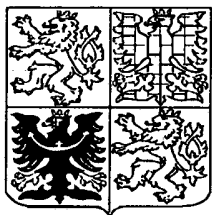


ČESKÁ
REPUBLIKA

(19)



ÚŘAD
PRŮMYSLVÉHO
VLASTNICTVÍ

ZVEŘEJNĚNÁ PŘIHLÁŠKA VYNÁLEZU

(12)

(22) 18.12.91
(32) 18.12.91, 19.12.90
(31) 91DK/9100396, 90/3000
(33) WO, DK
(40) 16.02.94

(21) 1192-93

(13) A3

5(51)

C 12 N 15/81
C 12 N 15/62
C 12 N 1/19

(71) Novo Nordisk A/S, Bagsvaerd, DK;
Christiansen Lars, Lyngby, DK;

(72) Christiansen Lars, Lyngby, DK;

(54) Způsob konstrukce syntetické vedoucí peptidové
sekvence

(57) Kvasinkový expresní clonovací vektor zahrnující sekvenci 5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS- - gen -3', kde znamená SP DNA sekvenci kodující signální peptid, X_n DNA sekvenci kodující n aminokyselin, přičemž n znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin, RS restriční endonukleasové rekognizační místo pro včlenění statistických DNA fragmentů, kteréžto místo je na spojení X_n a X_m, X_m DNA sekvenci kodující m aminokyselin, přičemž m znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin, (NZT)_p DNA sekvenci kodující Asn-Xaa-Thr, přičemž m znamená 0 nebo číslo 1, X_q DNA sekvenci kodující q aminokyselin, přičemž q znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin, PS DNA sekvenci kodující peptid definující kvasinkové zpracovatelské místo a gen sekvenci kodující heterologický polypeptid, je použitelný pro konstrukci syntetických peptidových sekvencí včleňováním statistických DNA fragmentů do "RS" místa, kultivací kvasničné buňky transformované vektorem a vyříděním kultury pro sekreci heterologického polypeptidu.

Oblast techniky

Dosavadní stav techniky

Byly navrženy různé způsoby exprese a vylučování v kvasinkách bílkovin heterologických ke kvasinkám. V evropské zveřejněné přihlášce vynálezu číslo 88 632 se popisuje způsob, kterým dochází v případě bílkovin heterologických ke kvasinkám, k expresi, zpracování a vylučování transformací kvasinkového organismu expresivním pojivem kotvicím DNA kodující žádanou bílkovinu a signální peptid, k přípravě kultury transformovaného organismu, k růstu kultury a k získání bílkoviny z kultivačního prostředí. Signální peptid může být signálním peptidem samotného žádaného peptidu, heterologickým signálním peptidem nebo hybridem nativního a heterologového signálního peptidu.

Problémem při použití signálních peptidů, heterologických ke kvasinkám, může být skutečnost, že heterologický signální peptid nezajišťuje účinnou translokaci a/nebo odštěpení po signálním peptidu.

S. cerevisiae MF α 1 (α -faktor) se syntetizuje jako preproforma 165 aminokyselin obsahující signální nebo prepeptid s 19 aminokyselinami následovaný "vedoucím" nebo propeptidem se 64 aminokyselinami majícím tři N-vázaná glykosylační místa následovaná (LysArg(Asp/Glu, Ala)₂₋₃ α -faktor)₄ (J. Kurjan a I. Herskowitz, Cell 30, str. 933 až 943, 1982). Signální vedoucí podíl preproMF α 1 se v široké míře používá k dosažení syntézy a sekrece heterologické bílkoviny v *S. cerevisiae*.

Použití signálních/vedoucích peptidů homologických ke kvasinkám je o sobě známe například z amerického patentového spisu číslo 4 546082, z evropské zveřejněné přihlášky vynálezu číslo 116201, 123294, 123544, 163529 a 123289 a z dánského zveřejněné přihlášky vynálezu číslo 3614/83.

Podle evropské přihlášky vynálezu číslo 123289 se používá *S. cerevisiae* α -faktorového prekursoru, přičemž podle světového patentového spisu WO 84/01153 se naznačuje použití *Saccharomyces cerevisiae* invertazového signálního peptidu a podle zveřejněné dánské přihlášky vynálezu číslo 3614/83 se používá *Saccharomyces cerevisiae* PHO5 signálního peptidu pro sekreci cizích bílkovin.

V americkém patentovém spise číslo 4 546082 a v evropských patentových spisech číslo 16201, 123294, 123544 a 163529 se popisuje způsob, podle kterého se α -faktorový signální vedoucí prvek ("signal-leader") ze *Saccharomyces cerevisiae* (MF α 1 nebo MF α 2) používá v sekrečním procesu vytlačených heterologických bílkovin v kvasinkách. Popisuje se fuze DNA sekvence kodující *Saccharomyces cerevisiae* MF α 1 signální vedoucí sekvenci na zakončení 5' genu pro žádanou sekreci bílkoviny a zpracování žádané bílkoviny.

Evropský patentový spis číslo 206783 popisuje systém pro sekreci polypeptidů ze *Saccharomyces cerevisiae*, přičemž α -faktorová vedoucí sekvence je seřiznuta k eliminaci čtyř α -faktorových peptidů obsažených na nativní vedoucí sekvenci, takže se uvolňuje

vedoucí peptid sám anelovaný na heterologický polypeptid prostřednictvím α -faktorového zpracovatelského místa LysArgGluAla-GluAla. Uvádí se, že tato konstrukce vede k účinnému způsobu přípravy menších peptidů (obsahujících méně než 50 aminokyselin). Pro sekreci a zpracování větších polypeptidů se nativní α -faktorová vedoucí sekvence seřízne k uvolnění jednoho nebo dvou α -faktorových peptidů mezi vedoucím peptidem a polypeptidem.

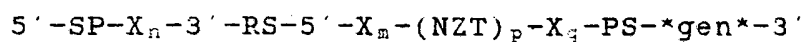
Počet sekretovaných bílkovin se směřuje tak, aby byly vystaveny proteolytickému zpracovatelskému systému, který může odštěpit peptidovou vazbu na karboxylovém zakončení dvou následujících bazických aminokyselin. Tato enzymatická aktivita je v *Saccharomyces cerevisiae* zakodována KEX 2 genem (D. A. Julius a kol., Cell 37, str. 1075, 1984b). Zpracování produktu KEX 2 genovým produktem je nutné pro sekreci aktivního *Saccharomyces cerevisiae* faktoru $\alpha 1$ (MF $\alpha 1$ nebo α -faktor) není však zahrnut v sekreci aktivního *S.cerevisiae* faktoru α .

Sekrece a správné zpracování polypeptidu, který má být výsledkem sekrece, se v některých případech dosahuje, jestliže se kultivují kvasinkové organismy, které jsou transformovány vektorem, konstruovaným jak shora uvedeno. V mnoha případech však hladina sekrece je velmi nízká nebo k sekreci vůbec nedochází nebo proteolytické zpracování může být nesprávné nebo nekompletní. Záměrem vynálezu je proto vytvořit vedoucí peptidy, které zajišťují účinnější expresi a/nebo zpracování heterologových polypeptidů.

