonde som har sekvensen av nukleinsyremolekylet ifølge ethvert av (m), (n), (p), (q), (r), (s), (t) og (u), og som koder et polypeptid som har squalenesyntaseaktivitet;

- 5 (w) nukleinsyremolekyler hvis komplementære streng hybridiserer under stringente betingelser med et nukleinsyremolekyl ifølge ethvert av (m), (n), (p), (q), (r), (s), (t), (u), (v), og som koder et polypeptid som har squalenesyntaseaktivitet.

Videre vedrører den foreliggende oppfinnelse en vektor ifølge en rekombinant som nevnt ovenfor inneholdende en rekombinant vektor som omfatter

- 10 antisenspoly-nukleotidet, hvor antisenspolynukleotidet er operativt linket til ekspresjonskontrollsekvenser som tillater ekspresjon i soppceller.

Den foreliggende oppfinnelse vedrører også en fremgangsmåte for å lage en rekombinant organisme omfattende innsetting av vektoren nevnt ovenfor i en vertsorganisme, hvor nevnte vertsorganisme hører til en stamme av *Phaffia*

- 15 *rhodozyma* eller *Xanthophylomyces dendrorhous*, samt en fremgangsmåte for å produsere karotenoider, som omfatter kultivering av den rekombinante organismen nevnt ovenfor, hvor nevnte karotenoider er valgt fra astaxanthin,  $\beta$ -caroten, lycopen, zeaxanthin, cantaxanthin.

Benevnelsene "gen(er)", "polynukleotid", "nukleinsyrese-kvens",

- 20 "nukleotidsekvens", "DNA-sekvens" eller "nukleinsyremolekyl(er)" refererer til en polymer form av nukleotider av hvilken som helst lengde, enten ribonukleotider eller deoksyribonukleotider. Denne benevnelsen refererer kun til molekylets primærstruktur.

Således innbefatter denne benevnelse dobbelt- og enkelt-trådet DNA og RNA. Det

- 25 innbefatter også kjente typer modifikasjoner, for eksempel metylering, "caps" substitusjon av en eller flere av de naturlig forekommende nukleotider med en analog. Fortrinnsvis omfatter DNA sekvensen en kodingssekvens som koder for det ovenfor definerte polypeptid.

En "kodende sekvens" er en nukleotidsekvens hvilket er transkribert til mRNA

- 30 og/eller translateret til et polypeptid når plassert under kontroll av passende regulatoriske sekvenser. Grensene av den kodende sekvens bestemmes ved et translasjonsstartkodon ved 5'-terminus og et translasjonsstoppkodon ved 3'-

terminus. En kodingssekvens kan innbefatte, men begrenses ikke til mRNA, cDNA, rekombinante nukleotidsekvenser eller genomisk DNA, mens introner også kan være tilstede ved enkelte forhold. SEQ ID: 1 viser det genomiske DNA i hvilket intronsekvensen settes inn i den kodende sekvens for squalenesyntasegen fra *Phaffia rhodozyma*.

Generelt omfatter genet flere deler hvilket har forskjellige funksjoner. I eukaryoter transkriberes gener som koder for korresponderende protein til prematur messenger RNA (pre-mRNA) forskjellig fra genene for ribosomalt RNA (rRNA), lite nukleært RNA (snRNA) og transfer RNA (tRNA). Selv om RNA polymerase II (PolII) spiller en sentral rolle i denne transkripsjonshendelse kan PolII i seg selv ikke starte transkripsjon uten cis-element dekkende en oppstrøms region inneholdende en promoter og en oppstrøms aktiveringssekvens (UAS) og en trans-virkende proteinfaktor. Først gjenkjenner et transkripsjons-initieringskompleks hvilket omfatter flere basisproteinkomponenter promotersekvensen i den 5'-tilgrensende region av genet som skal uttrykkes. I denne hendelsen er enkelte ytterligere deltakere nødvendig i tilfelle av gen hvilket uttrykkes under enkelte spesifikke reguleringer, som for eksempel en varmesjokk respons, eller adaptasjon til næringsmangel, og så videre. I et slikt tilfelle er det nødvendig at en UAS finnes i den 5'-ikke-translaterte oppstrømsregionen rundt promotersekvensen, og enkelte positive eller negative regulatorproteiner gjenkjenner og binder til UAS. Styrken av bindingen av transkripsjonsinitieringskomplekset til promotersekvensen påvirkes av en slik binding av den trans-virkende faktor rundt promotor og dette muliggjør regulering av transkripsjonsaktivitet.

Etter aktivering av et transkripsjonsinitieringskompleks ved fosforylering initierer et transkripsjonsinitieringskompleks transkripsjon fra transkripsjonsstartsetet. Enkelte deler av transkripsjonsinitieringskomplekset frigjøres som et elongeringskompleks fra promotorregionen til 3'-retningen av genet (dette trinn kalles en promoterklaringshendelse) og elongeringskomplekset fortsetter transkripsjonen til det når en termineringssekvens som er lokalisert i den 3'-tilgrensende nedstrømsregion av genet. Pre-mRNA således generert modifiseres i nukleus ved tilsetning av cap-struktur ved capsetet hvilket nesten samsvarer med transkripsjonsstartsetet og ved tilsetning av polyA-stykker ved polyA-signalet hvilket er lokalisert ved den 3'-tilgrensende nedstrømsregion. Deretter fjernes intronstrukturer fra kodingsregion og eksondeler kombineres til resulterende i en åpen leseramme hvis sekvens samsvarer med den primære aminosyresekvens av et korresponderende protein. Denne modifisering i hvilket et modent mRNA genereres er nødvendig for en stabil

genekspresjon. cDNA med generell-le begreper samsvarer med DNA sekvensen hvilket revers-transkriberes fra denne modne mRNA sekvens. Det kan syntetiseres eksperimentelt ved revers transkriptase avledet fra virale arter ved anvendelse av et modent mRNA som et templat.

- 5 For å uttrykke et gen hvilket er avledet fra en eukaryot anvendes ofte en prosedyre i hvilket cDNA kloner inn i en ekspresjonsvektor for *E. coli*. Dette er et resultat av det faktum at spesifisitet av intronstruktur varierer mellom organismene og en manglende evne til å gjenkjenne intronsekvensen fra andre arter. Faktisk har en prokaryot ingen intronstruktur i sin egen genetiske bakgrunn. Selv i gjær er den
- 10 genetiske bakgrunn forskjellig mellom *Ascomycetes*, som *Saccharomyces cerevisiae* tilhører, og *Basidiomycetes*, som *P. rhodozyma* tilhører, for eksempel kan ikke intronstrukturen av *actingen* fra *P. rhodozyma* gjenkjennes eller spleises av den *ascomycetøse* gjær, *Saccharomyces cerevisiae*. Intron-strukturen av enkelte typer av genene virker å være involvert i reguleringen av ekspresjonen av deres
- 15 respektive gen. Det kan være viktig å anvende et genomisk fragment hvilket har sine introner i tilfelle av selvkloning av genet av interesse hvis intronstruktur involverer en slik regulering av sin egen genekspresjon.

For å applisere en genteknologisk fremgangsmåte for en stammeforbedringsstudie er det nødvendig å studere den genetiske mekanisme i hendelser som for eksempel

20 transkripsjon og translasjon. Det er viktig å bestemme en genetisk sekvens som for eksempel dens UAS, promoter, intronstruktur og terminator for å studere den genetiske mekanisme.

I henhold til det som beskrives her ble genet som koder for squalenesyntase (SQS) genet fra *P. rhodozyma* inkluderte dets 5'- og 3'-tilgrensende regioner så vel som

25 dets intronstruktur bestemt.

Oppfinnelsen omfatter videre polynukleotider som er forskjellig fra en av nukleotidsekvensene vist i SEKV.ID.NR.: 2 (og deler derav) på grunn av degenerering av den genetiske kode og således koder for en squalene-syntase som den kodet for av nukleotidsekvenser vist i SEKV.ID.NR.: 2. Videre har

30 polynukleotidet en nukleotid-sekvens som koder for et protein som har en aminosyresekvens som vist i SEKV.ID.NR.: 3. I en fortsatt videre utførelsesform koder polynukleotidet for et fullengde *Phaffia rhodozyma* protein hvilket hovedsakelig er homolog med en aminosyresekvens i henhold til SEKV.ID.NR.: 3.

I tillegg vil det verdsettes av de med kunnskaper i faget at DNA-sekvens polymorfisme som fører til endringer i aminosyresekvensens kan forekomme innen en populasjon (for eksempel *P. rhodozyma* populasjonen). Slik genetisk polymorfisme i squalenesyntasegenet kan forekomme mellom individer innen en  
5 populasjon på grunn av naturlig variasjon.

Benevnelsene "gen" og "rekombinant gen" refererer til nukleinsyremolekyler omfattende en åpen leseramme som koder for en squalenesyntase, fortrinnsvis en squalene-syntase fra *P. rhodozyma*.

Slike naturlige variasjoner kan typisk resultere i 1-5 % varians i  
10 nukleotidsekvensen av squalenesyntasegenet. Hvilke som helst og alle slike nukleotidvariasjoner og resulterende aminosyrepolymerfisme i squalenesyntase som er resultat av naturlig variasjon og som ikke endrer den funksjonelle aktivitet av squalenesyntase er ment å være innen omfanget av oppfinnelsen.

Polynukleotider korresponderende til naturlige varianter og ikke-*P. rhodozyma*  
15 homologer av squalenesyntase cDNAet kan isoleres på basis av deres homologi med *P. rhodozyma* squalenesyntase polynukleotider brakt for dagen her ved anvendelse av polynukleotidet , eller en del derav, som en hybridiseringsprobe i henhold til standard hybridiseringsteknikker under stringente hybridiseringsbetingelser. Som det følger av dette er, i en annen utførelsesform, et  
20 polynukleotid i det minste 15 nukleotider langt.

Fortrinnsvis hybridiseres det under stringente betingelser med nukleinsyremolekylet omfattende en nukleotidsekvens i henhold til polynukleotidet, for eksempel SEKV. ID. NR.: 2. I andre utførelsesformer er nukleinsyren i det minste 20, 30, 50, 100, 250 eller flere nukleotider langt. Benevnelsen "hybridiserer under stringente  
25 betingelser" er definert ovenfor og er ment å beskrive betingelser for hybridisering og vasking under hvilke nukleotidsekvenser i det minste 60 % identisk med hverandre typisk forblir hybridisert med hverandre. Fortrinnsvis er tilstandene slik at sekvenser i det minste omtrent 65 % eller 70 %, mer fortrinnsvis i det minste omtrent 75 % eller 80 %, og enda mer fortrinnsvis i det minste omtrent 85 %, 90  
30 % eller 95 % eller mer identisk med hverandre typisk forblir hybridisert med hverandre. Fortrinnsvis samsvarer polynukleotidet som hybridiserer under stringente betingelser med en sekvens i henhold til SEKV. ID. NR.: 2 med et naturlig forekommende nukleinsyremolekyl.

Polynukleotidsekvensen omfatter SEKV. ID. NR.: 2 og fragmenter derav som har polynukleotidsekvenser hvilket hybridiserer til SEKV. ID. NR.: 2 under stringente betingelser hvilket er tilstrekkelig for å identifisere spesifikk binding til SEKV. ID. NR.: 2. For eksempel kan enhver kombinasjon av de følgende hybridiserings- og vaskebetingelser brukes for å oppnå den nødvendige spesifikke binding:

Høystringent hybridisering: 6X SSC; 0,5 % SDS, 100 µg/ml denaturert laksesperma DNA, 50 % formamid, inkuber natten over med forsiktig vugging ved 42 °C.

Høystringent vask: 1 vask i 2X SSC, 0,5 % SDS ved romtemperatur i 15 minutter, etterfulgt av en andre vask i 0,1 X SSC, 0,5 % SDS ved romtemperatur i 15 minutter.

Lavstringent hybridisering: 6X SSC, 0,5 % SDS, 100 µg/ml denaturert laksesperma DNA, 50 % formamid, inkuber natten over med forsiktig vugging ved 37 °C.

Lavstringent vask: 1 vask i 0,1X SSC, 0,5 % SDS ved rom-temperatur i 15 minutter.

Moderate stringente betingelser kan oppnås ved å variere temperaturen ved hvilket hybridiseringsreaksjonen foregår og/eller vaskebetingelsene som fremlagt ovenfor. Det foretrekkes her å anvende høystringente hybridiserings og vaskebetingelser for å definere antisensaktiviteten overfor squalenesyntasegen fra *P. rhodozyma*.

Benevnelsen "homologi" betyr at de respektive nukleinsyremolekyler eller kodede proteiner er funksjonelt og/eller strukturelt ekvivalent. Nukleinsyremolekylene som er homologe med nukleinsyremolekylene beskrevet ovenfor og som er derivater av nevnte nukleinsyremolekyler er, for eksempel variasjoner av nevnte nukleinsyremolekyler hvilket representerer modifikasjoner som har den samme biologiske funksjon, i særdeleshet som koder for proteiner med den samme eller hovedsakelig den samme biologiske funksjon. De kan være naturlig forekommende variasjoner, som for eksempel sekvenser fra andre plantevarieteter eller arter, eller mutasjoner.

Disse mutasjoner kan forekomme naturlig eller kan oppnås ved mutageneseteknikker. De alleliske variasjoner kan være naturlig forekommende alleliske varianter så vel som syntetisk produserte eller genteknologisk fremstilte varianter. Strukturelle ekvivalenter kan, for eksempel identifiseres ved å teste

bindingen av nevnte polypeptid til antistoff. Strukturelle ekvivalenter har lignende immunologiske karakteristikk, for eksempel omfatter lignende epitoper.

Et "naturlig forekommende" nukleinsyremolekyl refererer til et RNA eller DNA molekyl som har en nukleotidsekvens som forekommer i naturen (for eksempel  
5 som koder for et naturlig protein). Fortrinnsvis koder polynukleotidet for en naturlig P. rhodozyma squalenesyntase.

I tillegg til naturlig forekommende varianter av squalenesyntasesekvensen som kan eksistere i populasjonen vil den med fagkunnskaper videre forstå at endringer kan introduseres ved mutasjon i en nukleotidsekvens av polynukleotidet som koder for  
10 squalenesyntase, som derved fører til endringer i aminosyresekvensen av den kodede squalenesyntase, uten å endre den funksjonelle evnen av squalenesyntasen. For eksempel kan nukleotidsubstitusjoner som fører til aminosyresubstitusjoner ved "ikke-essensielle" amino-syrerester lages i en sekvens av polynukleotidet som koder for squalenesyntase, for eksempel SEKV.ID.NR.: 2.  
15 En "ikke-essensiell" aminosyrerest er en rest som kan være endret fra villtypesekvensen av en av squalenesyntasene uten å endre aktiviteten av nevnte squalenesyntase, mens en "essensiell" aminosyrerest er nødvendig for squalenesyntaseaktivitet. Dog behøver andre aminosyrerester (for eksempel de som ikke er konserverte eller kun semi-konserverte i domenet som har  
20 squalenesyntaseaktivitet) ikke være essensielle for aktivitet og er således sannsynligvis mottagelig for endring uten at det endrer squalenesyntaseaktivitet.

Som det følger av dette beskrives her polynukleotider som koder for squalenesyntase som inneholder endringer i aminosyrerester som ikke er essensielle for squalenesyntaseaktivitet. Slik squalenesyntase er forskjellig i  
25 aminosyresekvens fra en sekvens innbefattet i SEKV.ID.NR.: 3, men har fremdeles holdt tilbake squalenesyntaseaktiviteten beskrevet her. Polynukleotidet kan omfatte en nukleotidsekvens som koder for et polypeptid, karakterisert ved at polypeptidet omfatter en aminosyresekvens i det minste omtrent 60 % identisk med en aminosyresekvens i henhold til SEKV.ID.NR.: 3 og som er i stand til å delta i  
30 syntese av squalene. Fortrinnsvis er proteinet kodet for av nukleinsyremolekylet i det minste omtrent 60-65% identisk med sekvensen i SEKV.ID.NR.: 3, mer fortrinnsvis i det minste omtrent 60-70% identisk med en av sekvensene i SEKV.ID.NR.: 3, enda mer fortrinnsvis i det minste omtrent 70-80%, 80-90%, 90-95% homolog med sekvensen i SEKV.ID.NR.: 3, og mest foretrukket i det minste  
35 omtrent 96%, 97%, 98% eller 99% identisk med sekvensen i SEKV.ID.NR.: 3.

For å bestemme prosent homologi av to aminosyresekvenser (for eksempel en av sekvensens i henhold til SEKV.ID.NR.: 3 og en mutant form derav) eller av to nukleinsyrer, linjestilles sekvensene for optimale sammenligningsformål (for eksempel kan hull introduseres i sekvensen av et protein eller en nukleinsyre for optimal linjestilling med det andre protein eller den andre nukleinsyre).

5 Aminosyrerestene eller nukleotidene ved korresponderende aminosyreposisjoner eller nukleotidposisjoner sammenlignes deretter. Når en posisjon i en sekvens (for eksempel en av sekvensene i henhold til SEKV.ID.NR.: 2 eller 3) innehar den samme aminosyrerest eller nukleotid som den korresponderende posisjon i den andre sekvens (for eksempel en mutant form av sekvensen valgt), så er

10 molekylene homologe ved den posisjonen (dvs. at aminosyre eller nukleinsyre "homologi" er ekvivalent med aminosyre eller nukleinsyre "identitet"). Prosent homologi mellom de to sekvensene er en funksjon av antallet av identiske posisjoner delt av sekvensene (dvs. % homologi = antall av identiske

15 posisjoner/totalt antall av posisjoner x 100). Homologien kan bestemmes ved dataprogrammer som Blast 2,0 (Altschul SF, Nuc. Acid. Res., 25, 3389-3402, 1997). I denne oppfinnelse anvendes GENETYX-SV/RC software (Software Development Co., Ltd., Tokyo, Japan) ved å anvende dets standard algoritme som slik homologianalyse programvare. Denne programvare anvender Lipman-Pearson

20 fremgangsmåten i sin analytiske algoritme.

Et nukleinsyremolekyl som koder for en squalenesyntase homolog med en proteinsekvens i henhold til SEKV.ID.NR.: 3 kan fremstilles ved å introdusere en eller flere nukleotidsubstitusjoner, addisjoner eller delesjoner inn i en nukleotidsekvens av polynukleotidet, i særdeleshet av SEKV.ID.NR.: 2 slik at en

25 eller flere aminosyresubstitusjoner, addisjoner eller delesjoner introduseres inn i det kodede protein. Mutasjoner kan introduseres inn i sekvensene i henhold til, for eksempel SEKV.ID.NR.: 2 ved vanlige teknikker, som for eksempel setedirigert mutagenese og PCR-mediert mutagenese. Fortrinnsvis lages konservative aminosyresubstitusjoner ved en eller flere forutsagte ikke-essensielle

30 aminosyrerester. En "konservativ aminosyresubstitusjon" er en i hvilket aminosyreresten er erstattet med en aminosyrerest som har en lignende sidekjede. Familier av aminosyrerester som har lignende sidekjedder er definert innen faget. Disse familier innbefatter aminosyrer med basiske sidekjedder (for eksempel lysin, arginin, histidin), syresidekjedder (for eksempel asparginsyre, glutaminsyre),

35 uladede polare sidekjedder (for eksempel glysin, asparagin, glutamin, serin, treonin, tyrosin, cystein), upolare sidekjedder (for eksempel alanin, valin, leucin, isoleucin, prolin, fenylalanin, metionin, tryptofan), betaforgrenede sidekjedder (for eksempel

treonin, valin, isoleucin) og aromatiske sidekjeder (for eksempel tyrosin, fenyalanin, tryptofan, histidin). Således erstattes en forutsagt ikke-essensiell aminosyrerest i en squalene-syntase fortrinnsvis med en annen aminosyresyrerest fra den samme familie. Alternativt kan mutasjoner, i en annen utførelsesform, introduseres tilfeldig langs hele eller deler av en squalenesyntase kodende sekvens, som for eksempel ved metningsmutagenese og de resulterende mutanter kan screenes for en squalenesyntaseaktivitet beskrevet her for å identifisere mutanter som har bibeholdt squalenesyntaseaktivitet. Etterfølgende mutagenese av en av sekvensene i henhold til SEKV.ID.NR.: 2, kan det kodede protein uttrykkes rekombinant og aktiviteten av proteinet kan bestemmes ved anvendelse av for eksempel assayer beskrevet her.

Som det følger av dette er, i en foretrukket utførelses-form, polynukleotidet DNA eller RNA.

Et polynukleotid, for eksempel et nukleinsyremolekyl som har en nukleotidsekvens i henhold til SEKV.ID.NR.: 2, eller en del derav, kan isoleres ved anvendelse av standard molekylærbiologiske teknikker og sekvensinformasjonen brakt tilveie her. For eksempel kan squalenesyntase cDNA isoleres fra et bibliotek ved anvendelse av hele eller deler av en av sekvensene av polynukleotidet som hybridiseringsprobe og standard hybridiseringstekniker. Enn videre kan et polynukleotid omfattende hele eller en del av en av sekvensene av polynukleotidet isoleres ved polymerase kjedereaksjon ved anvendelse av oligonukleotidprimere designet på basis av denne sekvens (for eksempel et nukleinsyremolekyl omfattende hele eller en del av en av sekvensene av polynukleotid kan isoleres ved polymerase kjedereaksjon ved anvendelse av oligonukleotidprimere, for eksempel i henhold til SEKV.ID.NR.: 4, 5 eller 6, designet på basis av den samme sekvens av polynukleotidet. For eksempel kan mRNA isoleres fra celler, for eksempel *Phaffia* (for eksempel ved guanidiniumtiocyanat ekstraksjonsprosedyren av Chirgwin et al., og cDNA kan fremstilles ved anvendelse av revers transkriptase (for eksempel Moloney MLV revers transkriptase eller AMV revers transkriptase tilgjengelig fra Promega (Madison, USA)). Syntetiske oligonukleotidprimere for polymerase kjedereaksjon amplifisering kan designes på basis av en av nukleotidsekvensene vist i SEKV.ID.NR.: 2. Et polynukleotid kan amplifiseres ved anvendelse av cDNA eller, alternativt, genomisk DNA, som templat og passende oligonukleotidprimere i henhold til standard PCR-amplifiseringsteknikker. Polynukleotidet slik amplifisert kan klones inn i en passende vektor og karakteriseres ved DNA-sekvensanalyse. Videre kan oligonukleotider korresponderende til en squalenesyntase nukleotidsekvens

fremstilles ved standard synteseteknikker, for eksempel ved anvendelse av en automatisert DNA-syntetisator.

Benevnelsene "fragment", "fragment av en sekvens" eller "del av en sekvens" betyr en trunkert sekvens av den opprinnelige sekvens referert til. Den trunkerte sekvens  
 5 (nukleinsyre eller proteinsekvens) kan variere mye i lengde; minimumsstørrelsen er en sekvens som er tilstrekkelig stor til å bringe tilveie en sekvens med i det minste en sammenlignbar funksjon og/eller aktivitet referert til den opprinnelige sekvens, mens maksimal størrelse ikke er kritisk. I enkelte applikasjoner er maksimal størrelse vanligvis ikke hovedsakelig større enn nødvendig for å bringe tilveie den  
 10 ønskede aktivitet og/eller funksjon(er) av den opprinnelige sekvens.

Typisk vil den trunkerte aminosyresekvens være i området fra omtrent 5 til omtrent 60 aminosyrer lang. Mer typisk vil dog sekvensen være maksimalt omtrent 50 aminosyrer lang, fortrinnsvis maksimalt omtrent 30 aminosyrer. Det er vanligvis ønskelig å velge ut sekvenser på i det minste omtrent 10, 12 eller 15 aminosyrer,  
 15 opp til et maksimum av omtrent 20 eller 25 aminosyrer.

Benevnelsen "epitop" vedrører spesifikke immunreaktive seter innen et antigen, også kjent som antigene determi-nanter. Disse epitoper kan være en lineær oppstilling av monomerer i en polymerisk sammensetning - som for eksempel aminosyrer i et protein - eller bestå av eller omfatte en mer kompleks sekundær  
 20 eller tertiær struktur. De med fagkunnskaper vil innse at alle immunogener (dvs. substanser i stand til å frembringe en immunrespons) er antigener; dog er enkelte antigener, som for eksempel haptener, ikke immunogener, men kan gjøres immunogene ved kobling til et bærermateriale-molekyl. Benevnelsen "antigen" innbefatter referanser til en forbindelse til hvilket et antistoff kan frembringes  
 25 og/eller til hvilket antistoffet er spesielt immunoreaktivt.

Benevnelsen "en eller flere aminosyrer" vedrører i det minste en aminosyre, men ikke flere enn det antall aminosyrer hvilket ville resultere i en homologt under 60 % identitet. Fortrinnsvis er identiteten mer enn 70 % eller 80 %, mer foretrukket er 85%, 90% eller 95%, enda mer foretrukket er 96%, 97%, 98% eller 99% identitet.

30 Benevnelsen "squalenesyntase" eller "squalenesyntaseaktivitet" vedrører enzymatiske aktiviteter av et polypeptid som beskrevet nedenfor eller hvilket kan bestemmes i enzymassaymetoder. Videre omfattes også polypeptider som er inaktive i en assay her, men som gjenkjennes av et antistoff som spesifikt binder til

squalenesyntase, dvs. som har en eller flere squalenesyntase epitoper, under benevnelsen "squalenesyntase". I disse tilfeller refererer aktivitet til deres immunologiske aktivitet.

Benevnelsene "polynukleotid" og "nukleinsyremolekyl" relaterer også til "isolerte" polynukleotider eller nukleinsyremolekyler. Et "isolert" nukleinsyremolekyl er et 5 hvilket separeres fra andre nukleinsyremolekyler hvilket er tilstede i den naturlige kilde for nukleinsyren. Fortrinnsvis er en "isolert" nukleinsyre fri for sekvenser hvilket naturlig omgir nukleinsyren (dvs. sekvenser lokalisert ved 5'- og 3'-endene av nukleinsyren) i det genomiske DNA av organismen fra hvilket nukleinsyren 10 avledes.

For eksempel kan, i forskjellige utførelsesformer, PNO polynukleotid inneholde mindre enn omtrent 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb eller 0,1 kb av nukleotidsekvenser som naturlig omgir nukleinsyremolekylet i genomisk DNA av cellen fra hvilket nukleinsyren avledes (for eksempel en *Phaffia* celle). Enn videre 15 kan polynukleotidene, i særdeleshet et "isolert" nukleinsyremolekyl, som for eksempel et cDNA molekyl, i hovedsak være fritt for annet celle-materiale, eller kulturmedium når produsert ved rekombinante teknikker, eller kjemiske forløpere eller andre kjemikalier når kjemisk syntetisert.

Fortrinnsvis omfatter polypeptidet en av nukleotidsekvensene vist i SEKV.ID.NR.: 2. 20 Sekvensen i henhold til SEKV.ID.NR.: 2 samsvarer med *P. rhodozyma* squalenesyntase cDNAene .

Videre omfatter polynukleotidet et nukleinsyremolekyl hvilket er et komplement av en av nukleotidsekvensene av ovenfor nevnte polynukleotider eller en del derav. Et nukleinsyremolekyl hvilket er komplementært med en av nukleotidsekvensene vist i 25 SEKV.ID.NR.: 2 er et hvilket er tilstrekkelig komplementært med en av nukleotidsekvensene vist i SEKV.ID.NR.: 2 slik at det kan hybridisere med en av nukleotidsekvensene vist i SEQ ID N0:2, derved dannende et stabilt dupleks.

Polynukleotidet omfatter en nukleotidsekvens hvilket i det minste er omtrent 60%, fortrinnsvis i det minste om-trent 65-70%, mer fortrinnsvis i det minste omtrent 30 70-80%, 80-90% eller 90-95%, og ennå mer fortrinnsvis i det minste omtrent 95%, 96%, 97%, 98%, 99% eller mer homolog med en nukleotidsekvens vist i SEQ ID N0:2, eller en del derav. Polynukleotidet omfatter en nukleotidsekvens hvilket

hybridiserer, for eksempel hybridiserer under stringente betingelser som definert her, til en av nukleotidsekvensene vist i SEKV.ID.NR.: 2, eller en del derav.

Enn videre kan polynukleotidet omfatte kun en del av den kodende region av en av sekvensene i SEKV.ID.NR.: 2, for eksempel et fragment hvilket kan anvendes som

5 en probe eller primer eller et fragment som koder for en biologisk aktiv del av en squalenesyntase. Nukleotidsekvensene bestemt fra kloningen av squalenesyntasegenet fra *P. rhodozyma* gir anledning til dannelse av prober og primere designet for anvendelse ved identifisering og/eller kloning av squalenesyntasehomologer i andre celletyper og organismer. Proben/primeren

10 omfatter typisk hovedsakelig rensset oligonukleotid. Oligonukleotidet omfatter typisk en region av nukleotidsekvens som hybridiserer under stringente betingelser til i det minste omtrent 12, 15 fortrinnsvis omtrent 20 eller 25, mer fortrinnsvis omtrent 40, 50 eller 75 konsekutive nukleotider av en sense-tråd av en av sekvensens fremlagt, for eksempel i SEKV.ID.NR.: 2, en anti-sens-sekvens av en

15 av sekvensene, for eksempel fremlagt i SEKV.ID.NR.:2, eller naturlig forekommende mutanter derav. Primere basert på et nukleotid kan anvendes i PCR reaksjoner for å kloner squalenesyntasehomologer. Prober basert på squalenesyntasenukleotidsekvenser kan anvendes til å påvise transkripter eller genomiske sekvenser som koder for de samme eller homologe proteiner. Proben

20 kan videre omfatte en merkelappgruppe tilføyd dertil, for eksempel kan merkelappgruppen være en radioisotop, en fluorescerende forbindelse, et enzym, eller en enzym-ko-faktor. Slike prober kan anvendes som del av et genomisk markør testkit for å identifisere celler som uttrykker en squalenesyntase, som for eksempel ved å måle nivå av et squalenesyntetasekodende nukleinsyremolekyl i en

25 prøve av celler, for eksempel å påvise squalenesyntase mRNA nivåer eller å bestemme hvorvidt et genomisk squalenesyntasegen er mutert eller tatt bort.