Podstata vynálezu

Podstata způsobu konstrukce syntetické vedoucí peptidové sekvence pro sekreci heterologových polypeptidů v kvasinkách podle vynálezu spočívá v tom, že se

- a) vnáší statistický DNA fragment do kvasinkového expresního vektoru zahrnujícího sekvenci



kde znamená

SP DNA sekvenci kodující signální peptid

X_n DNA sekvenci kodující n aminokyselin, přičemž n zna-

- mená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
 RS restrikční endonukleové rekognizační místo pro včle-
 nění statistických DNA fragmentů, kteréžto místo je
 na spojení X_n a X_m ,
 X_m DNA sekvenci kodující m aminokyselin, přičemž m zna-
 mená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
 (NZT)_p DNA sekvenci kodující Asn-Xaa-Thr, přičemž m znamená
 0 nebo číslo 1,
 X_q DNA sekvenci kodující q aminokyselin, přičemž q zna-
 mená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
 PS DNA sekvenci kodující peptid definující kvasinkové
 zpracovatelské místo a

gen sekvenci kodující heterologický polypeptid,

- b) transformuje se kvasinková hostitelská buňka expresním vektor-
 em podle stupně (a),
- c) kultivuje se transformovaná hostitelská buňka podle stupně (b)
 za vhodných podmínek a
- d) vytřídí se kultura podle stupně (c) pro sekreci heterologové-
 ho polypeptidu.

S překvapením se totiž zjistilo, že je možno přesunout α -fak-
 torový vedoucí peptid mnoha různými DNA sekvencemi, čímž se do-
 sáhne sekrece heterologického polypeptidu v kvasinkách. Na zákla-
 dě tohoto objevu byl vyvinut způsob, kterým se statistické DNA
 fragmenty klonují do kvasinkových vektorů ve směru DNA sekvence
 kodující pro signální peptid a proti směru DNA sekvence kodující
 pro heterologový polypeptid. Po transformaci vektory se kvasinko-
 vé buňky vytřídí pro sekreci žádaného heterologického polypeptidu.

Vynález se proto týká především shora definovaného způsobu
 konstrukce syntetické vedoucí peptidové sekvence pro sekreci he-
 terologických polypeptidů v kvasinkách.

Výrazem "vedoucí peptid" se zde vždy míní peptid, jehož úlo-
 hou je umožnit řízení heterologického polypeptidu z endoplasmové-
 ho retikula na Golgi aparaturu a dále k sekretačnímu pojídlu pro
 sekreci do prostředí (to znamená export vytlačeného polypeptidu
 přes buněčnou stěnu nebo alespoň přes celulární membránu do pe-
 risplamové oblasti buňky). Výrazem "syntetický" v souvislosti

s vedoucími peptidy se zde vždy míní, že vedoucí peptid, konstruovaný podle vynálezu, nebyl v přírodě nalezen.

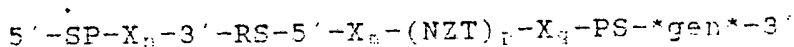
Výrazem "signální peptid" se zde vždy míní presekvence, která je převážně hydrofobní povahy a je obsažena jakožto N-zakončující sekvence v kvasinkách. Funkcí signálního peptidu je umožnit sekreci heterologické bílkoviny do endoplasmového retikula. Signální peptid se zpravidla v průběhu tohoto procesu odštěpí. Signální peptid může být heterologický nebo homologický s kvasinkovým organismem produkujícím bílkovinu, jak shora vysvětleno. Mnohem účinnějšího odštěpení signálního peptidu se však dosahuje, jestliže je homologický s příslušným kvasinkovým organismem.

Výrazem "heterologický polypeptid" se míní polypeptid, který se neprodukuje přírodním hostitelským kvasinkovým organismem. Při způsobu podle vynálezu je heterologickým polypeptidem s výhodou polypeptid, jehož sekrece transformovanými kvasinkovými buňkami se může snadno zjistit, například ustavením standardních způsobů, jako je imunologické vyřídění použitím protilátek reaktivních s příslušným polypeptidem (viz například Sambrook, Fritsch a Maniatis, Molecular Cloning: A Laboratory Manual (Molekulární klonování: Laboratorní příručka), Cold Spring Harbor, New York, 1989) nebo vyříděním pro specifickou biologickou účinnost heterologického peptidu. Pozitivní výsledek vyřídění indikuje, že byl konstruován vedoucí peptid, užitečný pro sekreci heterologického polypeptidu v kvasinkách.

Výrazem "statistický DNA fragment" se míní jakákoliv sekvence DNA o délce alespoň tří nukleotidů, například získaná digescí genomové DNA (jakéhokoliv organismu) s restriční endonukleasou nebo s restričními endonukleásami nebo přípravou syntetické DNA, například fosfoamidátovým způsobem, který popsal S.L. Beaucage a H.H. Caruthers, Tetrahedron Letters 22, str. 1859 až 1869, 1981.

Peptid Asn-Xaa-Thr kodovaný "(NXT)", je asparaginem vázané glykosylační místo. "Xaa" znamená kteroukoliv ze známých aminokyselin kromě prolinu.

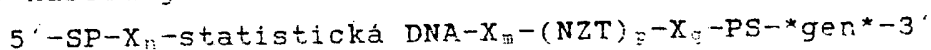
Vynález se také týká kvasinkového expresního klonovacího vektoru zahrnujícího následující sekvenci



kde jednotlivé symboly mají shora uvedený význam.

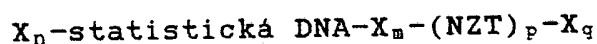
Tohoto vektoru se může používat pro způsob konstrukce syntetické vedoucí peptidové sekvence shora popsaným způsobem.

Vynález se také týká kvasinkového expresního vektorem zahrnujícího následující sekvenci



kde znamená výraz "statistická DNA" fragment včleněný do restričního endonukleosového rekognizačního místa ve spojení X_n a X_m a ostatní symboly mají shora uvedený význam.

V tomto vektoru vedoucí peptidová sekvence (identifikovaná způsobem podle vynálezu) je složen ze sekvence



kde jednotlivé symboly mají shora uvedený význam. Takového vektoru se může používat pro produkti žádaného heterologického polypeptidu.

Vynález se také týká způsobu přípravy heterologického polypeptidu v kvasinkách, přičemž se kultivuje kvasinková buňka, která je schopna exprese heterologického polypeptidu a která se transformuje kvasinkovým expresním vektorem, jak shora uvedeno, včetně vedoucí peptidové sekvence konstruované způsobem podle vynálezu ve vhodném prostředí k získání exprese a sekrece heterologického polypeptidu, přičemž se nakonec heterologický peptid získá z prostředí.

Délka statistického DNA fragmentu, včleněného do expresního vektoru, nemá rozhodující význam. Aby však byla délka vhodná pro manipulaci, má fragment s výhodou délku 16 až 600 páru bází. Především má fragment délku přibližně 15 až přibližně 300 párů bází. V současné době se míní, že vhodná délka fragmentu je přibližně 30 až přibližně 150 páru bází.

Statistický DNA fragment s výhodou koduje vysoký podíl polárních aminokyselin. Tyto aminokyseliny jsou voleny ze souboru zahrnujícího Glu, Asp, Arg, His, Thr, Ser, Asn a Gln. V této souvislosti se výrazem "vysoký podíl" míní, že DNA fragment koduje větší počet polárních aminokyselin než mají schopnost jiné DNA sekvence odpovídající délky. Nezávisle na tom nebo nad to mů-

že být výhodné, aby fragment kodoval alespoň jeden prolin.