Polynukleotidet koder for et polypeptid eller del derav hvilket innbefatter en aminosyresekvens hvilket er til-strekkelig homolog med en aminosyresekvens i henhold til SEKV.ID.NR.: 3 slik at proteinet eller del derav bibeholder evnen til å

30 delta i syntese av squalene, i særdeleshet en squalenesyntaseaktivitet som beskrevet i eksemplene i mikroorganismer eller planter. Språk-uttrykket "tilstrekkelig homolog" refererer til proteiner eller deler derav hvilket har aminosyresekvenser hvilket innbefatter et minimum antall identiske eller ekvivalente (for eksempel en aminosyrerest hvilket har en lignende sidekjede som

35 en aminosyrerest i en av sekvensene av polypeptidet, aminosyrerester til en aminosyresekvens i henhold til SEKV.ID.NR.: 3 slik at proteinet eller del derav er i

stand til å delta i syntese av squalene i mikroorgan-ismer eller planter. Eksempler av squalenesyntaseaktivitet beskrives også her.

Proteinet er i det minste omtrent 60-65%, fortrinnsvis i det minste omtrent 66-70%, og mer fortrinnsvis i det minste omtrent 70-80%, 80-90%, 90-95%, og mest  
5 foretruk-*ket* i det minste omtrent 96%, 97%, 98%, 99% eller mer homolog med en fullstendig aminosyresekvens i henhold til SEKV.ID.NR.: 3.

Deler av proteiner kodet for av squalenesyntasepolynukleotidet er fortrinnsvis biologisk aktive deler av en av squalenesyntasene.

Som nevnt her menes benevnelsen "biologisk aktiv del av squalenesyntase" å  
10 innbefatte en del, for eksempel et domene/motiv, som deltar i biosyntese av squalene eller har en immunologisk aktivitet slik at det binder til et antistoff som binder spesifikt til squalenesyntase. For å bestemme hvorvidt en squalenesyntase eller en biologisk aktiv del derav kan delta i metabolismen kan en assay av enzymatisk aktivitet utføres. Slike assaymetoder er vel kjente for de med  
15 kunnskaper i faget, som detaljert i eksemplene. Ytterligere nukleinsyrefragmenter som koder for biologisk aktive deler av en squalenesyntase kan fremstilles ved å isolere en del av en av sekvensene i SEKV.ID.NR.: 2, uttrykke den kodede del av squalenesyntasen eller peptidet (for eksempel ved rekombinant ekspresjon in vitro) og vurdere aktiviteten av den kodede del av squalenesyntasen eller peptidet.

20 Først ble et delvis genfragment inneholdende en del av SQS gen ved anvendelse av degenerert PCR-metode klonet. Den nevnte degenererte PCR er en fremgangsmåte for å klonet et gen av interesse hvilket har høy homologi av aminosyresekvens med det kjente enzym fra andre arter hvilket har en lik eller lignende funksjon. Degenerert primer, hvilket anvendes som en primer i degenerert PCR, ble designet  
25 ved en revers translasjon av aminosyresekvensen til korresponderende nukleotider ("degenerert"). I en slik degenerert primer anvendes generelt en blandet primer hvilket består av hvilken som helst A, C, G eller T, eller en primer inneholdende inosin ved en ubestemt kode. Her ble slike blandede primere anvendt som degenererte primere for å klonet genet ovenfor.

30 Et fullstendig gen inneholdende dets kodende region med dets introner så vel som dets reguleringsregion som for eksempel en promoter eller en terminator kan klonet fra et kromosom ved screening av genomisk bibliotek hvilket konstrueres i fagvektor eller plasmidvektor i passende vert, ved anvendelse av et delvis DNA

fragment oppnådd ved degenerert PCR som beskrevet ovenfor som en probe etter at den var merket. Generelt anvendes ofte, *E. coli* som en vertsstamme og *E. coli* vektor, en fagvektor som for eksempel  $\lambda$ -fag vektor, eller en plasmidvektor som for eksempel pUC-vektor for konstruksjon av bibliotek og en etterfølgende genetisk  
5 manipulering som for eksempel en sekvensering, en restriksjonsnedbrytning, en ligering og lignende. Her ble et EcoRI genomisk bibliotek av *P. rhodozyma* konstruert i derivatene av  $\lambda$ -vektor,  $\lambda$ DASHII. Innsettets størrelse, hvilken lengde av innsettet som må klones, ble bestemt ved Southern blot hybridisering for genet før konstruksjon av et bibliotek. Her ble et DNA anvendt som en probe merket med  
10 digoxigenin (DIG), et steroid haptent istedenfor konvensjonell <sup>32</sup>P-merkelapp, etterfølgende protokollen som ble forberedt av produsenten (Boehringer-Mannheim, Mannheim, Tyskland). Et genomisk bibliotek konstruert fra kromosomet av *P. rhodozyma* ble screenet ved anvendelse av et DIG-merket DNA fragment hvilket hadde en del av et gen av interesse som probe. Hybridiserte plakk ble plukket og  
15 anvendt for videre studier. I tilfellet med anvendelse av  $\lambda$ DASHII (innsettets størrelse var fra 9 kb til 23 kb), ble preparert  $\lambda$ DNA nedbrutt av EcoRI, etterfulgt av kloningen av EcoRI-innsettet inn i en plasmidvektor som for eksempel pUC19 eller pBluescriptII SK+. Et plasmid DNA således oppnådd ble undersøkt for dets sekvens.

I denne oppfinnelse ble det brukt den automatiserte fluorescerende DNA-  
20 sekvensator, ALFred systemet (Pharmacia, Uppsala, Sverige) ved anvendelse av en autosyklus sekvenseringsprotokoll i hvilket Taq DNA polymerase benyttes i flest sekvenseringstilfeller.

Etter bestemmelse av den genomiske sekvens ble en sekvens av en kodende region anvendt for kloning av cDNA av korresponderende gen. PCR-metoden ble også  
25 utnyttet for å klonen cDNA fragment. PCR-primere hvis sekvenser var identiske med sekvensen ved 5'- og 3'-endene av den åpne leseramme (ORF) ble syntetisert med addisjon av et passende restriksjonssete, og PCR ble utført ved å anvende disse PCR primere. Her ble en cDNA-pool anvendt som templat i denne PCR kloning av cDNA. Den nevnte cDNA-pool omfatter forskjellige cDNA arter hvilket ble  
30 syntetisert in vitro ved den virale revers transkriptase og Taq polymerase (CapFinder Kit fremstilt av Clontech, Palo Alto, U.S.A.) ved å anvende mRNAet oppnådd fra *P. rhodozyma* som templat. cDNA av interesse således oppnådd ble bekreftet i sin sekvens.

I en annen utførelsesform beskrives en metode for å lage en rekombinant vektor  
35 omfattende å sette inn et polynukleotid i en vektor.

Videre vedrører den foreliggende oppfinnelse en rekombinant vektor inneholdende polynukleotidet eller produsert ved nevnte fremgangsmåte.

Benevnelsen "vektor" refererer til et nukleinsyremolekyl i stand til å transportere et polynukleotid til hvilket det er lenket. En type av vektor er et "plasmid", hvilket refererer til en sirkulær dobbeltrådet DNA sirkel inn i hvilket ytterligere DNA segmenter kan liggeres. En annen type vektor er en viral vektor, hvori ytterligere DNA eller RNA segmenter kan liggeres inn i det virale genom. Enkelte vektorer er i stand til autonom replikasjon i en vertscelle inn i hvilket de er introdusert (for eksempel bakterielle vektorer med et bakteriell replikasjonsorigin og episomale mammaliske vektorer). Andre vektorer (for eksempel ikke-episomale mammaliske vektorer) er integrert inn i genomet av en vertscelle ved introduksjon inn i vertscellen, og replikeres derved sammen med vertsgenomet. Enn videre er enkelte vektorer i stand til å dirigere ekspresjonen av gener til hvilke de er operativt lenket. Slike vektorer refereres til her som "ekspresjonsvektorer". Generelt er ofte ekspresjonsvektorer for benyttelse i rekombinante DNA teknikker i form av plasmider. I den foreliggende patentbeskrivelse kan "plasmid" og "vektor" anvendes omvekslende siden plasmidet er den vanligst anvendte formen for vektor. Dog er det her ment å innbefatte andre former for ekspresjonsvektorer, som for eksempel virale vektorer (for eksempel replikasjonsdefekte retrovirus, adenovirus og adeno-assosierte virus), hvilket tjener ekvivalente funksjoner.

Det beskrives her også cosmider, virus, bakteriofager og andre vektorer konvensjonelt anvendt innen genteknologi som inneholder et nukleinsyremolekyl. Fremgangsmåter hvilket er velkjente for de med kunnskaper i faget kan anvendes for å konstruere forskjellige plasmider og vektorer. Alternativt kan nukleinsyremolekylene og vektorene rekonstitueres i liposomer for avlevering til målceller.

En annen foretrukket utførelsesform vedrører en vektor i hvilket polynukleotidet operativt lenkes til ekspresjonskontrollsekvenser hvilket gir adgang til ekspresjon i prokaryote eller eukaryote vertsceller. Karakteren av slike kontrollsekvenser er forskjellig avhengig av vertsorganismen. I prokaryoter innbefatter generelt kontrollsekvenser promotere, ribosomalt bindingssete og terminatorer. I eukaryote innbefatter generelt kontrollsekvenser promotere, terminatorer og, i enkelte tilfeller, forsterkere, transaktivatorer; eller transkripsjonsfaktorer.

Benevnelsen "kontrollsekvens" menes å innbefatte, i minimum, komponenter hvilket nærværet av er nødvendig for ekspresjon, og kan også innbefatte ytterligere fordelaktige komponenter.

5 Benevnelsen "operativt lenket" refererer til en juxtaposisjon karakterisert ved at komponentene på den måte beskrevet er i et forhold som tillater de å fungere på deres tiltenkte måte. En kontrollsekvens "operativt lenket" til en kodingssekvens liggeres på en slik måte at ekspresjon av den kodende sekvens oppnås ved betingelser kompatible med kontrollsekvensene. I tilfelle kontrollsekvensen er en promoter, er det innlysende for en trent person at dobbelttrådet nukleinsyre  
10 anvendes.

Slike regulatoriske sekvenser er kjent for den faglærte. Regulatoriske sekvenser innbefatter de hvilket dirigerer konstitutiv ekspresjon av en nukleotidsekvens i mange typer vertsceller og de hvilket dirigerer ekspresjon av nukleotidsekvensen kun i visse vertsceller eller under særskilte forutsetninger. Det vil forstås av de med  
15 kunnskaper i faget at designet av ekspresjonsvektoren kan avhenge av slike faktorer som valget av vertscelle som skal transformeres, ønsket nivå av ekspresjon av protein, osv. Ekspresjonsvektorene kan introduseres inn i vertsceller for derved å produsere proteiner eller peptider, inklusive fusjonsproteiner eller peptider, kodet for av polynukleotider som beskrevet her.

20 De rekombinante ekspresjonsvektorer kan være designet for ekspresjon av squalenesyntase i prokaryote eller eukaryote celler. For eksempel kan gener som koder for polynukleotidet uttrykkes i bakterielle celler som for eksempel E. coli, insektceller (ved anvendelse av baculovirus ekspresjonsvektorer), gjær og andre soppceller, alger, ciliater av typene: Holotrichia, Peritrichia, Spirotrichia, Suctoria,  
25 Tetrahymena, Paramecium, Colpidium, Glaucoma, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella, og Stylonychia, særskilt Stylonychia lemnae med etterfølgende vektorer, en transformeringsmetode som beskrevet i WO 98/01,572 og multicellulære planteceller eller mammaliske celler. Passende vertsceller er kjent for den faglærte. Alternativt kan den rekombinante  
30 ekspresjonsvektor transkriberes og translateres in vitro, for eksempel ved anvendelse av T7 promoter regulatoriske sekvenser og T7 polymerase.

Ekspresjon av proteiner i prokaryoter er som oftest utført med vektorer inneholdende konstitutive eller induserbare promotere som dirigerer ekspresjon av enten fusjons- eller ikke-fusjonsproteiner. Fusjonsvektorer tilsetter et antall

aminosyrer til et protein kodet for deri, vanligvis til den amino-terminale ende av det rekombinante protein, men også til C-terminus eller fusjonert innen passende regioner i proteinene. Slike fusjonsvektorer tjener typisk tre formål: 1) å øke ekspresjon av rekombinant protein; 2) å øke løseligheten av det rekombinante protein; og 3) å understøtte rensingen av det rekombinante protein ved å virke som en ligand i affinitetsrensing. Ofte introduseres et proteolytisk spaltingssete i fusjonsekspresjonsvektorer ved overgangen mellom fusjonsdelen og det rekombinante protein for å muliggjøre separasjon av det rekombinante protein fra fusjonsdelen påfølgende rensing av fusjonsproteinet. Slike enzymer, og deres kognate gjenkjennelsessekvenser, innbefatter Faktor Xa, trombin og enterokinase.

Typiske fusjonsekspresjonsvektorer innbefatter pGEX (Pharmacia Biotech Inc.), pMAL (New England Biolabs, Beverly, USA) og pRIT5 (Pharmacia, Piscataway, USA) hvilket fusjonerer henholdsvis glutathion S-transferase (GST), maltose E bindende protein, eller protein A, til det målrettede rekombinante protein. I en utførelsesform kloner den kodende sekvens av polypeptidet kodet for polynukleotidet inn i en pGEX ekspresjonsvektor for å konstruere en vektor som koder for et fusjonsprotein omfattende, fra N-terminus til G-terminus, GST-trombin spaltingssete-X protein. Fusjonsproteinet kan renses ved affinitetskromatografi ved anvendelse av glutathion-agarose resin, for eksempel kan rekombinant squalenesyntase ufusjonert med GST gjenvinnes ved spalting av fusjonsproteinet med trombin.

Eksempler av passende induserbare ikke-fusjon E. coli ekspresjonsvektorer innbefatter pTrc og pET 11d. Målgenekspresjon fra pTrc-vektoren er avhengig av vert RNA polymerase transkripsjon fra en hybrid trp-lac fusjonspromoter. Målgenekspresjon fra pET 11d-vektoren er avhengig av transkripsjon fra en T7 gn10-lac fusjonspromoter mediert ved en ko-uttrykt viral RNA polymerase (T7 gnl). Denne virale polymerase stilles til rådighet av vertsstammer BL21(DE3) eller HMS174(DE3) fra en resident  $\lambda$ -profag innehavende et T7gnl-gen under transkripsjonell kontroll av lacUV 5 promoteren.

En strategi for å maksimere rekombinant proteinekspresjon er å uttrykke proteinet i vertsbakterier med forringet kapasitet til å proteolytisk spalte det rekombinante protein. En annen strategi er å endre nukleinsyresekvensen av nukleinsyren som skal settes inn i en ekspresjonsvektor slik at de individuelle kodoner for hver aminosyre er de for-trinnsvis benyttet i bakterien valgt for ekspresjon, som for

eksempel *E. coli*. Slik endring av nukleinsyre-sekvensene kan utføres ved standard DNA synteseteknikker.

Videre kan squalenesyntasevektoren være gjærekspresjons-vektor. Eksempler av vektorer for ekspresjon i gjæren *S. cerevisiae* innbefatter pYepSecI, pMFa, pJRY88  
5 og pYES2 (Invitrogen, San Diego, USA). Vektorer og metoder for konstruksjon av vektorer passende for anvendelse i andre fungi, som for eksempel de filamentøse sopper, er vel kjent for den faglærte.

Alternativt kan polynukleotidet introduseres i insektceller ved anvendelse av baculovirus ekspresjonsvektorer. Baculovirusvektorer tilgjengelig for ekspresjon av  
10 proteiner i dyrkede insektceller (for eksempel Sf9 celler) innbefatter pAc-seriene og pVL-seriene.

Alternativt introduseres polynukleotidet i mammaliske celler ved anvendelse av en mammalisk ekspresjonsvektor. Eksempler av mammaliske ekspresjonsvektorer innbefatter pCDM8 og pMT2PC. Når anvendt i mammaliske celler bringes  
15 ekspresjonsvektorens kontrollfunksjoner ofte tilveie ved virale regulatoriske elementer. For eksempel er vanlig benyttede promotere avledet fra polyoma, Adenovirus 2, cytomegalovirus og Simian Virus 40. Andre passende ekspresjonssystemer for både prokaryote og eukaryote celler er kjent for den faglærte.

20 Den rekombinante mammaliske ekspresjonsvektor kan være i stand til å dirigere ekspresjon av nukleinsyren fortrinnsvis i en spesiell celletype (for eksempel anvendes vevsspesifikke regulatoriske elementer for å uttrykke nukleinsyren). Vevsspesifikke regulatoriske elementer er kjent innen faget. Ikke-begrensede eksempler av passende vevsspesifikke promotere innbefatter albuminpromoteren  
25 (leverspesifikk), lymfoidspesifikke promotere, i særdeleshet promotere av T-cellerreseptorer og immunoglobuliner, neuronspesifikke promotere (for eksempel neurofilamentpromoteren), pankreasspesifikke promotere, og brystkjertelspesifikke promotere (for eksempel melkemysepromoter; US 4,873,316 og EP 264,166). Utviklingsregulerte promotere er også omfattet, for eksempel de murine hox-  
30 promotere og fetoprotein promoteren.

Således uttrykt SQS gen kan verifiseres for sin aktivitet for eksempel ved enzymassaymetode. Enkelte eksperimentelle protokoller beskrives i litteraturen: Det følgende er den ene av fremgangsmåtene som anvendes for bestemmelse av

squalenesyntaseaktivitet: squalenesyntaseaktiviteter bestemmes ved å følge omdanningen av [1-3H] FPP til squalene. Reaksjonsblandinger (500 ml) innbefatter 50 mM Tris-HCl (pH 7,4), 2 mM KF, 1 mM MgCl<sub>2</sub>, 1 mM NADPH, enzym, 10 mM [1-3H] FPP (370 MBq/mmol, 3,7 kBq/ml; New England Nuclear, Boston, MA).

- 5 Reaksjoner startes ved å tilsette [1-3H] FPP. Etter 10 minutters inkubering ved 37 °C, avsluttes reaksjoner ved å tilsette 1 ml etanol. Etter at 1 ml H<sub>2</sub>O tilsettes ristes blandingene kraftig med 3 ml petroleumeter i 30 min. Ekstraherte lipider evaporeres og resuspenderes i 25 ml klorform. Prøver appliseres på ark med plast bakgrunn (Silica gel 60, F254; Merck, Rahway, NJ) for tynnsjiktskromatografi
- 10 (TLC), og utvikles i heptan i 15 minutter. Radioaktiviteter inkludert i squalenefraksjonen måles ved væskescintilla-sjonstilling. Når ekspresjonsvektor for *S. cerevisiae* anvendes kan en komplementeringsanalyse på beleilig måte utnyttes ved anvendelse av kondisjonell squalenesyntase mutant ERG 9 stamme avledet fra *S. cerevisiae* som vertsstamme for stadfesting av dets aktivitet
- 15 (Merkulov et al., *Yeast*, 16, 197-206, 2000).

- Etterfølgende bekreftelse av enzymaktivitet vil et uttrykt protein renses og benyttes for å frembringe antistoff mot det rensede enzym. Antistoff således frembrakt vil anvendes for karakterisering av ekspresjonen av det korresponderende enzym i en stammeforbedringsstudie, en optimeringsstudie av dyrkingsbetingelsen, og
- 20 lignende.

I en videre utførelsesform beskrives et antistoff som binder spesifikt til polypeptidet eller deler, dvs. spesifikke fragmenter eller epitoper av et slikt protein.

- Antistoffene kan anvendes for å identifisere og isolere andre squalenesyntaser og gener. Disse antistoffene kan være monoklonale antistoff, polyklonale antistoff eller
- 25 syntetiske antistoff så vel som fragmenter av antistoff, som for eksempel Fab, Fv eller scFv-fragmenter osv. Monoklonale antistoff kan fremstilles, for eksempel ved teknikker kjent for den faglærte, hvilket omfatter fusjonen av myelomaceller fra mus til miltceller avledet fra immuniserte pattedyr.

- Videre kan antistoff eller fragmenter derav mot de tidligere nevnte peptider oppnås
- 30 ved anvendelse av fremgangsmåter som er kjent for den faglærte. Disse antistoffene kan anvendes, for eksempel for immunutfelling og immunolokalisering av proteiner så vel som for overvåking av syntese av slike proteiner, for eksempel i rekombinante organismer, og for identifisering av forbindelser vekselvirkende med proteinet. For eksempel kan overflate plasmaresonans som benyttet i BIAcore

systemet anvendes til å øke effektiviteten av fagantistoff seleksjoner, derved oppnående en høyere stigning av affinitet fra et enkelt bibliotek av fagantistoff hvilket binder til en epitop på proteinet. I mange tilfeller er bindingsfenomenet av antistoff mot antigener ekvivalent med annen ligand/anti-ligand binding.

- 5 Her ble genfragmentet for squalenesyntase klonet fra *P. rhodozyma* med formålet å senke dets ekspresjonsnivå i *P. rhodozyma* ved genetiske fremgangsmåter ved anvendelse av det klonede genfragment.

Her bringes tilveie en prosess for produksjon av karo-tenoider hvori et gen som koder for squalenesyntase modifiseres i en passende vert, som for eksempel *P. rhodozyma* for å senke dets ekspresjon, og dyrking av en slik transformant i et passende medium under passende dyrkingsbetingelser.

10

For å senke en genekspresjon med genetiske metoder kan enkelte strategier benyttes. En av disse er en gen-disrupsjonsmetode. I denne fremgangsmåten ligeres et delfragment av målgenet som skal oppløses til en

15 legemiddelresistenskassett på integreringsvektoren som ikke kan replikere i vertsorganismen. Et legemiddelresistensgen som koder for enzymet som gjør vertsorganismen i stand til å overleve i nærvær av et toksisk antibiotikum anvendes ofte som den utvelgbare markør. G418 resistensgenet satt inn i pGB-Ph9 er et eksempel på et legemiddelresistensgen hvilket virker i *P. rhodozyma*.

20 Næringskomplementeringsmarkør kan også anvendes i vertsorganismen hvilket har en passende auxotrofimarkør. *P. rhodozyma* ATCC24221 stamme som be-høver cytidin for vekst er et eksempel av en auxotrof. Ved å anvende CTP syntetase som donor DNA for ATCC24221 kan et vertsvectorsystem som anvender en næringskomplementering etableres.

25 Etter transformering av vertsorganismene og rekombinering mellom målgenfragmentet på vektoren og dets korresponderende genfragment på kromosomet av vertsorganismene, integreres integreringsvektoren inn i vertskromosomet ved enkeltkryss rekombinering. Som et resultat av denne rekombinasjon, vil legemiddelresistenskassetten settes inn i målgenet hvis

30 translaterte produkt kun syntetiseres i sin trunkerte form hvilket ikke har enzymatisk funksjon. På lignende måte ble to deler av målgenet også anvendt for gendisrupsjonsstudie i hvilket legemiddelresistensgenet kan settes inn mellom to slike delfragmenter av målgenene på integreringsvektoren. I tilfelle av denne type vektor forventes doble rekombinasjonshendelser mellom genfragmentene satt inn i

integreringsvektoren og de korresponderende genfragmenter på kromosomet av  
 vertsorganismen. Selv om frekvensen av denne doble kryss-over rekombinasjonen  
 er lavere enn enkelt-kryss rekombinasjoner, er nullfenotype av målgenet ved  
 dobbel-tkryss rekombinasjonen mer stabil enn ved den enklel-kryss  
 5 rekombinasjonen.

Denne strategi ble anvendt for å konstruere den lycoplen-produserende  
 rekombinant av *Candida utilis* hvilket inneholder bakterielle karotenogene gener på  
 plasmidet (Shimada et al., (Applied and Environmental Microbiology, 64 (7), 2676-  
 2680, 1998)). ERG9 gen som koder for squalenesyntase?`? ble klonet fra *C. utilis*  
 10 og dets gendisruptant ble indusert etter dobbelt-kryss rekombinering av ERG9 gen  
 på kromosomet av den lycoplen-produserende *C. utilis*. Shimada et al. rapporterte  
 at disrupsjon av ERG9 gen ga en positiv effekt på karotenogenese ved den  
 rekombinante *C. utilis* særskilt avledet fra vertsorganismen i hvilket 3-  
 hydroksymetylgultaryl-CoA reduktase var amplifisert på det ribosomale DNA locus  
 15 multikopierte på kromosomet av vertsorganismen.

På den annen side har denne strategien vanskeligheter i tilfeller hvor genet har en  
 funksjon som er essensiell og disrupsjon er dødelig for vertsorganismen som for  
 eksempel squalenesyntasegen. I referansen ovenfor (Shimada et al.,) ble  
 disrupsjon laget på hvilken som helst av kopiene av squalenesyntasegenet innen de  
 20 to kopier av disse på vertskromosomet. I en slik konstruksjon ble det ikke bekreftet  
 at det senkede nivå av squalenesyntaseaktivitet var tilstrekkelig for å øke strømmen  
 av karbon inn i den karotenoide reaksjonsvei.

I et slikt tilfelle kan andre strategier appliseres for å senke (ikke ødelegge)  
 genekspresjon. En av disse er en konvensjonell mutagenese og å screene for  
 25 mutanter hvis ekspresjon for squalenesyntase synker. I denne fremgangsmåten  
 fusjoneres en passende rekombinant i hvilket et passende reporter-gen med  
 promotorregion av squalenesyntasegen fra vertsorganismen muteres og mutanter  
 som viser en svakere aktivitet av reporter-genprodukt kan screenes. I slike  
 mutanter forventes det at deres ekspresjon av squalene-syntaseaktivitet senket  
 30 ved mutasjonen ligger i promotorregion av reporter-gen eller trans-virkende region  
 hvilket kan påvirke ekspresjon av squalenesyntasegen annet enn mutasjonen som  
 ligger i promotorgenet i seg selv. I tilfelle av mutasjon forekommende ved  
 promotorregion av reporterfusjon kan slik mutasjon isoleres ved sekvensen av den  
 korresponderende region. Således isolert mutasjon kan introduseres i en rekke  
 35 forskjellige karotenoider, særskilt astaxanthinproduserende mutanter avledet fra *P.*

rhodozyma ved en rekombinasjon mellom den originale promoter for squalenesyntasegen på kromosomet og det muterte promoterfragment. For å ekskludere mutasjoner forekommende ved trans-virkende region kan en mutasjon også induseres av en in vitro mutagenese av et cis-element i promotorregion. I

5 denne tilnæringsmåten mutageniseres en genkassett, inneholdende et reporter-gen hvilket er fusjonert med en promotorregion avledet fra et gen av interesse ved dets 5'-ende og en terminatorregion fra et gen av interesse ved dets 3'-ende, og introduseres deretter inn i *P. rhodozyma*. Ved å detektere forskjell mellom aktiviteten av reporter-gen kan det screenes for en effektiv mutasjon. En

10 slik mu-tasjon kan introduseres i sekvensen av den native pro-moterregion på kromosomet ved den samme metode som i tilfellet av en in vivo mutasjonstilnærming. Men, disse fremgangsmåter har enkelte bakdeler som for eksempel tidsforbrukende prosesser.

En annen strategi for å senke en genekspresjon er en antisensmetode. Denne

15 fremgangsmåten appliseres ofte for å senke genekspresjon selv når teleomorfiske organismer som for eksempel *P. rhodozyma* anvendes som vertsorganismer, til hvilke mutasjons- og gendisrupsjonsmetoder vanligvis med vanskelighet kan appliseres. Anti-sensemotoden er en fremgangsmåte for å senke ekspresjon av gen av interesse ved å introdusere et kunstig genfragment, hvis sekvens er

20 komplementær med cDNA fragment av genet av interesse.

Et "antisens" nukleinsyremolekyl omfatter en nukleotid-sekvens hvilket er komplementær med et "sens" nukleinsyremolekyl som koder for et protein, for eksempel komplementær med den kodende tråd av et dobbeltrådet cDNA molekyl eller komplementær med en mRNA sekvens. Som det følger av dette kan et

25 antisens-nukleinsyremolekyl hydrogenbinde med et sens-nukleinsyremolekyl. Antisens-nukleinsyremolekylet kan være komplementært med en fullstendig squalenesyntase-kodende tråd, eller med kun en del derav. Som det følger av dette kan et antisens-nukleinsyremolekyl være antisens med en "kodende region" av den kodende tråd av en nukleotidsekvens som koder for en squalenesyntase.

30 Benevnelsen "kodende region" refererer til regionen av nukleotidsekvensen omfattende kodoner hvilket translateres til aminosyrerester. Videre er antisens-nukleinsyremolekylet antisens med en "ikke-kodende region" av den kodende tråd av en nukleotidsekvens som koder for squalenesyntase. Benevnelsen "ikke-kodende region" refererer til 5' og 3' sekvenser hvilket omslutter den kodende

35 region som ikke translateres til et polypeptid (dvs. også referert til som 5' og 3' ikke-translaterte regioner).

Gitt de kodende trådsekvenser som koder for squalenesyntase brakt for dagen her kan antisens-nukleinsyremolekyler designes i henhold til Watson og Cricks basesammenpassingsregler. Antisens-nukleinsyremolekylet kan være komplementært med den fullstendige kodende region av squalenesyntase mRNA, men kan også være et oligonukleotid hvilket er antisens med kun en del av den kodende eller ikke-kodende region av squalenesyntase mRNA. For eksempel kan antisens-oligonukleotid være komplementært med regionen omsluttende translasjonsstartsete av squalenesyntase mRNA. Et antisens oligonukleotid kan være, for eksempel omtrent 5, 10, 15, 20, 25, 30, 35, 40, 45 eller 50 nukleotider langt. Et antisens-nukleinsyremolekyl kan konstrueres ved anvendelse av kjemisk syntese og enzymatisk ligeringsreaksjoner ved anvendelse av fremgangsmåter kjent innen faget. For eksempel kan et antisens-nukleinsyremolekyl (for eksempel et antisens oligonukleotid) kjemisk syntetiseres ved anvendelse av naturlig forekommende nukleotider eller på annen måte modifiserte nukleotider designet for å øke den biologiske stabilitet av molekylerne eller for å øke den fysiske stabilitet av duplekset dannet mellom antisens- og sens- nukleinsyrer, for eksempel kan fosforo-tioatderivater og acridinsubstituerte nukleotider anvendes. Eksempler av modifiserte nukleotider som kan anvendes for å frembringe anti-sens nukleinsyre innbefatter 5-fluoruracil, 5-bromuracil, 5-kloruracil, 5-ioduracil, hypoxantin, xantin, 4-acetylcytosin, 5-(karboksy-hydroksyl-metyl)uracil, 5-karboksymetylaminometyl-2-tiouridin, 5-karboksymetylaminometyluracil, dihydrouracil, beta-D-galactosylqueosin, inosin, N6-isopentenyladenin, 1-metylguanin, 1-metylinosin, 2,2-dimetylguanin, 2-metyl-adenin, 2-metylguanin, 3-metylcytosin, 5-metylcytosin, N6-adenin, 7-metylguanin, 5-metylaminometyluracil, 5-metoksyaminometyl-2-tiouracil, beta-D-mannosylqueosin, 5'-metoksykarboksymetyluracil, 5-metoksyuracil, 2-metyltio-N6-isopentenyladenin, uracil-5-oksyeddiksyre (v), wybutoxosin, pseudouracil, queosin, 2-thiocytosin, 5-metyl-2-tiouracil, 2-tiouracil, 4-tiouracil, 5-metyl-uracil, uracil-5-oksyeddiksyre metylester, uracil-5-oksyeddiksyre (v), 5-metyl-2-tiouracil, 3-(3-amino-3-N-2-karboksypropyl)uracil, (acp3)<sub>w</sub>, og 2,6-diaminopurin. Alternativt kan antisens-nukleinsyren produseres biologisk ved anvendelse av en ekspresjonsvektor i hvilket et polynukleotid er subklonet i en antisensretning (dvs. RNA transkribert fra det innsatte polynukleotid vill være i antisensretning i forhold til et målpolynukleotid av interesse, beskrevet videre i den følgende underseksjon).