V sekvenci

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

znamená n a/nebo m a/nebo q s výhodou 1 nebo větší číslo než 1. Především všechny symboly n, m a q znamenají 1 nebo větší číslo než 1 a ostatní symboly mají shora uvedený význam.

Jsou některé skutečnosti (viz světový patentový spis WO 89/02463), které potvrzují, že přítomnost na asparagin vázaného glykosylačního místa ve vedoucí sekvenci může podporovat sekreční účinnost vedoucího peptidu. V sekvenci

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

znamená prosto symbol p s výhodou 1, přičemž ostatní symboly mají shora uvedený význam.

Signální peptidová sekvence (SP) může kodovat jakýkoliv signální peptid, který zajišťuje účinné řízení vytlačovaného heterologického polypeptidu do sekreční cesty buňky. Signálním peptidem může být přirodně se vyskytující signální peptid nebo jeho funkční podíl nebo to může být syntetický peptid. Vhodnými signálními peptidy jsou α-faktorové signální peptidy, signální peptid myších slinných žláz, modifikovaný karboxypeptidasový signální peptid, signální peptid kvasinek BAR1 nebo lipasový signální peptid *Humicola lanuginosa* nebo jeho derivát. Amylasová signální sekvence myšších slinných žláz je popsána v publikaci O. Hagenbüchle a kol., *Nature* 289, str. 643 až 646, 1981. Karboxypeptidasová signální sekvence je popsána v publikaci L.A. Valls a kol., *Cell* 48, str. 887 až 897, 1987. Signální peptid kvasinek BAR1 je popsán ve světovém patentovém spise WO 87/02670. Lipasový signální peptid *Humicola lanuginosa* je popsán v evropském patentovém spise číslo EP 305 216.

Kvasinkovým zpracovatelským místem, kodovaným DNA sekvencí PS může být vhodně jakákoliv párová kombinace Lys a Arg, například Lys-Arg, Arg-Lys, Lys-Lys nebo Arg-Arg, která umožňuje výrobu heterologického polypeptidu KEX2 proteasou *Saccharomyces cerevisiae* nebo ekvivalentní proteasou jiného druhu kvasinek (D.A. Julius a kol., *Cell* 37, od str. 1075, 1984). Pokud není KEX2 proces vhodný, například se má dosáhnout štěpení polypeptidového produk-

tu, má se volit zpracovatelské místo pro jinou proteasu místo zahrnutí aminokyselinové kombinace, která není zjištěna v polypeptidovém produktu, například procesní místo pro FX_a , Ile-Glu-Gly-Arg (jak popisuje Sambrook, Fritsch a Maniatis, Molecular Cloning: A Laboratory Manual (Molekulární clonování: Laboratorní příručka), Cold Spring Harbor, New York, 1989).

Heterologickou bílkovinou, produkovanou způsobem podle vynálezu, může být jakákoliv bílkovina, která se výhodně produkuje kvasinkami. Jakožto příklady takových bílkovin se uvádějí aprotinin, tkáňový faktorový inhibitor nebo jiné proteasové inhibitory, insulin nebo insulinové prekursory, lidský nebo hovězí růstový faktor, interleukin, glukagon, tkáňový plasminogenový aktivátor, transformační růstový faktor α nebo β , růstový faktor odvozený od krevních destiček, enzymy nebo jejich funkční analogy. Výrazem "funkční analog" se zde vždy míní polypeptid s podobnou funkcí jako má nativní bílkovina (míní se vztah k přírodě spíše než hladina biologické účinnosti nativní bílkoviny). Polypeptid může být strukturálně podobný nativní bílkovině a může být odvozen od nativní bílkoviny přidáním jedné nebo několika aminokyselin buď na jedno nebo na obě C a N-zakončení nativní bílkoviny, náhradou jedné nebo několika aminokyselin na jednom nebo na četných odlišných místech v nativní aminokyselinové sekvenci, vypuštěním jedné nebo několika aminokyselin na jednom nebo na obou zakončeních nativní bílkoviny na jednom nebo na několika místech v aminokyselinové sekvenci nebo vložением jedné nebo několika aminokyselin na jednom nebo na několika místech nativní aminokyselinové sekvence. Takové modifikace jsou dobře známé pro četné shora uvedené bílkoviny.

Statistický DNA fragment a sekvence

$5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'$

kde jednotlivé symboly mají shora uvedený význam, se mohou připravovat synteticky o sobě známými způsoby například fosfoamidátovým způsobem, který popsali S.L. Beaucage a M.H. Caruthers, Tetrahedron Letters 22, str. 1859 až 1869, 1981 nebo způsobem, který popsal Matthes a kol., EMBO Journal 3, str. 801 až 805, 1984. Fosfoamidátovým způsobem se syntetizují oligonukleotidy, napřík-

lad v automatickém DNA syntetizeru, čistí se, tepelně se hybridizují, ligatují a clonují do kvasinkového expresního vektoru. Připomíná se, že se sekvence

$$5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'$$

kde jednotlivé symboly mají shora uvedený význam, nemusí připravovat jedinou operaci, ale může se vytvářet ze dvou nebo z několika oligonukleotidů, připravených synteticky takovým způsobem.

Statistický DNA fragment nebo jedna nebo několik částí sekvence

$$5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'$$

kde jednotlivé symboly mají shora uvedený význam, mohou být také genomického nebo cDNA původu, například získané přípravou genomického nebo cDNA souboru ("library") a vytríděním pro DNA sekvence kodující pro uvedené páry (zpravidla SP nebo *gen*) hybridizací za použití syntetických oligonukleotidových sond spolu se standardními způsoby (Sambrook, Fritsch a Maniatis, Molecular Cloning: A Laboratory Manual (Molekulární clonování: Laboratorní příručka), Cold Spring Harbor, New York, 1989). V takovém případě genomická nebo cDNA sekvence, kodující signální peptid, se může vázat na genomickou nebo cDNA sekvenci kodující heterologickou bílkovinu, načež se DNA sekvence může modifikovat včleněním syntetických oligonukleotidů kodujících o sobě známým způsobem sekvenci

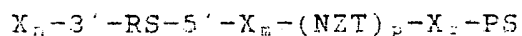
$$X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS$$

kde jednotlivé symboly mají shora uvedený význam.

Statistický DNA fragment a/nebo sekvence

$$5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'$$

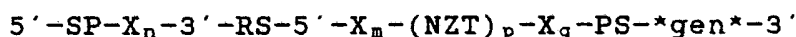
kde jednotlivé symboly mají shora uvedený význam, může být směsného původu syntetického a genomického, směsného původu syntetického a cDNA nebo směsného původu genomického a cDNA, přičemž se příprava provádí o sobě známým způsobem tepelnou hybridizací fragmentů syntetického, genomického nebo cDNA původu (podle vhodnosti), přičemž fragmenty odpovídají různým částem DNA sekvence jako celku. Proto je možno mít představu, že DNA sekvence, kodující signální peptid nebo heterologický polypeptid, může být genomického nebo cDNA původu, zatímco sekvence



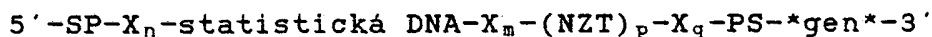
může být připravena synteticky.