Antisens-nukleinsyremolekyler administreres typisk til en celle eller genereres in situ slik at de hybridiserer med og binder til cellulært mRNA og/eller genomisk DNA

som koder for en squalenesyntase for derved å inhibere ekspresjon av proteinet, for eksempel ved å inhibere transkripsjon og/eller translasjon. Hybridiseringen kan være ved konvensjonell nukleotidkomplementaritet for å danne et stabilt dupleks, eller, for eksempel i tilfelle et antisens-nukleinsyremolekyl som binder til DNA

5 duplekser, gjennom spesifikk vekselvirkning i hovedkløften av dobbelheliksen. Antisensmolekylet kan modifiseres slik at det spesifikt binder til en reseptor eller et antigen uttrykt på en valgt celleoverflate, for eksempel ved å lenke antisens-nukleinsyremolekylet til et peptid eller et antistoff hvilket binder til en celleoverflatereseptor eller antigen. Antisens-nukleinsyremolekylet kan også

10 avleveres til celler ved anvendelse av vektorene beskrevet her. For å oppnå tilstrekkelig intracellulær konsentrasjon av anti-sensmolekylene foretrekkes vektorkonstrukt i hvilke antisens-nukleinsyremolekylet plasseres under kontroll av sterke prokaryote, virale eller eukaryote, inklusive plante, promotere.

I videre utførelsesform kan antisens-nukleinsyremolekylet være et  $\alpha$ -anomerisk nukleinsyremolekyl. Et  $\alpha$ -anomerisk nukleinsyremolekyl danner spesifikke

15 dobbelttrådede hybrider med komplementær RNA i hvilket, i motsetning til de vanlige  $\beta$ -enheter, trådene løper parallelt med hverandre. Antisens-nukleinsyremolekylet kan også omfatte et 2'-o-metylribonukleotid eller en kimerisk RNA-DNA analog.

Videre kan antisens-nukleinsyremolekylet være et ribozym. Ribozym er

20 katalytiske RNA molekyler med ribonukleaseaktivitet som er i stand til å spalte en enkelttrådet nukleinsyre, som for eksempel et mRNA, til hvilket de har en komplementær region. Således kan ribozym (for eksempel "hammerhead ribozym") anvendes for å katalytisk spalte squalenesyntase mRNA tran-skripter

25 for derved å inhibere translasjon av mRNA. Et ribozym som har spesifisitet for en squalenesyntasekodende nukleinsyremolekyl kan designes med basis i polynukleotidsekvensen . For eksempel kan et derivat av en Tetrahymena L-19IVS RNA konstrueres i hvilket nukleotidsekvensen av det aktive sete er komplementært med nukleotidsekvensen som skal spaltes i en som koder for mRNA (US

30 4,987,071 og US 5,116,742.) Alternativt kan squalenesyntase mRNA anvendes for å selektere et katalytisk RNA som har en spesifikk ribonukleaseaktivitet fra en samling av RNA molekyler.

Anvendelse av antisensmetode for å konstruere en karo-tenoidoverproduserende stamme av *P. rhodozyma* ble eksemplifisert i EP 1,158,051.

I en utførelsesform beskrives her en metode for å frem-bringe en rekombinant vertscelle omfattende å introdu-sere vektoren eller polynukleotidet inn i en vertscelle.

Vektor DNA kan introduseres inn i prokaryote eller euka-ryote celler via  
 5 konvensjonelle transformasjon eller transfeksjonsteknikker. Med benevnelsene  
 "transformasjon" og "transfeksjon" er konjugering og transduksjon ment å referere  
 til en rekke forskjellige fagkjente teknikker for å introdusere fremmed nukleinsyre  
 (for eksempel DNA) inn i en vertscelle, inklusive kalsiumfosfat eller kalsiumklorid  
 sam-utfelling, DEAE-dextran-mediert transfeksjon, lipofek-sjon, naturlig  
 10 kompetens, kjemisk-mediert overføring eller elektroporering. Passende  
 fremgangsmåter for transformering eller transfektering av vertsceller inklusive  
 planteceller er kjent for den faglærte.

For stabil transfeksjon av mammaliske celler er det kjent at, avhengig av  
 ekspresjonsvektoren og transfeksjonsteknikken anvendt, kun en liten andel av  
 15 celler kan integrere det fremmede DNA inn i sitt genom. For å kunne identifisere og  
 velge disse integranter introduseres generelt et gen som koder for en utvelgbar  
 markør (for eksempel resistens overfor antibiotika) inn i vertscellene sammen med  
 genet av interesse. Foretrukne utvelgbare markører innbefatter de som tilfører  
 resistens overfor legemidler, som for eksempel G418, hygromycin og metotrexat.  
 20 Nukleinsyre som koder for en utvelgbar markør kan introduseres inn i en vertscelle  
 på den samme vektor som den som koder for polypeptidet eller kan introduseres på  
 en separat vektor. Celler stabilt transfektert med den introduserte nukleinsyre kan  
 identifiseres ved, for eksempel legemiddelseleksjon (for eksempel celler som har  
 inkorporert det utvelgbare markørgen vil overleve, mens de andre celler vil dø).

25 For å lage en homolog rekombinant mikroorganisme fremstilles en vektor hvilket  
 inneholder i det minste en del av polynukleotidet inni i hvilket en delesjon, addisjon  
 eller substitusjon er introdusert for derved å endre, for eksempel funksjonelt  
 forstyrre squalenesyntasegenet. Fortrinnsvis er dette squalenesyntasegenet et *P.*  
*rhodozyma* squalenesyntasegen, men det kan være en homolog fra en relatert eller  
 30 forskjellig kilde. Alternativt kan vektoren designes slik at, ved homolog  
 rekombinering, det endogene squalenesyntasegen muteres eller på annen måte  
 endres men fremdeles koder for funksjonelt protein (for eksempel kan den  
 oppstrøms regulatoriske region endres for derved å endre ekspresjonen av den  
 endogene squalenesyntase). For å lage en punktmutasjon via homo-log  
 35 rekombinasjon kan også DNA-RNA hybrider anvendes kjent som chimeraplasi.

Vektoren introduseres inn i en celle og celler i hvilket det introduserte polynukleotidgen homologt har rekombinert med det endogene squalenesyntasegen velges ved anvendelse av teknikker kjent innen faget.

Videre kan vertsceller produseres hvilke inneholder seleksjonssystemer hvilket  
 5 legger til rette for regulert ekspresjon av det introduserte gen. For eksempel tillater inklusjon av polynukleotidet på en vektor som plasserer det under kontroll av lac-operonet ekspresjon av polynukleotidet kun i nærvær av IPTG. Slike regulatoriske systemer er velkjent innen faget.

Fortrinnsvis er det introduserte nukleinsyremolekyl fremmed for vertscellen.

10 Med "fremmed" er det ment at nukleinsyremolekylet enten er heterologt med, respektivt til vertscellen, dette betyr avledet fra en celle eller organisme med en forskjellig genomisk bakgrunn, eller er homolog med hensyn til vertscellen, men lokalisert i et forskjellig genomisk miljø enn den naturlig forekommende motpart av nevnte nukleinsyremolekyl. Dette betyr at, hvis nukleinsyremolekylet er homologt  
 15 med hensyn til vertscellen, det ikke er lokalisert i sin naturlige plassering i genomet av nevnte vertscelle, i særdeleshet er det omsluttet av forskjellige gener. I dette tilfelle kan nukleinsyremolekylet være enten under kontroll av sin egen promoter eller under kontroll av en heterolog promoter. Vektoren eller nukleinsyremolekylet hvilket er tilstede i vertscellen kan enten integreres inn i genomet av vertscellen  
 20 eller det kan vedlikeholdes ekstrakromosomalt i enkelte former. I dette hensyn skal det også forstås at nukleinsyremolekylet kan anvendes til å gjenopprette eller danne et mutant gen via homolog rekombinasjon.

Som det følger av dette beskrives i en annen utførelses-form en vertscelle genteknologisk fremstilt med poly-nukleotidet eller vektoren.

25 Benevnelsene "vertscelle" og "rekombinant vertscelle" anvendes omvekslende her. Det forstås at slike benevnelser refererer ikke bare til den spesielle måicelle, men til avstamninger eller potensielle avstamninger av en slik celle. Fordi enkelte modifikasjoner kan forekomme i etterfølgende generasjoner på grunn av enten mutasjon eller miljømessige påvirkninger behøver slike avstamninger faktisk, å  
 30 være identiske med opphavscellen, men inkluderes fortsatt innen omfanget av benevnelsen.

For eksempel kan et polynukleotid introduseres i bakterielle celler så vel som insektceller, soppceller eller mammaliske celler (som for eksempel Chinese hamster ovarium celler (CHO) eller COS celler), alger, ciliater, planteceller, sopp eller andre mikroorganismer som *E. coli*. Andre passende vertsceller er kjent for de med kunnskaper i faget. Foretrukne er *E. coli*, baculovirus, *Agrobacterium* eller soppceller er, for eksempel de av genuset *Saccharomyces*, for eksempel de av artene *S. cerevisiae* eller *P. rhodozyma* (*Xanthophylomyces dendrorhous*).

I tillegg beskrives i en utførelsesform en metode for produksjon av soppransformerer omfattende introduksjon av polynukleotidet eller vektoren inn i genomet av nevnte soppcelle.

For ekspresjon av nukleinsyremolekylene i sens- eller antisens-retning i planteceller plasseres molekylene under kontroll av regulatoriske elementer hvilket sikrer uttrykk i soppceller. Disse regulatoriske elementer kan være heterologe eller homologe med hensyn til nukleinsyremolekylet som skal uttrykkes så vel som med hensyn til soppartene som skal transformeres.

Generelt omfatter slike regulatoriske elementer en promotor aktiv i soppceller. For å oppnå konstitutiv ekspresjon i soppceller anvendes fortrinnsvis konstitutive promotere, som for eksempel glycerinaldehyd-3-dehydrogenasepromoter fra *P. rhodozyma* (WO 97/23,633). Induserbare promotere kan brukes for å være i stand til å eksakt kontrollere ekspresjon. Et eksempel av induserbare promotere er promotor av gener som koder for varmesjokkproteiner. Også en amylasegenpromoter hvilket er en kandidat for slike induserbare promotere er beskrevet (EP 1,035,206). De regulatoriske elementer kan videre omfatte transkripsjonelle og/eller translasjonelle forsterkere som er virksomme i soppceller. Videre kan de regulatoriske elementer bestå av transkripsjonstermineringssignaler, som for eksempel et poly-A signal, hvilket fører til tilsetning av en poly-A hale til transkriptet hvilket kan forbedre dets stabilitet.

Fremgangsmåter for introduksjon av fremmed DNA inn i soppceller er også vel kjent innen faget. Disse inkluderer, for eksempel transformering med LiCl-metode, fusjon av protoplaster, elektroporering, biolistiske fremgangsmåter lik partikkelbombardement og andre metoder kjent innen faget. Fremgangsmåter for fremstilling av passende vektorer er kjent for den fagkyndige. Fremgangsmåter for transformering ved anvendelse av biolistiske fremgangsmåter er vel kjente for personen med kunnskaper innen faget.

Benevnelsen "transformering" refererer til overføring av et eksogent polynukleotid inn i en vertscelle, uten hen-syn til fremgangsmåten anvendt for overføringen. Polynukleotidet kan være transient eller stabilt introdusert inn i vertscellen og kan opprettholdes ikke-integrert, for eksempel som et plasmid eller som kimeriske  
5 lenker, eller alternativt, kan integreres inn i vertsgenomet.

Generelt kan sopp hvilket kan være modifisert og hvilket enten viser overekspressjon av et protein eller en reduksjon av syntesen av et slikt protein avledes fra hvilken som helst ønsket soppart.

Videre beskrives i en utførelsesform en soppcelle omfattende polynukleotidet,  
10 vektoren oppnåelig ved fremgangsmåten beskrevet ovenfor.

Således beskrives også transgene soppceller hvilket inneholder (fortrinnsvis stabilt integrert inn i genomet) et polynukleotid linket til regulatoriske elementer hvilket legger til rette for ekspressjon av polynukleotidet i soppceller og karakterisert ved at polynukleotidet er fremmed for den transformerte soppcelle. For betydningen av  
15 fremmed se supra.

Nærvær og ekspressjon av polynukleotidet i de transformerte soppceller modulerer, fortrinnsvis senker syntese av squalene og fører til økning av karotenoidproduksjon, særskilt astaxanthinproduksjon i således oppnådde transformerte soppceller; fortrinnsvis i *P. rhodozyma* celler.

20 Således beskrives her også transformerte soppceller.

Som det følger av dette moduleres, på grunn av den endrede ekspressjon av squalenesyntase, cellers metaboliske reaksjonsveier med tanke på produksjonsutbytte og/eller produksjonseffektivitet.

Benevnelsene "produksjon" eller "produktivitet" er kjent innen faget og innbefatter  
25 konsentrasjonen av fermentasjonsprodukt (for eksempel fettsyrer, karotenoider, (po-ly)sakkarider, vitaminer, isoprenoide, lipider, voksester, og/eller polymerer som polyhydroksyalkanoater og/eller dets metabolismeprodukter eller videre ønsket finkjemikalie som nevnt her) dannet innen en gitt tidsperiode og et gitt fermentasjonsvolum (for eksempel kg produkt/tid/liter).

Benevnelsen "effektivitet" av produksjon innbefatter tiden nødvendig for å oppnå et spesielt nivå av produksjon (for eksempel hvor lenge det tar for cellen for å oppnå en spesiell hastighet av output av et nevnt endret utbytte, i særdeleshet, til karotenoider, (poly)sakkarider, lipider, vitaminer, isoprenoider, osv.).

- 5 Benevnelsen "utbytte" eller "produkt/karbonutbytte" er kjent innen faget og innbefatter effektiviteten av omdanningen av karbonkilden til produktet (dvs. acetyl-CoA, fettsyrer, karotenoider, vitaminer, isoprenoider, lipider osv, og/eller videre forbindelser som definert ovenfor og hvilket biosyntesen av baseres på nevnte produkter). Dette skrives generelt som, for eksempel kg produkt per kg
- 10 karbonkilde. Ved å øke utbyttet eller produksjonen av forbindelsen økes kvantiteten av samlede molekyler, eller av anvendbare samlede molekyler av forbindelsen i en gitt mengde kultur over en gitt tidsperiode.

- Benevnelsene "biosyntese" (hvilket anvendes synonymt med "syntese" av "biologisk produksjon" i celler, vev, plan-ter, osv.) eller en "biosyntetisk
- 15 reaksjonsvei" er kjent innen faget og innbefatter syntese av en forbindelse, fortrinnsvis en organisk forbindelse, av en celle fra intermediære forbindelser i hva som kan være en multitrinn og høyst regulert prosess.

- Språkformen "metabolisme" er kjent innen faget og innbefatter totaliteten av de biokjemiske reaksjoner som finner sted i en organisme. Metabolismen av en spesiell
- 20 forbindelse omfatter, deretter, (for eksempel metabolismen av acetyl-CoA, en fettsyre, hexose, lipid, isoprenoid, vitamin, karotenoid osv.) de fullstendige biosyntetiske, modifierende, og degraderingsreaksjonsveier i cellen relatert til denne forbindelse.

- En slik genteknologisk fremstilt *P. rhodozyma* ville dyrkes i et passende medium og
- 25 evaluert med tanke på dens produktivitet og/eller utbytte av karotenoider, særskilt astaxanthin. En hyperproducent av astaxanthin således valgt ville bekreftes med hensyn til forholdet mellom dens produktivitet og nivået av gen eller proteinekspresjon som introduseres ved en slik genteknologisk fremgangsmåte.

Den foreliggende oppfinnelse illustreres videre med eksempler beskrevet nedenfor.

- 30 De følgende materialer og metoder ble benyttet i eksemplet beskrevet nedenfor:

## Stammer

*P. rhodozyma* ATCC96594 (re-depositert under aksesjonsnr. ATCC 74438 den 8. april, 1998 i samsvar med Budapest-konvensjonen)

5 *E. coli* DH5 $\alpha$ : F-,  $\Phi$ 80d, lacZ $\Delta$ M15,  $\Delta$ (lacZYA-argF)U169, hsd(rk-, M $\kappa$ +), recA1, endA1, deoR, thi-1, supE44, gyrA96, relA1 (Toyobo, Osaka, Japan)

*E. coli* XL1 MRA (P2):  $\Delta$ (mcrA)183,  $\Delta$ (mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, gyrA96, relA1, lac(P2 lysogen) (Stratagene, La Jola, U.S.A.)

## Vektorer

$\lambda$ DASHII (Stratagene)

10 pBluescriptII KS- (Stratagene)

pMOSBlue T-vektor (Amersham, Buckinghamshire, U.K.)

## Media

*P. rhodozyma* stamme ble rutinemessig vedlikeholdt i YPD-medium (DIFCO, Detroit, U.S.A.). *E. coli* stamme ble vedlikeholdt i LB-medium (10 g Bacto-trypton, 5 g  
15 gjærekstrakt (DIFCO) og 5 g NaCl per liter). NZY-medium (5 g NaCl, 2 g MgSO<sub>4</sub>-7H<sub>2</sub>O, 5 g gjærekstrakt (DIFCO), 10 g NZ amin type A (WAKO, Osaka, Japan) per liter) anvendes for  $\lambda$ -fag propagering i en myk agar (0,7 % agar (WAKO)). Når et agarmedium ble forberedt ble 1,5 % agar (WAKO) supplementert.

## Fremgangsmåter

20 Restriksjonsenzymmer og T4 DNA-ligase ble anskaffet fra Takara Shuzo (Ohtsu, Japan).

Isolering av kromosomalt DNA fra *P. rhodozyma* ble utført ved anvendelse av QIAGEN Genomisk Kit (QIAGEN, Hilden, Tyskland) etterfølgende protokollen tatt frem av leverandøren. Mini-prep av plasmid DNA fra transformert *E. coli* ble utført  
25 med det automatiske DNA-isolerings systemet (PI-50, Kurabo, Co. Ltd., Osaka, Japan). Midi-prep av plasmid DNA fra en *E. coli* transformant ble utført ved

anvendelse av QIAGEN kolonne(QIAGEN). Isolering av  $\lambda$ -DNA ble utført ved "Wizard lambda preps DNA purification system" (Promega, Madison, U.S.A.) etterfølgende protokollen frembrakt av leverandøren. Et DNA fragment ble isolert og rensset fra agarose ved anvendelse av QIAquick eller QIAEX II (QIAGEN). Manipulering av  $\lambda$ -  
5 fagderivater ble gjort i henhold til protokollen frembrakt av leverandøren (Stratagene).

Isolering av total RNA fra *P. rhodozyma* ble utført med fenolmetoden ved anvendelse av Isogen (Nippon Gene, Toyama, Japan). mRNA ble rensset fra total RNA således oppnådd ved anvendelse av mRNA separasjonskit (Clontech). cDNA  
10 ble fremstilt ved anvendelse av CapFinder cDNA konstruksjonskit (Clontech).

In vitro pakking ble utført ved anvendelse av "Gigapack III gold packaging extract" (Stratagene).

Polymerase kjedereaksjon (PCR) utføres med den termiske syklisator fra Perkin Elmer modell 2400. Hver PCR betingelse er beskrevet i eksempler. PCR primere ble  
15 anskaffet fra en kommersiell leverandør. Fluorescerende DNA primere for DNA sekvensering ble anskaffet fra Pharmacia. DNA sekvensering ble utført med den automatiserte fluorescerende DNA-sekvensator (ALFred, Pharmacia).

Kompetente celler av DH5 $\alpha$  ble anskaffet fra Toyobo (Ja-pan).

Eksempel 1: Isolering av mRNA fra *P. rhodozyma* og konstruksjon av cDNA  
20 bibliotek

For å konstruere et cDNA bibliotek av *P. rhodozyma* ble total RNA isolert ved fenolekstraksjonsmetoden rett etter celledisrupsjonen og mRNAet fra *P. rhodozyma* ATCC96594 stamme ble rensset ved anvendelse av mRNA separeringskit (Clontech). Celler av stamme ATCC96594 fra 10 ml av en to-dagers-kultur i YPD-medium ble  
25 høstet ved sentrifugering (1500 x g i 10 min.) og vasket en gang med ekstraksjonsbuffer (10 mM Na-citrat / HCl (pH 6,2) inneholdende 0,7 M KCl). Etter suspensering i 2,5 ml ekstraksjonsbuffer ble cellene oppløst ved fransk presse homogenisator (Ohtake Works Corp., Tokyo, Japan) ved 1500 kgf/cm<sup>2</sup> og øyeblikkelig blandet med to ganger volumet av isogen (Nippon gene) i henhold til  
30 fremgangsmåten spesifisert av produsenten. I dette steg ble 400  $\mu$ g total RNA samlet.

Deretter ble dette total RNA rensset ved anvendelse av mRNA separeringskit (Clontech) i henhold til fremgangsmåten spesifisert av produsenten. Endelig ble 16 µg av mRNA fra *P. rhodozyma* ATCC96594 stamme oppnådd.

For å konstruere cDNA bibliotek ble CapFinder PCR cDNA konstruksjonskit (Clontech) anvendt i henhold til frem-gangsmåten spesifisert av produsenten. Et µg rensset mRNA ble applisert for en første streng syntese etterfulgt av PCR amplifisering. Etter denne amplifisering ved PCR ble 1 mg cDNA pool oppnådd.

Eksempel 2: Kloning av et partielt SQS (squalenesyntase) gen fra *P. rhodozyma*

For å klonere et partielt SQS gen fra *P. rhodozyma*, ble en degenererende PCR metode utnyttet. Arter og aksesjonsnummer til database hvis sekvens for squalenesyntase ble anvendt for multippel linjestillingsanalyse er som følger.

	<i>Ustilago maydis</i>	Q92459 (SwissProt)
	<i>Schizosaccharomyces pombe</i>	P36596 (SwissProt)
	<i>Saccharomyces cerevisiae</i>	M63979 (GenBank)
15	<i>Rattus norvegicus</i>	Q02769 (SwissProt)
	<i>Mus musculus</i>	P53798 (SwissProt)
	<i>Candida albicans</i>	P78589 (SwissProt)
	<i>Homo sapiens</i>	I38245 (Pir)
	<i>Arabidopsis thaliana</i>	U79159, AF004396
20	<i>Leishmania major</i>	U30455
	<i>Glycyrrhiza glabra</i>	D86410

To blandede primere hvis nukleotidsekvenser ble designet og syntetisert basert på felles sekvens av kjente squalenesyntasegener fra andre arter, dvs. *squ1* (senseprimer) (SEKV.ID.NR.: 4), *squ4* (antisenseprimer) (SEKV.ID.NR.:5) og *squ5*

(antisenseprimer) (SEKV.ID.NR.: 6) (i sekvensene "n" betyr nukleotider a, c, g eller t, "r" betyr nukleotider a eller g, og "y" betyr nukleotider c eller t).

Etter PCR reaksjonen ved 25 sykluser av 95 °C i 30 sekunder, 45 °C i 30 sekunder og 72 °C i 15 sekunder ved anvendelse av "ext." (Takara Shuzo) som en DNA  
 5 polymerase og cDNA pool oppnådd i eksempel 1 som et templat, ble reaksjonsblandingen anvendt for agarose gelelektroforese. Hvert PCR bånd som hadde en ønsket lengde ble gjenvunnet fra PCR reaksjonsblandingen i hvilket kombinasjonen som anvendte henholdsvis squ1 og squ4, og squ1 og squ5, og renses ved QIAquick (QIAGEN) i henhold til produsentens fremgangsmåte og  
 10 deretter ligert til pMOSBlue-T-vektor (Amersham). Etter transformering av kompetent *E. coli* DH5 $\alpha$ , ble 6 hvite kolonier valgt og plasmider ble isolert med automatisk DNA isoleringssystem. Som et resultat av sekvensering ble det funnet at 3 kloner hadde en sekvens hvis avledede aminosyresekvens lignet kjente squalene-syntasegener. Disse isolerte cDNA kloner ble benevnt pSQS1007 avledet  
 15 fra PCR reaksjonen ved anvendelse av squ1 og squ5 og pSQS1006 avledet fra PCR reaksjonen ved anvendelse av squ1 og squ4, og pSQS1006 ble anvendt for videre screeningsstudier.

#### Eksempel 3: Isolering av genomisk DNA fra *P. rhodozyma*

For å isolere et genomisk DNA fra *P. rhodozyma*, ble QIAGEN genomisk kit anvendt  
 20 i henhold til fremgangsmåten spesifisert av produsenten.

Celler av *P. rhodozyma* ATCC96594 fra 100 ml av over-natt kultur i YPD-medium ble høstet ved sentrifugering (1500 x g i 10 min.) og vasket en gang med TE buffer (10 mM Tris / HCl (pH 8,0) inneholdende 1 mM EDTA). Etter suspensering i 8 ml Y1-buffer av QIAGEN genomisk kit, lyticase (SIGMA, St. Louis, U.S.A.) ble tilsatt  
 25 ved en konsentrasjon av 2 mg/ml for å oppløse celler ved enzymatisk degradering og reaksjonsblandingen ble inkubert i 90 minutter ved 30 °C og deretter brakt videre til det neste ekstraksjonssteg. Endelig ble 20  $\mu$ g genomisk DNA oppnådd.

#### Eksempel 4: Southern blot hybridisering ved anvendelse av pSQS1006 som en probe

30 Southern blot hybridisering ble utført for å klonere et genomisk fragment hvilket inneholder SQS gen fra *P. rhodozyma*. To  $\mu$ g genomisk DNA ble nedbrutt ved EcoRI og underlagt agarose gelelektroforese etterfulgt av syrebehandling og alkalisk

behandling. Det denaturerte DNA ble overført til nylonmembran (Hybond N+, Amersham) ved anvendelse av transblot (Joto Rika, Tokyo, Japan) i en time. DNAet som ble overført til nylonmembran ble fiksert ved en varmebehandling (80 °C, 90 min). En probe ble forberedt ved å merke et templat DNA (EcoRI-nedbrutt pSQS1006) med DIG multiprimingsmetode (Boehringer Mannheim). Hybridisering ble utført med fremgangsmåten spesifisert av produsenten. Som et resultat ble et hybridisert bånd visualisert i området fra 9,0 to 23,0 kilobaser (kb).

#### Eksempel 5: Kloning av et genomisk fragment inneholdende SQS gen

Fire µg av det genomiske DNA ble nedbrutt ved EcoRI og underlagt agarose gelelektroforese. Deretter ble DNAer hvis lengde er i området fra 9,0 til 20,0 kb gjenvunnet ved QIAEX II gelekstraksjonskit (QIAGEN) i henhold til fremgangsmåten spesifisert av produsenten. Det rensede DNA ble ligert til 0,5 µg EcoRI-nedbrutt og CIAP (kalv intestin alkalisk fosfatase)-behandlet λDASH II (Stratagene) ved 16 °C natten over, og pakket ved "Gigapack III gold packaging extract" (Stratagene). Det pakkede ekstrakt ble infektert i E. coli MRA(P2) stamme og overlagt med NZY-medium helt på LB-agarmedium. Omtrent 5000 plakker ble screenet ved anvendelse av EcoRI-nedbrutt pSQS1006 som probe. Fem plakker ble hybridisert til den merkede probe.

Dette λDASH II derivat inneholdende putativ SQS gen fra P. rhodozyma ble forberedt ved anvendelse av "Wizard lambda preps DNA purification system" (Promega). Deretter ble PCR gjennomført ved anvendelse av disse λDASH II derivater som et templat og to primere, squ9 og squ10 som primere. Disse squ9 og squ10 primere ble designet basert på den interne sekvens av pSQS1006: squ9 (senseprimer) (SEKV.ID.NR.: 7) og squ10 (antisenseprimer) (SEKV.ID.NR.: 8).

Som et resultat av PCR under de samme PCR betingelser som beskrevet i eksempel 2 ble et forventet 0,5 kb band oppnådd. Det ble foreslått at alle disse λDASH II derivater kunne inneholde et putativt SQS gen fra P. rhodozyma. Omtrentlig 20,0 kb EcoRI insettfragment i en av disse λDASH II derivater ble rensed ved anvendelse av QIAEX II (QIAGEN) og underlagt subkloning inn i pBluescriptII KS-vektor (Stratagene) ved anvendelse av DH5α som vertsstamme og frembrakte pSQ1229pSQS1229.

#### Eksempel 6: Sekvensering av et genomisk fragment inneholdende SQS gen

pSQS1229 ble sekvensert med primervandringsprosedyre ved anvendelse av "AutoRead sekvensering kit" (Pharmacia).

Som et resultat av sekvensering ble en nukleotidsekvens omfattende 4807 basepar av et genomisk fragment innehol-dende SQS gen fra *P. rhodozyma* inneholdende  
5 dets promoter (1549 bp) og terminator (836 bp) bestemt (SEKV.ID.NR.: 1).