Výhodná konstrukce DNA, kodující insulinové prokursory, je objasněna na sekvenčním seznamu ID číslo 1 - 13 (Sequence Listing ID Nos. 1-13) nebo na jeho modifikacích. Příklady vhodných modifikací DNA sekvence jsou nukleotidové substituce, které nevedou k růstu na jinou aminokyselinovou sekvenci bílkoviny, které však mohou odpovídat kodovanému použití kvasinkového organismu, do kterého se vkládá DNA konstrukce nebo nukleotidové substituce, které umožňují růst do odlišné aminokyselinové sekvence a tak možnou odlišnou strukturu bílkoviny. Jakožto jiný případ možné modifikace se uvádí včlenění tří nebo násobku tří nukleotidů do sekvence, adice tří nebo násobku tří nukleotidů na jedno zakončení sekvence a vyuštění tří nebo násobku tří nukleotidů buď ze zakončení nebo z vnitřku sekvence.

Rekombinantním expresním vektorem nesoucím sekvenci



nebo



kde jednotlivé symboly mají vždy shora uvedený význam, může být jakýkoliv vektor, který je schopen replikace v kvasinkovém organismu. Ve vektoru kterákoliv DNA sekvence může být operativně napojena na vhodný promotor sekvence. Promotorem může být jakákoliv DNA sekvence, která vykazuje transkripční aktivitu v kvasinkách a může být odvozena od genů kodujících bílkoviny buď homologické nebo heterologické ke kvasinkám. Promotor je s výhodou odvozen od genu kodujícího bílkovinový homolog ke kvasinkám. Jakožto příklady vhodných promotorů se uvádějí promotory *Saccharomyces cerevisiae* MF α 1, TPI, ADH nebo PGK.

Shora uvedené sekvence mají být také operativně spojeny se vhodným terminátorem, jako je například TPI terminátor (viz T. Alber a G. Kawasaki, J. Mol. Appl. Genet. 1, str. 419 až 434, 1982).

Rekombinantní expresní vektor podle vynálezu dále zahrnuje DNA sekvenci, umožňující replikaci vektoru v kvasinkách. Jakožto příklady takových sekvencí se uvádějí kvasinkové plasmidové 2 μ

replikační geny REP 1-3 a počátek replikace. Vektor také může zahrnovat selektovatelný signální znak, například *Schizosaccharomyces pombe* TPI gen, jak popisuje P.R. Russell, Gene 40, str. 125 až 130, 1985.

Procesy, používané k ligaci sekvence

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

kde jednotlivé symboly mají shora uvedený význam, statistický DNA fragment, promotor a terminátor a jejich včleňování do vhodných kvasinkových vektorů obsahujících informaci nutnou pro replikaci kvasinek jsou pracovníkům v oboru dobře známy (viz například Sambrook, Fritsch a Maniatis, Molecular Cloning: A Laboratory Manual (Molekulární clonování: Laboratorní příručka), Cold Spring Harbor, New York, 1989). Je jasné, že se vektor může konstruovat buď nejdříve přípravou DNA konstrukce obsahující celou sekvenci

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

kde jednotlivé symboly mají shora uvedený význam, a následně včleněním tohoto fragmentu do vhodného expresního vektoru nebo následným včleněním DNA fragmentů obsahujících genetickou informaci pro jednotlivé prvky (jako například signální peptid, sekvence

X_n-3'-RS-5'-X_m-(NZT)_p-X_q

kde jednotlivé symboly mají shora uvedený význam, nebo heterologický polypeptid) s následnou ligací.

Kvasinkový organismus, používaný při způsobu podle vynálezu může být jakýkoliv vhodný kvasinkový organismus, který při kultivování produkuje velká množství heterologického žádaného polypeptidu. Jakožto příklady takových kvasinkových organismů se mohou uvést kmeny kvasinek druhu *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe* a *Saccharomyces uvarum*. Transformace kvasinkových buněk se může například realizovat protoplastovou formací následovanou transformací o sobě známým způsobem. Použitým prostředím ke kultivaci buněk může být jakékoliv prostředí pro růst kvasinkového organismu. Sekretovaná heterologická bílkovina, jejíž značný podíl je obsažen v prostředí ve správně produkované formě, se může získat o sobě známým způsobem včetně oddělení kvasinkových buněk z prostředí odstředěním

nebo filtrací, vysrážením bílkovinových složek supernatantu nebo odfiltrováním za použití soli, například síranu amonného a následným čištěním různými chromatografickými procesy, jako je například ionexová chromatografie a afinitní chromatografie.

Vynález blíže objasňují připojené výkresy:

na obr. 1 je schéma konstrukce pMT742Sm,

na obr. 2 je schéma konstrukce pLaC202,

na obr. 3 je DNA sekvence a odvozená aminokyselinová sekvence na clonovacím místě v pLaC202 pro statistické DNA fragmenty (připomíná se, že sekvence se štěpí v ClaI místě a že ligace bez včlenění statistické DNA vede ke změně ve čtecím rámci),

na obr. 4 je schéma konstrukce pLSC6315D#.

Vynález rovněž objasňují následující příklady praktického provedení, které však nejsou míněny jako omezení vynálezu.

Příklady provedení vynálezu

Plasmidové a DNA materiály

Všechny expresní plasmidy jsou C-POT typu. Všechny plasmidy jsou popsány ve zveřejněné evropské přihlášce vynálezu číslo 171142 a jsou charakterizovány obsahem Schizosaccharomyces pombe triosovým fosfátovým isomerasovým genem (POT) pro účely plasmidové selekce a stabilizace. Plasmid, obsahující POT gen, je dostupný jakožto uložený E. coli kmen (ATCC 39685). Plasmidy dále obsahují S. cerevisiae triosový fosfátový isomerasový promotor a terminátor (P_{TPI} a T_{TPI}). Jsou identické s pMT742 (M. Engel-Mitani a kol., Gene 73, str. 113 až 120, 1988) (viz obr. 1) s výjimkou pro oblast definovanou Sph-XbaI restrikčními místy zahrnujícími P_{TPI} a kodujícími oblast pro signal/vedoucí prvek/produkt.

P_{TPI} se modifikuje se zřetelem na sekvenci nalezenou v pMT742 jedině k usnadnění konstrukční operace. Vnitřní SphI restrikční místo se eliminuje SphI stěpením, odstraněním ocasů a religací. Kromě toho DNA sekvence ve směru proti promotoru a bez jakéhokoli napadání promotoru, se odstraňuje Bal31 exonukleaso-

vým zpracováním s následnou adicí SphI restričního místového spojovníku. Tato promotorová konstrukce na 373 bp SphI-EcoRI fragmentu se označuje pMT7426 (obr. 1).

Soubor různých DNA fragmentů zaujímá případně místo v menším E. coli plasmidu typu pT7 shora popsaného (viz světový patentový spis WO 89/02463) pouze modifikovaného se zřetelem na shora popsaný P_{TP1}, to je pT7 6. Pro statistické clonování, dále popsané, se používá genomové DNA různých původů. S. cerevisiae DNA se izoluje ze kmene MT633 (deponován 7. prosince 1990 v Německé sbírce mikroorganismů a buněčných kultur podle ustanovení Budapeštské dohody na Mezinárodním shromáždění o deponování mikroorganismů pro účely patentovaných postupů pod číslem DSM 6278). A. oryzae DNA se izoluje ze kmene A1560 (IFO 4177).