Den kodende region var 2422 basepar lang bestående av 9 eksoner og 8 introner. Introner var spredt gjennom hele den kodende region uten 5' eller 3' bias. Det ble funnet at en åpen leseramme (SEKV.ID.NR.:2) omfatter 512 aminosyrer (SEKV.ID.NR.: 3) hvis sekvens er slående lik den kjente aminosyresekvens av  
10 squalenesyntase fra andre arter (51,3 % identitet med squalenesyntase fra *Schizosaccharomyces pombe*) som et resultat homologisøking ved GENETYX-SV/RC programvare (Software Development Co., Ltd., Tokyo, Japan).

#### Eksempel 7: Konstruksjon av antisenseplasmid for SQS gen

Et antisense genfragment hvilket dekker det fullstendige strukturgen for SQS gen  
15 amplifiseres ved PCR fremgangsmåte og klones deretter inn i integrasjonsvektor i hvilket antisense SQS gen transkriberes ved sin egen SQS promoter i *P. rhodozyma*. Slike primere innbefatter asymmetriske gjen-kjennelsessekvens for restriksjonsenzym, Sfil (GGCCNNNNNGGCC), men deres asymmetriske overhengsekvens designes for å være forskjellige. Dette muliggjør en dirigert  
20 kloning inn i ekspressjonsvektor hvilket har den samme asymmetriske sekvens ved sin ligeringssekvens. Anvendelsen av en slik konstruksjon er eksemplifisert i EP 1,158,051.

For promotor og terminatorfragment som kan drive transkripsjonen av antisense SQS genet, klones SQS promoter og terminator fra kromosomet ved anvendelse av  
25 sekvensinformasjonen opplistet i SEKV.ID.NR.: 1.

Deretter fusjoneres SQS terminator fragment med G418 re-sistenskassett ved å ligere DNA fragmentet inneholdende SQS terminator med G418 resistens kassett av pG418Sa330 (EP 1,035,206) til passende vektor som for eksempel pBluescriptII KS- (Stratagene).

30 Deretter ble 3,1 kb SacI fragment inneholdende ribosomalt DNA (rDNA) locus (Wery et al., *Gene*, 184, 89-97, 1997) settes inn nedstrøms for G418 kassetten på

således fremstilte plasmid. rDNA fragmentet forekommer i multikopier på kromosomet av eukaryot. Integreringshendelsen via rDNA fragmentet ville resultere i multikopierte integrering i den anvendte vertsorganismens kromosom og dette muliggjør overekspresjon av fremmede gener hvilket finnes i ekspresjonsvektor. Deretter settes SQS promoter inn oppstrøms for SQS terminator for å konstruere ekspresjonsvektor hvilket virker i *P. rhodozyma*. Endelig avsluttes antisense SQS konstrukt avsluttet ved å sette inn 1,5kb SfiI fragment inneholdende antisense SQS inn i således frembrakt ekspresjonsvektor funksjonell i *P. rhodozyma*. En lignende plasmidkonstruksjon eksemplifiseres i EP 1,158,051.

#### 10 Eksempel 8: Transformering av *P. rhodozyma* med SQS-antisense vektor

SQS-antisense vektoren således frembrakt transformeres inn i *P. rhodozyma* villtype stamme, ATCC96594 ved biolistisk transformering i henhold til protokollen beskrevet i EP 1,158,051.

#### Eksempel 9: Karakterisering av antisense SQS rekombinant av *P. rhodozyma*

15 Antisense SQS rekombinant av *P. rhodozyma*, ATCC96594 dyrkes i 50 ml YPD-medium i 500 ml Erlenmeyer flaske ved 20 °C i 3 dager ved anvendelse av deres såkulturet hvilket vokser i 10 ml YPD-medium i prøverør (21 mm diameter) ved 20 °C i 3 dager. For analyse av produsert karotenoid trekkes passende volum kulturbuljong ut og anvendes for analyse av deres vekst, produktivitet av karotenoider, særskilt astaxanthin. For analyse av vekst måles optisk tetthet ved 20 660 nm ved anvendelse av UV-1200 fotometer (Shimadzu Corp., Kyoto, Japan) i tillegg til bestemmelse av tørrvekt av celledensitet ved å tørke cellene avledet fra 1 ml buljong etter mikrosentrifugering ved 100 °C i en dag.

For analyse av innhold av astaxanthin og totale karotenoider, høstes celler fra 1,0 ml buljong etter mikrosentrifugering og anvendt for ekstraksjon av karotenoidene fra celler av *P. rhodozyma* ved oppløsning med glasskuler. Etter ekstraksjon fjernes oppløste celler ved sentrifugering og det resulterende analyseres for karotenoidinnhold med HPLC. HPLC-betingelsen anvendt er som følger: HPLC kolonne; Chrompack Lichrosorb si-60 (4,6 mm, 250 mm); temperatur; romtemperatur; eluent; aceton / heksan (18/82) tilsett 1 ml/l vann til eluent; injeksjonsvolum; 10 µl; strømningshastighet; 2,0 ml/minutt; deteksjon; UV ved 30 450 nm.

Referanseprøver kan oppnås fra Hoffmann La-Roche (Basel, Sveits) (særskilt astaxanthin) eller andre kommersielle leverandører (WAKO, SIGMA, osv.)

## P a t e n t k r a v

1. Rekombinant organisme hvis genekspresjon av squalene-syntase er redusert sammenlignet med vertsorganismen, som derved er i stand til å fremstille karotenoider i et økt nivå i forhold til vertsorganismen, karakterisert ved at den
  - 5 rekombinante organismen er en soppcelle som inneholder et antisenspolynukleotid mot et nukleinsyremolekyl valgt fra gruppen bestående av
    - (a) nukleinsyremolekyler som koder minst den modne formen av polypeptidet vist i SEKV. ID. NR.: 3;
    - (b) nukleinsyremolekyler omfattende kodesekvensen som vist i SEKV. ID. NR.: 2;
    - 10 (c) nukleinsyremolekyler hvis nukleotidsekvens er degenerert som et resultat av den genetiske koden til en nukleotidsekvens av (a) eller (b);
    - (e) nukleinsyremolekyler som koder et polypeptid avledet fra polypeptidet hvis sekvens har en identitet på 51.3 % eller mer til aminosyresekvensen av polypeptidet kodet av et nukleinsyremolekyl av (a) eller (b);
    - 15 (g) nukleinsyremolekyler omfattende et polynukleotid som har en sekvens av et nukleinsyremolekyl amplifisert fra et Phaffia nukleinsyrebibliotek som anvender primerene vist i SEKV. ID. NR.: 4, 5 og 6;
    - (j) nukleinsyremolekyler som koder et polypeptid som har squalenesyntaseaktivitet, hvor nevnte polypeptid er gjenkjent av antistoffer som har blitt dyrket mot et
      - 20 polypeptid kodet av et nukleinsyremolekyl ifølge ethvert av (a), (b), (c) og (g);
      - (k) nukleinsyremolekyler oppnåelige ved screening av et passende bibliotek under stringente betingelser med en sonde som har sekvensen av nukleinsyremolekylet ifølge ethvert av (a), (b), (c), (e), (g) og (j), og som koder et polypeptid som har squalenesyntaseaktivitet.
- 25 2. Rekombinant organisme hvis genekspresjon av squalene-syntase er redusert sammenlignet med vertsorganismen, som derved er i stand til å fremstille karotenoider i et økt nivå i forhold til vertsorganismen, karakterisert ved at den rekombinante organismen er en soppcelle som inneholder et antisenspolynukleotid mot et nukleinsyremolekyl valgt fra gruppen bestående av:

- (m) nukleinsyremolekyler omfattende nukleotidsekvensen som vist i SEKV. ID. NR.: 1;
- (n) nukleinsyremolekyler hvis nukleotidsekvens er degenerert som et resultat av den genetiske koden til en nukleotidsekvens av (m);
- 5 (p) nukleinsyremolekyler som koder et polypeptid avledet fra polypeptidet hvis sekvens har en identitet på 51.3 % eller mer til aminosyresekvensen av polypeptidet kodet av et nukleinsyremolekyl av (m);
- (q) nukleinsyremolekyler omfattende et fragment kodet av et nukleinsyremolekyl ifølge ethvert av (m), (n) eller (p) og som har squalenesyntaseaktivitet;
- 10 (r) nukleinsyremolekyler omfattende et polynukleotid som har en sekvens av et nukleinsyremolekyl amplifisert fra et Phaffia nukleinsyrebibliotek som anvender primerene vist i SEKV. ID. NR.: 4, 5 og 6;
- (s) nukleinsyremolekyler som koder et polypeptid som har squalenesyntaseaktivitet, hvor nevnte polypeptid er et fragment av et polypeptid
- 15 kodet ifølge ethvert av (m), (n), (p), (q) og (r);
- (t) nukleinsyremolekyler omfattende minst 15 nukleotider av et polynukleotid av (m) eller (a);
- (u) nukleinsyremolekyler som koder et polypeptid som har squalenesyntaseaktivitet, hvor nevnte polypeptid er gjenkjent av antistoffer som
- 20 har blitt dyrket mot et polypeptid kodet av et nukleinsyremolekyl ifølge ethvert av (m), (n), (p), (q), (r) og (s);
- (v) nukleinsyremolekyler oppnåelige ved screening av et passende bibliotek under stringente betingelser med en sonde som har sekvensen av nukleinsyremolekylet ifølge ethvert av (m), (n), (p), (q), (r), (s), (t) og (u), og som koder et polypeptid
- 25 som har squalenesyntaseaktivitet;
- (w) nukleinsyremolekyler hvis komplementære streng hybridiserer under stringente betingelser med et nukleinsyremolekyl ifølge ethvert av (m), (n), (p), (q), (r), (s), (t), (u), (v), og som koder et polypeptid som har squalenesyntaseaktivitet.

3. Rekombinant organisme ifølge krav 1 eller 2, inneholdende en rekombinant vektor som omfatter antisenspoly-nukleotidet.
4. Vektor ifølge krav 3, hvor antisenspolynukleotidet er operativt linket til ekspresjonskontrollsekvenser som tillater ekspresjon i soppceller.
- 5 5. Fremgangsmåte for å lage en rekombinant organisme omfattende innsetting av vektoren ifølge krav 4 i en vertsorganisme, hvor nevnte vertsorganisme hører til en stamme av *Phaffia rhodozyma* eller *Xanthophylomyces dend-rohous*.
- 10 6. Fremgangsmåte for å produsere karotenoider, som omfatter kultivering av den rekombinante organismen ifølge ethvert av krav 1 til 3, hvor nevnte karotenoider er valgt fra astaxanthin,  $\beta$ -caroten, lycopen, zeaxanthin, cantaxanthin.

## SEKVENSLISTE

<110> DSM IP ASSETS B.V.  
<120> SQS gene  
<130> NDR5218  
<140> PCT/EP03/10573  
<141> 2003-09-23  
<150> EP 02021619.8  
<151> 2002-09-27  
<160> 8  
<170> PatentIn version 3.2  
<210> 1  
<211> 4807  
<212> DNA  
<213> Phaffia rhodozyma  
  
<220>  
<221> 5'UTR  
<222> (1469) .. (1470)  
  
<220>  
<221> exon  
<222> (1550) .. (1577)  
  
<220>  
<221> Intron  
<222> (1578) .. (1752)  
  
<220>  
<221> exon  
<222> (1753) .. (1766)  
  
<220>  
<221> Intron  
<222> (1767) .. (1882)  
  
<220>  
<221> exon  
<222> (1883) .. (2071)  
  
<220>  
<221> Intron  
<222> (2072) .. (2182)  
  
<220>  
<221> exon  
<222> (2183) .. (2397)  
  
<220>  
<221> Intron  
<222> (2398) .. (2474)  
  
<220>  
<221> exon  
<222> (2475) .. (3087)

<220>  
 <221> Intron  
 <222> (3088)..(3230)

<220>  
 <221> exon  
 <222> (3231)..(3356)

<220>  
 <221> Intron  
 <222> (3357)..(3453)

<220>  
 <221> exon  
 <222> (3454)..(3475)

<220>  
 <221> Intron  
 <222> (3476)..(3564)

<220>  
 <221> exon  
 <222> (3565)..(3881)

<220>  
 <221> Intron  
 <222> (3882)..(3958)

<220>  
 <221> exon  
 <222> (3959)..(3970)

<220>  
 <221> polyA\_site  
 <222> (4106)..(4107)

<400> 1  
 gttcctgttc agtcaaagag tgggaaaaac atgaaagtaa aaagatgtaa tgaaagaagg 60  
 ggtcagaaca tcggagatac aatggcccat agaggaagga aagctactta ccagaaacca 120  
 gtgaggtttg cctaggaagt aatcccttcg tttctcaaag atatcttttt tgaaagcatc 180  
 gatgaacgac atgtcgaacc catctccatc ctcgaaatca agtttactcg atttagacct 240  
 ttccagcttt tctgctctct ccagtttcgc agctttctct tcgggaagaa gctctccgcc 300  
 agtcgatggt ctgtcgacag gagaccagta gaaggcggaa ccgacaattt tggatggatc 360  
 ggaggacagg gtggctttaa caaatcggta gtaocggagga tcgaacggcg cttctctcgt 420  
 tcgaaggttg actcctcttg ctatgtgtat gagagcatat ccgttgatgt ctcagttaaa 480  
 atttcctttt ctttctaccc ggagagtaag acacacaaag aatcacgaag aatatgatga 540  
 ctgaccgatc cgaatatcta gcgcagggtg cttctctact ggttccattc ttcgaacgat 600  
 aggttcatgt ttgaaagcat tgatcctagt tgccctatc tgaggccagt ctgccaatgt 660  
 agcaggctca atgatcactt ggggtttgtg catcttgatg ttcaaccaag tgtcgcaacg 720  
 gtcgagattc tttttcttc ttttggtcga gaaaaaaaaa cggcttcgct tcgcacgcgc 780  
 gcgggggatc acccgcatat taagcgggat gacgctcatc aaccggccaa gtgttcttca 840

tcataggtga aggttaaac ggaatggata ggaggagcta accacgtttt tattttaatt	900
cgacttgggc agcctcgtcc atagtgtctg atggttatat cgatcatagaa aggcagcgc	960
tggcgggttc gtcattggccg tgatcatctg ctttggtaga cattgtccat cagtcacctc	1020
aatgacagtt tcccagcgc atcactaaga cacaaacgta tccagcagc catgtccatc	1080
actgaagaag gtagggctc gtcgagccag tgcaaccaga gttacagatg aacatcaggc	1140
cttgatcaga cccgacttat gaatatggcc gttattgtac acttcttggg gtcctcag	1200
ctgctcttcc gtgtttttca ctttcttcc ggatcaaacy agactgctcg tgtatctatc	1260
tgtgcttgcc atatgagcat cccatgcctc tgctcaaag atgctggagc tacgatccat	1320
cagagacgac acaaaacggg gttgtatgaa ctctacattt cctaagtta ttggaatttt	1380
ctgtaatgcg ttcttcatct ttctctaag cttttttgta gtccgtcttt tcaacctg	1440
cagcgtttcg cgtgtcttct ttctccttg acggatca ctttcttctc tcttctcgtt	1500
cttctctcgg tcttctctc tctctctcgg tctgaacatc agcatc atg ggc ata	1558
	Met Gly Ile
	1
tca gat tac ctg gtt ctg g gtcagttctg tcttttgttt gattcttacc	1607
Ser Asp Tyr Leu Val Leu	
5	
ttcttgccgg cggctgcctg tcttgggtat atcatcagca atgagaaaca tgatgttccc	1667
cccgcgtcaa tcaactgacct tttggctctc tacttcttcc ctgtcgaatt gatcctgatt	1727
gatacgtgtg ccggctgctt aacag ct ttc acg cat cct gtaggtgttt	1776
	Ala Phe Thr His Pro
tatcgtatgc ttcattgtga tgtttagtca cgcggactga cctggccggg tgattttctg	1836
tatgatcgtc tgtgctaccg tcttcttggg aaatccttcc catcag gcc gat ctg	1891
	Ala Asp Leu
	15
cga gct tta atg cag tac gcg atc tgg cat gag cct cga agg aat atc	1939
Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn Ile	
20 25 30	
act gca cag gag gaa cat gca aca tcc ggt tgg gac cga gaa act atg	1987
Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr Met	
35 40 45	
aag gaa tgt tgg aag tat ttg gat ctg act tca aga agt ttc gca gct	2035
Lys Glu Cys Trp Lys Tyr Leu Asp Leu Thr Ser Arg Ser Phe Ala Ala	
50 55 60 65	
gtc atc aaa gag ttg gac gga gat ctt acc cga gtc gtacgtgttt	2081
Val Ile Lys Glu Leu Asp Gly Asp Leu Thr Arg Val	
70 75	
tcatcttctc tctccttga gatctggctg cctccgcatt ttcttgttgc agaagggtca	2141
gaagctgaca acaccatctc tactgttccg gacacggcta g atc tgt tta ttc tat	2197