Konečně se používají četné syntetické DNA fragmenty, které jsou všechny syntetizovány na automatickém DNA syntetizeru (Applied Biosystems model 380A) za použití fosforamiditové chemie a obchodně dostupných reakčních činidel (S.L. Beaucage a M.H. Caruthers, Tetrahedron Letters 22, str. 1859 až 1869, 1981). Oligonukleotidy se čistí poolyakrylamidovou gelovou elektroforesou za denaturačních podmínek. Před tepelnou hybridizací se kinasují komplementární páry takových DNA jednoduchých řetězců za použití T4 polynukleotidové kinasy a ATP.

Všechny jiné způsoby a materiály jsou obecně známe (viz například (J.Sambrook a kol., Molecular Cloning: A Laboratory Manual (Molekulární clonování: Laboratorní příručka), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).

Příklad 1

Konstrukce pLaC202

490 bp SphI-ApaI pT7 1966 (viz světový patentový spis WO 89/02463, obr. 5) a 179 bp HinfI-XbaI fragment pT7.αMI3 (viz světový patentový spis WO 89/02463, obr. 1) spojeny v 11 kb XbaI-SphI fragment pMT742 přes syntetický adaptor:

NOR367/373 CAACCATCGATAACACCACTTTGGCTAAGAG

CCGGGTTGGTAGCTATTGTGGTGAAACCGATTCTCTAA

resultují v plasmid pLaC202 (obr. 2 a 3, jakož také sekvence ID No.1).

Tento vektor, obsahující specifické *Cla*I místo, tvoří jedno provedení statistického DNA klonovacího vektro, ve kterém produkováný gen koduje pro insulinový prekursor MI3 (B(1-29)-Ala-Ala-Lys-A(1-21)). Následující příklady se týkají vedoucích prvků clonovaných touto konstrukcí.

Příklad 2

Konstrukce pLSC6315 a pLSC5210

Celá DNA se izoluje z *S. cerevisiae* kmenu MT663 a digeruje se s *Taq*I, *Hin*PI nebo *Taq*I + *Hin*PI. Produkty se oddělí podle velikosti na 1% agarosovém gelu a fragmenty menší než 600 bp se izolují z každé ze tří digescí.

pLaC202, předem digerovaný s *Cly*I, u něhož se předešlo samoligaci s telecí střevní alkalickou fosfatázou (CIAP) defosforylací se mísí s fragmentovými fondy shora popsány a liguje se. Kmen *E. coli* MT172 (MT172 = MC 1000 *m⁺r⁻ara⁺ leuB-6*; MC 1000 (viz M. Casadaban a S. Cohen, *J. Mol. Biol.* 138, str. 179, 1980)) se transformuje se shora uvedenou ligační směsí a získají se vhodné 5000 *Ap^R* transformanty pro každou směs. Připraví se rekombinantní plasmidy z každého ze tří typů ve fondech zahrnujících všech 5000 transformantů. Těchto plasmidových fondů se používá k transformaci *S. cerevisiae* kmenu MT663 a resultující TOI transformanty se imunotřídí pro sekreci MI3.

Z překvapivě velkého počtu pozitivních transformantů se reisoluje osm zjevně nejúčinnějších a z nich izolovaný plasmidový obsah.

Jako výsledek této operace se očekává, že většina kvasinkových získaných transformantů má heterofenní populaci plasmidů a k získání pravých clonů se proto provádí stupeň plasmidové reisolace. Plasmidové preparace z každého z osmi reizolovaných kvasinkových transformantů se používá ke transformaci *E. coli* kmene MT172 až *Ap^R*. Plasmidů z 12 *E. coli* transformantů pro každý z osmi kvasinkových isolátů se individuálně používá k transformaci

kvasinkového kmene MT663, TPI a MI3 sekreteační transformanty se identifikují immunotříděním.

Sekvencování inzertů osmi izolovaných derivátů pLaC202 vykazuje tři odlišné sekvence, dvě z nich pLSC6315 a pLSC5210 jsou nejúčinnějším nosičem MI3 sekrece. Sekvence clonované DNA a lemovací oblasti jsou ukázány na sekvenčním seznamu ID číslo 2 a 4.

Příklad 3

Modifikace pLSC6315

pLSC6315 se volí pro další modifikaci clonované syntetické vedoucí sekvence.

pLSC6315 se digeruje s ApaI endonukleasou s následným zpracováním endonukleasovou Bal31. Po fenolové extrakci se rezultující DNA digeruje s XbaI a izolují se DNA fragmenty menší než originální 367 bp ApaI-XbaI fragment.

pLaC202 se digerují s ClaI a jednořetězcové CG generované ocasy se odstraní s následnou digescí XbaI a izolací 11 kb XbaI-]ClaI[fragmentu ("][" znamená, že jsou jednořetězcové ocasy odříznuty). Tento fragment se mísí s pLSC6415 fragmenty izolovanými shora a ligatovanými (obr. 6).

Proces transformace a třídění, popsany v příkladu 2, se opakuje a pLSC6315D3 a pLSC6315D7 se izoluje jakožto plasmidy podporující MI3 sekreci mnohem účinněji než originální pLSC6315 (viz sekvenční seznam ID číslo 6 nebo 8).

Příklad 4

Konstrukce pLAO1 - 5

Celá DNA z mikroorganismu *Aspergillus oryzae* kmen A1560 se zpracovává jako shora popsáno pro mikroorganismus *S. cerevisiae* DNA v příkladu 2 a clonování a reclonování se provádí přesně, jako je popsáno v příkladu 2, s tou výjimkou, že počet *E. coli* transformantů v prvním clonování se sníží na přibližně 3000 na ligační směs. Tento pokus vede k izolaci pěti klonu *A. oryzae* DNA, které v pLaC202 kontextu zprostředkovávají sekreci insulino-

vého prekursoru MI3 z *S. cerevisiae*.

Sekvencování inzertů vykazuje pět odlišných sekvencí, dvě z nichž (pLAO2 a pLAO5) jsou mnohem účinnější, pokud je záměrem sekrece MI3. Sekvence DNA inzertů v pLAO2 a pLAO5 jsou ukázány spolu s lemovacími oblastmi v sekvenčním seznamu ID číslo 10 nebo 12.

Příklad 5

Kvasinkové kmeny, kotvící plasmidy, jak shora popsáno, se nechávají růst v prostředí YPD (F. Sherman a kol., *Methods in Yeast Genetics* (Způsoby v genetice kvasinek), Cold Spring Harbor Laboratory 1981). Pro každý kmen se 6 individuálních 5 ml kultur protřepává při teplotě 30 °C po dobu 60 hodin s konečným OD₆₀₀ přibližně 15. Po odstředění se supernatant odstraní pro analýzu HPLC, čímž se měří koncentrace sekretovaného inzulinového prekursoru způsobem, který popsal Leo Snel a kol., *Chromatographia* 24, str. 329 až 332, 1987).

V tabulce I jsou uvedeny expresní hladiny inzulinového prekursoru, MI3, použitím vedoucích sekvencí izolovaných způsobem podle vynálezu, jakožto procento hladiny získané s transformanty pMT7426, za použití MF α (1) vedoucího prvku *S. cerevisiae*.

Tabulka I

pMT742	100 %
pLSC6315	100 %
pLSC5210	60 %
pLSC6315D3	175 %
pLSC6315D7	120 %
pLAO2	120 %
pLAO5	60 %

Průmyslová využitelnost

Kvasinkový expresní vektor zahrnující sekvenci pro konstrukci syntetických vedoucích peptidových sekvencí.

Sekvenční seznam

(1) Obecná informace

- (i) Přihlašovatel: Novo Nordisk A/S
- (ii) Název vynálezu: Způsob konstrukce syntetické vedoucí peptidové sekvence
- (iii) Počet sekvencí: 13
- (iv) Korespondenční adresa:
 - (A) adresa: Novo Nordisk A/S, patentové oddělení
 - (B) ulice: Novo Alle
 - (C) město: Bagsvaerd
 - (E) země: Dánsko
 - (F) ZIP: DK-2880
- (v) Počítačem čitelná forma:
 - (A) typ prostředí: floppy disk
 - (B) počítač: IBM PC kompatibilní
 - (C) operační systém: PC-DOS/MS-DOS
 - (D) software: PatentInRelease # 1,0, verze # 1,25
- (vi) Současné údaje o přihlášce vynálezu:
 - (A) číslo přihlášky vynálezu
 - (B) datum podání
 - (C) mezinárodní klasifikace
- (viii) Informace o zástupci
 - (A) jméno: Thaloe-Madsen, Birgit
 - (B) číslo: 3540.204-WO
- (ix) Telekomunikační informace:
 - (A) telefon: + 45 4444 8888
 - (B) telefax: + 45 4449 3256
 - (C) telex" 37304

(2) Informace pro SEQ ID NO:1:

- (i) Charakteristiky sekvence:
 - (A) délka: 335 páru bází
 - (B) typ: nukleová kyselina
 - (C) řetězovitost: jeden řetězec
 - (D) topologie: lineární
- (ii) Typ molekuly: cDNA
- (xi) Popis sekvence: SEQ ID NO:1:

GAATTCATTTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTTC ATACACAATA	60
TAAACGATTA AAAGAATGAA AGTCCTCCTG CTGCTTTCCC TCATTGGATT CTGCTGGGCC	120
CAACCATOGA TAACACCACT TTGGCTAAGA GATTGGTTAA CCAACACTTG TGCGGTTCCTC	180
ACTTGGTTGA AGCTTTGTAC TTGGTTTGGG GTGAAAGAGG TTTCTTCTAC ACTCCTAAGG	240
CTGCTAAGGG TATTGTGAA CAATGCTGTA CCTCCATCTG CTCCTTGTAC CAATTGGAAA	300
ACTACTGCAA CTAGACGCAG CCGCAGGCT CTAGA	335

(2) Informace pro SEQ ID NO:2:

(i) Charakteristiky sekvence:

- (A) délka: 492 páru bází
- (B) typ: nukleová kyselina
- (C) řetězcovitost: jeden řetězec
- (D) topologie: lineární

(ii) Typ molekuly: cDNA

(ix) Charakteristiky

- (A) jméno/klíč: CDS
- (B) lokace: 76..468

(ix) Charakteristiky

- (A) jméno/klíč: sig peptid
- (B) lokace: 76..309

(ix) Charakteristiky

- (A) jméno/klíč: mat peptid
- (B) lokace: 310..468

(xi) Popis sekvence: SEQ ID NO:2:

GAATTCATTTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTTC ATACACAATA	60
TAAACGATTA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC	111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe	
-78 -75 -70	
TGC TGG GCC CAA CCA TCG ATA GAT GGA ACA CAT TTT CCG AAC AAC AAT	159
Cys Trp Ala Gln Pro Ser Ile Asp Gly Thr His Phe Pro Asn Asn Asn	
-65 -60 -55	

GTC CCA ATA GAC ACA AGA AAA GAA GGA CTA CAG CAT GAT TAC GAT ACA Val Pro Ile Asp Thr Arg Lys Glu Gly Leu Gln His Asp Tyr Asp Thr -50 -45 -40 -35	207
GAA ATT TTG GAG CAC ATT GGA AGC GAT GAG TTA ATT TTG AAT GAA GAG Glu Ile Leu Glu His Ile Gly Ser Asp Glu Leu Ile Leu Asn Glu Glu -30 -25 -20	255
TAT GTT ATT GAA AGA ACT TTG CAA GCC ATC GAT AAC ACC ACT TTG GCT Tyr Val Ile Glu Arg Thr Leu Gln Ala Ile Asp Asn Thr Thr Leu Ala -15 -10 -5	303
AAG AGA TTC GTT AAC CAA CAC TTG TGC GGT TCC CAC TTG GTT GAA GCT Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala 1 5 10	351
TTG TAC TTG GTT TGC GGT GAA AGA GGT TTC TTC TAC ACT CCT AAG GCT Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala 15 20 25 30	399
GCT AAG GGT ATT GTC GAA CAA TGC TGT ACC TCC ATC TGC TCC TTG TAC Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr 35 40 45	447
CAA TTG GAA AAC TAC TGC AAC TAGAAGCAGC CGCAGGCTC TAGA Gln Leu Glu Asn Tyr Cys Asn 50	492

(2) Informace pro SEQ ID NO:3:

- (i) Charakteristiky sekvence:
 - (A) délka: 131 aminokyselin
 - (B) typ: aminokyselina
 - (D) topologie: lineární
- (ii) Typ molekuly: bílkovina
- (xi) Popis sekvence: SEQ ID NO:3:

Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln -78 -75 -70 -65
Pro Ser Ile Asp Gly Thr His Phe Pro Asn Asn Asn Val Pro Ile Asp -60 -55 -50
Thr Arg Lys Glu Gly Leu Gln His Asp Tyr Asp Thr Glu Ile Leu Glu -45 -40 -35
His Ile Gly Ser Asp Glu Leu Ile Leu Asn Glu Glu Tyr Val Ile Glu -30 -25 -20 -15

Arg Thr Leu Gln Ala Ile Asp Asn Thr Thr Leu Ala Lys Arg Phe Val
-10 -5 1

Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val
5 10 15

Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile
20 25 30

Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn
35 40 45 50

Tyr Cys Asn

(2) Informace pro SEQ ID NO:4:

- (i) Charakteristiky sekvence:
 - (A) délka: 420 párů bází
 - (B) typ: nukleová kyselina
 - (C) řetězcovitost: jeden řetězec
 - (D) topologie: lineární
- (ii) Typ molekuly: cDNA
- (ix) Charakteristiky
 - (A) jméno/klíč: CDS
 - (B) lokace: 76..396
- (ix) Charakteristiky
 - (A) jméno/klíč: sig peptid
 - (B) lokace: 76..237
- (ix) Charakteristiky
 - (A) jméno/klíč: mat peptid
 - (B) lokace: 238..396
- (xi) Popis sekvence: SEQ ID NO:4:

GAATTCATTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTTT ATACACAATA	60
TAAACGATTAA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC	111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe	
-54 -50 -45	
TGC TGG GCC CAA CCA TCG CTA TTG GAG TCA CTT ACG CTC GTT GAT GTT	159
Cys Trp Ala Gln Pro Ser Leu Leu Glu Ser Leu Thr Leu Val Asp Val	
-40 -35 -30	
GAC GCA CTG TCG GAT ATT GAT GTA CTT GTT GAG TCT GAA ACG CTT GTG	207
Asp Ala Leu Ser Asp Ile Asp Val Leu Val Glu Ser Glu Thr Leu Val	
-25 -20 -15	