His Thr Thr Asp Ala Leu Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser  
 300 305 310  
 gtt ttc aac ttt tgt gct atc ccg gct gtc atg tcg att gca acg ttg 3018  
 Val Phe Asn Phe Cys Ala Ile Pro Ala Val Met Ser Ile Ala Thr Leu  
 315 320 325 330  
 gag cta tgc ttc atg aac cca gcg gtg ttc caa cga aac ata aaa atc 3066  
 Glu Leu Cys Phe Met Asn Pro Ala Val Phe Gln Arg Asn Ile Lys Ile  
 335 340 345  
 aga aag gga gaa gcc gtc gag gtgcggttcgc gcgtttctggtt tctacctttc 3117  
 Arg Lys Gly Glu Ala Val Glu  
 350  
 ataacattgg aggttcttga ctcttaagcg tcttccaatc tgatgcctcc aattatcatc 3177  
 atttttgtct tttttgcttt cctcttgttt cttttcggcg tgattcaatc cag ctc 3233  
 Leu  
 att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt aga gat 3281  
 Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe Arg Asp  
 355 360 365 370  
 tat gct cga aag att cat gcc aag gct att cct aca gat cct aac ttc 3329  
 Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro Asn Phe  
 375 380 385  
 atc aag ttg agc gtt gcg tgt ggt cga gtgagttgat cgatcgatcc 3376  
 Ile Lys Leu Ser Val Ala Cys Gly Arg  
 390 395  
 atcttttgtt ttgatcatcg cgagacttga ctgatcgatt actcaaaaca tcatcgcttc 3436  
 tccttcttgc tctctag atc gaa caa tgg gct gag cac t gtatgttcct 3485  
 Ile Glu Gln Trp Ala Glu His  
 400  
 ccgccccctcc ttcaagtttc ctctcgcttc atctttgttg agaagagggga tctgatgtat 3545  
 ctttctttgt tcggatcag ac tac ccc tca ttt atg atg att cgg cct tcg 3596  
 Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser  
 405 410  
 aat gac cct caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca 3644  
 Asn Asp Pro Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser  
 415 420 425  
 gga gac gct cgt tta agg ata gcc tct aag aag gct gag atc acc gcc 3692  
 Gly Asp Ala Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala  
 430 435 440 445  
 gct gct ctt gtc agg aag aaa gcc gcg gat cac gct aag tgg aga gag 3740  
 Ala Ala Leu Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu  
 450 455 460  
 tcc aag gga ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag 3788  
 Ser Lys Gly Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu  
 465 470 475  
 gat gtt aat tgg gta ttg atc ggc ggt atg atc gtt gga ttg ttg ctc 3836  
 Asp Val Asn Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu  
 480 485 490

```

gtg atg ggc gtg ctc ggt ttg gct atc gct tgg gtt gtt ctt cag      3881
Val Met Gly Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln
    495                      500                      505

gtgCGTtctt ccaaGagcc tttctctcat gaacacgcac ataggttgat ctaattctat  3941

cttactctgt catacag ttt gag caa taa tctcaagatt ctagtccatc      3990
                Phe Glu Gln
                510

ctttcgctca acgatctgct ttttctcctt ctctttctcc gtctttctctg gtttcttttc  4050

ttactttctg ggatcttctt ttttgaatcc tccgatccaa tgtaatctgc ataccctcgc  4110

tttagtagaa accgatcctt cattcgatct tggcgaaaat ctaagcaaag agaatcactt  4170

ttgtctaata aaatttctt taaagagtcg gctttttctt gtggcgaagc ttcatcccgt  4230

cttctctgg accatctctt ctcaatattc tttgtgctac tatatgatca agttctttga  4290

aatcaaagaa gaacatgtat ttgattttga ggttccaaga atacaaccgg cccaagtcgt  4350

tcttgcagt tttcatcaga cagcacatat ctctctctct ctctatagaa gccgtatggg  4410

gccaatcgac tctcatgggt agaccgtgcc cttttgacac ggggagaaag agaacgaaag  4470

gacacttgac cgattcgtta ataaagccgt cccacctttt tctttaatgg caattcaaga  4530

agagaaaaac aaccctcgcg cgcactcgag tagtcgatca gaccttccga acgacagata  4590

tcatttgctg aaatcgaccg gattttaaag ctgctgccag gtcggtgaat ccccttaggt  4650

gatctccttg taaaagatg ttgggcacgg acttttcgac ccggatgaga acgtcgtgaa  4710

gagtttgaaa aagattatca acataatgtg tctttttttc ttttttcttt cgtaactctc  4770

tagagaacga ggagacgtac ggtctgattt gttatcg                          4807

```

```

<210> 2
<211> 1536
<212> DNA
<213> Phaffia rhodozyma

```

```

<220>
<221> CDS
<222> (1)..(1536)

```

```

<400> 2
atg ggc ata tca gat tac ctc gtt ctg gct ttc acg cat cct gcc gat      48
Met Gly Ile Ser Asp Tyr Leu Val Leu Ala Phe Thr His Pro Ala Asp
1          5          10          15

ctg cga gct tta atg cag tac gcg atc tgg cat gag cct cga agg aat      96
Leu Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn
                20          25          30

atc act gca cag gag gaa cat gca aca tcc ggt tgg gac cga gaa act      144
Ile Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr
                35          40          45

atg aag gaa tgt tgg aag tat ttg gat ctg act tca aga agt ttc gca      192

```

Met	Lys	Glu	Cys	Trp	Lys	Tyr	Leu	Asp	Leu	Thr	Ser	Arg	Ser	Phe	Ala	
	50					55					60					
gct	gtc	atc	aaa	gag	ttg	gac	gga	gat	ctt	acc	cga	gtc	atc	tgt	tta	240
Ala	Val	Ile	Lys	Glu	Leu	Asp	Gly	Asp	Leu	Thr	Arg	Val	Ile	Cys	Leu	
65					70					75					80	
ttc	tat	ctc	gct	ctt	cga	gga	ctg	gat	acc	att	gag	gat	gac	atg	agt	288
Phe	Tyr	Leu	Ala	Leu	Arg	Gly	Leu	Asp	Thr	Ile	Glu	Asp	Asp	Met	Ser	
				85					90					95		
cta	tct	aat	gat	gtg	aag	ctt	ccc	ctg	ctt	cgg	aca	ttc	tgg	gaa	aag	336
Leu	Ser	Asn	Asp	Val	Lys	Leu	Pro	Leu	Leu	Arg	Thr	Phe	Trp	Glu	Lys	
			100					105					110			
ctt	gac	tcc	cct	ggg	tgg	acc	ttt	act	gga	tcc	ggt	cca	aat	gag	aag	384
Leu	Asp	Ser	Pro	Gly	Trp	Thr	Phe	Thr	Gly	Ser	Gly	Pro	Asn	Glu	Lys	
		115					120					125				
gat	aga	gag	ctt	ctt	gtt	cac	ttc	gat	gtg	gcc	atc	gcc	gag	ttt	gcc	432
Asp	Arg	Glu	Leu	Leu	Val	His	Phe	Asp	Val	Ala	Ile	Ala	Glu	Phe	Ala	
	130					135					140					
aac	ttg	gac	gtc	aac	tct	cgg	aac	gtc	att	cga	gac	atc	act	cgc	aag	480
Asn	Leu	Asp	Val	Asn	Ser	Arg	Asn	Val	Ile	Arg	Asp	Ile	Thr	Arg	Lys	
145					150					155					160	
atg	ggt	aac	ggt	atg	gcc	gac	ttt	gct	tct	ctc	tct	acg	ccc	tcc	aag	528
Met	Gly	Asn	Gly	Met	Ala	Asp	Phe	Ala	Ser	Leu	Ser	Thr	Pro	Ser	Lys	
				165					170					175		
cct	gtg	gcc	gag	gtc	cag	tcg	acc	gaa	gat	ttc	aac	cta	tac	tgt	cat	576
Pro	Val	Ala	Glu	Val	Gln	Ser	Thr	Glu	Asp	Phe	Asn	Leu	Tyr	Cys	His	
			180						185					190		
tac	gtc	gct	gga	ctc	gtc	ggc	gag	gga	ctc	tcc	cga	ctc	ttt	gtc	gcg	624
Tyr	Val	Ala	Gly	Leu	Val	Gly	Glu	Gly	Leu	Ser	Arg	Leu	Phe	Val	Ala	
		195					200						205			
acc	gag	aag	gaa	cga	cca	ttc	ttg	gcc	aac	cag	atg	gta	ctt	tca	aac	672
Thr	Glu	Lys	Glu	Arg	Pro	Phe	Leu	Ala	Asn	Gln	Met	Val	Leu	Ser	Asn	
	210					215					220					
tcg	ttc	gga	ctc	ctt	ctc	caa	aag	aca	aac	atc	ctt	cga	gat	att	cg	720
Ser	Phe	Gly	Leu	Leu	Leu	Gln	Lys	Thr	Asn	Ile	Leu	Arg	Asp	Ile	Arg	
225					230					235					240	
gag	gac	gcc	gac	gaa	ggt	cgt	ggc	ttc	tgg	cca	aga	gag	atc	tgg	gcc	768
Glu	Asp	Ala	Asp	Glu	Gly	Arg	Gly	Phe	Trp	Pro	Arg	Glu	Ile	Trp	Ala	
				245					250					255		
aac	ccg	atc	tat	act	gcg	cat	gca	ccg	ggc	aca	agg	ttt	aac	tcg	ttg	816
Asn	Pro	Ile	Tyr	Thr	Ala	His	Ala	Pro	Gly	Thr	Arg	Phe	Asn	Ser	Leu	
			260					265					270			
act	gac	ctg	gtc	aag	aaa	gaa	aac	atc	gac	aaa	gga	tca	atg	tgg	gtg	864
Thr	Asp	Leu	Val	Lys	Lys	Glu	Asn	Ile	Asp	Lys	Gly	Ser	Met	Trp	Val	
		275					280					285				
ttg	agt	gcg	atg	aca	ctc	gac	gcg	atc	acc	cat	act	acc	gac	gca	ctg	912
Leu	Ser	Ala	Met	Thr	Leu	Asp	Ala	Ile	Thr	His	Thr	Thr	Asp	Ala	Leu	
	290					295					300					
gac	tac	ctc	tca	ctt	cta	aag	aac	cag	agt	gtt	ttc	aac	ttt	tgt	gct	960

Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser Val Phe Asn Phe Cys Ala  
 305 310 315 320  
 atc ccg gct gtc atg tcg att gca acg ttg gag cta tgc ttc atg aac 1008  
 Ile Pro Ala Val Met Ser Ile Ala Thr Leu Glu Leu Cys Phe Met Asn  
 325 330 335  
 cca gcg gtg ttc caa cga aac ata aaa atc aga aag gga gaa gcc gtc 1056  
 Pro Ala Val Phe Gln Arg Asn Ile Lys Ile Arg Lys Gly Glu Ala Val  
 340 345 350  
 gag ctc att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt 1104  
 Glu Leu Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe  
 355 360 365  
 aga gat tat gct cga aag att cat gcc aag gct att cct aca gat cct 1152  
 Arg Asp Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro  
 370 375 380  
 aac ttc atc aag ttg agc gtt gcg tgt ggt cga atc gaa caa tgg gct 1200  
 Asn Phe Ile Lys Leu Ser Val Ala Cys Gly Arg Ile Glu Gln Trp Ala  
 385 390 395 400  
 gag cac tac tac ccc tca ttt atg atg att cgg cct tcg aat gac cct 1248  
 Glu His Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser Asn Asp Pro  
 405 410 415  
 caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca gga gac gct 1296  
 Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser Gly Asp Ala  
 420 425 430  
 cgt tta agg ata gcc tct aag aag gct gag atc acc gcc gct gct ctt 1344  
 Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala Ala Ala Leu  
 435 440 445  
 gtc agg aag aaa gcc cgg gat cac gct aag tgg aga gag tcc aag gga 1392  
 Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu Ser Lys Gly  
 450 455 460  
 ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag gat gtt aat 1440  
 Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu Asp Val Asn  
 465 470 475 480  
 tgg gta ttg atc ggc ggt atg atc gtt gga ttg ttg ctc gtg atg ggc 1488  
 Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu Val Met Gly  
 485 490 495  
 gtg ctc ggt ttg gct atc gct tgg gtt gtt ctt cag ttt gag caa taa 1536  
 Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln Phe Glu Gln  
 500 505 510

<210> 3  
 <211> 511  
 <212> PRT  
 <213> Phaffia rhodozyma

<400> 3

Met Gly Ile Ser Asp Tyr Leu Val Leu Ala Phe Thr His Pro Ala Asp  
 1 5 10 15

Leu Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn





<220>  
<223> Primer

<220>  
<221> misc\_feature  
<222> (3)..(3)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (4)..(4)  
<223> y = c or t

<220>  
<221> misc\_feature  
<222> (6)..(6)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (12)..(12)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (15)..(15)  
<223> n = a, c, g or t

<400> 4  
gcnytngaya cngtngarga ygayatg

27

<210> 5  
<211> 26  
<212> DNA  
<213> Artificial

<220>  
<223> Primer

<220>  
<221> misc\_feature  
<222> (3)..(3)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (9)..(9)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (15)..(15)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (18)..(18)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature

<222> (21)..(21)  
<223> n = a, c, g or t

<400> 5  
atngccatna cytgnggnat ngrca

26

<210> 6  
<211> 29  
<212> DNA  
<213> Artificial

<220>  
<223> Primer

<220>  
<221> misc\_feature  
<222> (3)..(3)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (6)..(6)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (9)..(9)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (12)..(12)  
<223> n = a, c, g or t

<220>  
<221> misc\_feature  
<222> (15)..(15)  
<223> n = a, c, g or t

<400> 6  
ccnacngtnc cngcnacrta rtgrcarta

29

<210> 7  
<211> 20  
<212> DNA  
<213> Artificial

<220>  
<223> Primer

<400> 7  
aatgatgtga agcttcccct

20

<210> 8  
<211> 20  
<212> DNA  
<213> Artificial

<220>  
<223> Primer

<400> 8  
ccagatctct cttggccaga

20