CTT GTC GAT AAC ACC ACT TTG GCT AAG AGA TTC GTT AAC CAA CAC TTG	255
Leu Val Asp Asn Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu	
-10 -5 1 5	
TGC GGT TCC CAC TTG GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA	303
Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg	
10 15 20	
GGT TTC TTC TAC ACT CCT AAG GCT GCT AAG GGT ATT GTC GAA CAA TGC	351
Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys	
25 30 35	
TGT ACC TCC ATC TGC TCC TTG TAC CAA TTG GAA AAC TAC TGC AAC	396
Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn	
40 45 50	
TAGACGCAGC CGGCAGGCTC TAGA	420

(2) Informace pro SEQ ID NO:5:

- (i) Charakteristiky sekvence:
 (A) délka: 107 aminokyselin
 (B) typ: aminokyselina
 (D) topologie: lineární
- (ii) Typ molekuly: bílkovina
- (xi) Popis sekvence: SEQ ID NO:5:

Met Lys Val Phe Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln
-54 -50 -45 -40

Pro Ser Leu Leu Glu Ser Leu Thr Leu Val Asp Val Asp Ala Leu Ser
-35 -30 -25

Asp Ile Asp Val Leu Val Glu Ser Glu Thr Leu Val Leu Val Asp Asn
-20 -15 -10

Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His
-5 1 5 10

Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr
15 20 25

Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile
30 35 40

Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
45 50

(2) Informace pro SEQ ID NO:6:

- (i) Charakteristiky sekvence:
 - (A) délka: 453 párů bází
 - (B) typ: nukleová kyselina
 - (C) řetězcovitost: jeden řetězec
 - (D) topologie: lineární
- (ii) Typ molekuly: cDNA
- (ix) Charakteristiky
 - (A) jméno/klíč: CDS
 - (B) lokace: 76..420
- (ix) Charakteristiky
 - (A) jméno/klíč: sig peptid
 - (B) lokace: 76..270
- (ix) Charakteristiky
 - (A) jméno/klíč: mat peptid
 - (B) lokace: 271..420
- (xi) Popis sekvence: SEQ ID NO:6:

GAATTCATTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTTT ATACACAATA	60
TAAACGATTA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC	111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe	
-65 -60 -55	
TGC TGG GCC CAA CCA ATA GAC ACA AGA AAA GAA GGA CTA CAG CAT GAT	159
Cys Trp Ala Gln Pro Ile Asp Thr Arg Lys Glu Gly Leu Gln His Asp	
-50 -45 -40	
TAC GAT ACA GAA ATT TTG GAG CAC ATT GGA AGC GAT GAG TTA ACC CCG	207
Tyr Asp Thr Glu Ile Leu Glu His Ile Gly Ser Asp Glu Leu Thr Pro	
-35 -30 -25	
AAT GAA GAG TAT GTT ATT GAA AGA ACT TTG CAA GCC ATC GAT AAC ACC	255
Asn Glu Glu Tyr Val Ile Glu Arg Thr Leu Gln Ala Ile Asp Asn Thr	
-20 -15 -10	
ACT TTG GCT AAG AGA TTC GTT AAC CAA CAC TTG TGC GGT TCC CAC TTG	303
Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu	
-5 1 5 10	
GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA GGT TTC TTC TAC ACT	351
Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr	
15 20 25	

CCT AAG GCT GCT AAG GGT ATT GTC GAA CAA TGC TGT ACC TCC ATC TGC 399
Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
30 35 40

TCC TIG TAC CAA TTG GAA AAC TACTGCAACT AGAAGCAGCC CGCAGGCTCT 450
Ser Leu Tyr Gln Leu Glu Asn
45 50

AGA 453

(2) Informace pro SEQ ID NO:7:

(i) Charakteristiky sekvence:

(A) délka: 115 aminokyselin

(B) typ: aminokyselina

(D) topologie: lineární

(ii) Typ molekuly: bílkovina

(xi) Popis sekvence: SEQ ID NO:7:

Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln
-65 -60 -55 -50

Pro Ile Asp Thr Arg Lys Glu Gly Leu Gln His Asp Tyr Asp Thr Glu
-45 -40 -35

Ile Leu Glu His Ile Gly Ser Asp Glu Leu Thr Pro Asn Glu Glu Tyr
-30 -25 -20

Val Ile Glu Arg Thr Leu Gln Ala Ile Asp Asn Thr Thr Leu Ala Lys
-15 -10 -5

Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
1 5 10 15

Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala
20 25 30

Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln
35 40 45

Leu Glu Asn
50

(2) Informace pro SEQ ID NO:8:

(i) Charakteristiky sekvence:

(A) délka: 459 párů bází

(B) typ: nukleová kyselina

(C) řetězcovitost: jeden řetězec

(D) topologie: lineární

- (ii) Typ molekuly: cDNA
- (ix) Charakteristiky
 - (A) jméno/klíč: CDS
 - (B) lokace: 76..435
- (ix) Charakteristiky
 - (A) jméno/klíč: sig peptid
 - (B) lokace: 76..276
- (ix) Charakteristiky
 - (A) jméno/klíč: mat peptid
 - (B) lokace: 277..435
- (xi) Popis sekvence: SEQ ID NO:8:

GAATTCATTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTTC ATACACAATA	60
TAAACGATTAA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC	111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe	
-67 -65 -60	
TGC TGG GCC CAA CCT GTC CCA ATA GAC ACA AGA AAA GAA GGA CTA CAG	159
Cys Trp Ala Gln Pro Val Pro Ile Asp Thr Arg Lys Glu Gly Leu Gln	
-55 -50 -45 -40	
CAT GAT TAC GAT ACA GAA ATT TTG GAG CAC ATT GGA AGC GAT GAG TTA	207
His Asp Tyr Asp Thr Glu Ile Leu Glu His Ile Gly Ser Asp Glu Leu	
-35 -30 -25	
ACC COG AAT GAA GAG TAT GTT ATT GAA AGA ACT TTG CAA GCC ATC GAT	255
Thr Pro Asn Glu Glu Tyr Val Ile Glu Arg Thr Leu Gln Ala Ile Asp	
-20 -15 -10	
AAC ACC ACT TTG GCT AAG AGA TTC GTT AAC CAA CAC TTG TGC GGT TCC	303
Asn Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser	
-5 1 5	
CAC TTG GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA GGT TTC TTC	351
His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe	
10 15 20 25	
TAC ACT CCT AAG GCT GCT AAG GGT ATT GTC GAA CAA TGC TGT ACC TCC	399
Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser	
30 35 40	
ATC TGC TCC TTG TAC CAA TTG GAA AAC TAC TGC AAC TAGACGCAGC	445
Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn	
45 50	
COGCAGGCTC TAGA	459

(2) Informace pro SEQ ID NO:9:

- (i) Charakteristiky sekvence:
 - (A) délka: 120 aminokyselin
 - (B) typ: aminokyselina
 - (D) topologie: lineární
- (ii) Typ molekuly: bílkovina
- (xi) Popis sekvence: SEQ ID NO:9:

Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln
-67 -65 -60 -55

Pro Val Pro Ile Asp Thr Arg Lys Glu Gly Leu Gln His Asp Tyr Asp
-50 -45 -40

Thr Glu Ile Leu Glu His Ile Gly Ser Asp Glu Leu Thr Pro Asn Glu
-35 -30 -25 -20

Glu Tyr Val Ile Glu Arg Thr Leu Gln Ala Ile Asp Asn Thr Thr Leu
 -15 -10 -5

Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu
 1 5 10

Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys
 15 20 25

Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu
30 35 40 45

Tyr Gln Leu Glu Asn Tyr Cys Asn
 50

(2) Informace pro SEQ ID NO:10:

- (i) Charakteristiky sekvence:
 - (A) délka: 408 párů bází
 - (B) typ: nukleová kyselina
 - (C) řetězcovitost: jeden řetězec
 - (D) topologie: lineární
- (ii) Typ molekuly: cDNA
- (ix) Charakteristiky
 - (A) jméno/klíč: CDS
 - (B) lokace: 76..384
- (ix) Charakteristiky
 - (A) jméno/klíč: sig peptid

(B) lokace: 76..225

(ix) Charakteristiky

(A) jméno/klíč: mat peptid

(B) lokace: 226..384

(xi) Popis sekvence: SEQ ID NO:10:

GAATTCATTC AAGAATAGTT CAAACAAGAA GATTACAAAC TATCAATTC ATACACAATA	60
TAAACGATTA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC	111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe	
-50 -45 -40	
TGC TGG GCC CAA CCA TCG ATC TTG GAT TAT GTT GAC TTG GGT GCG GAA	159
Cys Trp Ala Gln Pro Ser Ile Leu Asp Tyr Val Asp Leu Gly Ala Glu	
-35 -30 -25	
CTG ATC TCC ATT CGT GGG TAT GAT AAC CTC AAC GAC GCG ATC GAT AAC	207
Leu Ile Ser Ile Arg Gly Tyr Asp Asn Leu Asn Asp Ala Ile Asp Asn	
-20 -15 -10	
ACC ACT TTG GCT AAG AGA TTC GTT AAC CAA CAC TTG TGC GGT TCC CAC	255
Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His	
-5 1 5 10	
TTG GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA GGT TTC TTC TAC	303
Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr	
15 20 25	
ACT CCT AAG GCT GCT AAG GGT ATT GTC GAA CAA TGC TGT ACC TCC ATC	351
Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile	
30 35 40	
TGC TCC TTG TAC CAA TTG GAA AAC TAC TGC AAC TAGAOGCAGC COGCAGGCTC	404
Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn	
45 50	
TAGA	408

(2) Informace pro SEQ ID NO:11:

(i) Charakteristiky sekvence:

(A) délka: 103 aminokyselin

(B) typ: aminokyselina

(D) topologie: lineární

(ii) Typ molekuly: bílkovina

(xi) Popis sekvence: SEQ ID NO:11:

Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln
 -50 -45 -40 -35

Pro Ser Ile Leu Asp Tyr Val Asp Leu Gly Ala Glu Leu Ile Ser Ile
 -30 -25 -20

Arg Gly Tyr Asp Asn Leu Asn Asp Ala Ile Asp Asn Thr Thr Leu Ala
 -15 -10 -5

Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala
 1 5 10

Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala
 15 20 25 30

Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr
 35 40 45

Gln Leu Glu Asn Tyr Cys Asn
 50

(2) Informace pro SEQ ID NO:12:

- (i) Charakteristiky sekvence:
 - (A) délka: 372 párů bází
 - (B) typ: nukleová kyselina
 - (C) řetězcovitost: jeden řetězec
 - (D) topologie: lineární
- (ii) Typ molekuly: cDNA
- (ix) Charakteristiky
 - (A) jméno/klíč: CDS
 - (B) lokace: 76..348
- (ix) Charakteristiky
 - (A) jméno/klíč: sig peptid
 - (B) lokace: 76..189
- (ix) Charakteristiky
 - (A) jméno/klíč: mat peptid
 - (B) lokace: 190..348
- (xi) Popis sekvence: SEQ ID NO:12:

GAATTCATTC AAGAATAGTT CAACAAGAA GATTACAAAC TATCAATTTT ATACACAATA 60
TAAACGATTA AAAGA ATG AAA GTC TTC CTG CTG CTT TCC CTC ATT GGA TTC 111
Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe
-38 -35 -30
TGC TGG GCC CAA CCA TCG CAC ACT ACC ATC GGC ACC GCA ACT GAC AAA 159
Cys Trp Ala Gln Pro Ser His Thr Thr Ile Gly Thr Ala Thr Asp Lys
-25 -20 -15
AAC ATC GAT AAC ACC ACT TTG GCT AAG AGA TTC GTT AAC CAA CAC TTG 207
Asn Ile Asp Asn Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu
-10 -5 1 5
TGC GGT TCC CAC TTG GTT GAA GCT TTG TAC TTG GTT TGC GGT GAA AGA 255
Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
10 15 20
GGT TTC TTC TAC ACT CCT AAG GCT GCT AAG GGT ATT GTC GAA CAA TGC 303
Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys
25 30 35
TGT ACC TCC ATC TGC TCC TTG TAC CAA TTG GAA AAC TAC TGC AAC 348
Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
40 45 50
TAGAOGCAGC COGCAGGCTC TAGA 372

(2) Informace pro SEQ ID NO:13:

(i) Charakteristiky sekvence:

(A) délka: 91 aminokyselin

(B) typ: aminokyselina

(D) topologie: lineární

(ii) Typ molekuly: bílkovina

(xi) Popis sekvence: SEQ ID NO:13:

Met Lys Val Phe Leu Leu Leu Ser Leu Ile Gly Phe Cys Trp Ala Gln
-38 -35 -30 -25
Pro Ser His Thr Thr Ile Gly Thr Ala Thr Asp Lys Asn Ile Asp Asn
-20 -15 -10
Thr Thr Leu Ala Lys Arg Phe Val Asn Gln His Leu Cys Gly Ser His
-5 1 5 10
Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr
15 20 25
Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile
30 35 40
Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
45 50

P A T E N T O V É N Á R O K Y

1. Způsob konstrukce syntetické vedoucí peptidové sekvence pro sekreci heterologických polypeptidů v kvasinkách v y z n a - č u j í c í s e t í m , že se

a) vnáší statistický DNA fragment do kvasinkového expresního vektoru zahrnujícího sekvenci

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

kde znamená

SP DNA sekvenci kodující signální peptid,

X_n DNA sekvenci kodující n aminokyselin, přičemž n znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

RS restriční endonukleasové rekognizační místo pro včlenění statistických DNA fragmentů, kteréžto místo je na spojení X_n a X_m,

X_m DNA sekvenci kodující m aminokyselin, přičemž m znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

(NZT)_p DNA sekvenci kodující Asn-Xaa-Thr, přičemž m znamená 0 nebo číslo 1,

X_q DNA sekvenci kodující q aminokyselin, přičemž q znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

PS DNA sekvenci kodující peptid definující kvasinkové zpracovatelské místo a

gen sekvenci kodující heterologický polypeptid,

b) transformuje se kvasinková hostitelská buňka expresním vektorem podle stupně (a),

c) kultivuje se transformovaná hostitelská buňka podle stupně (b) za vhodných podmínek a

d) vytrídí se kultura podle stupně (c) pro sekreci heterologového polypeptidu.

2. Způsob podle nároku 1, v y z n a č u j í c í s e t í m , že statistický DNA fragment, začleňovaný do vektoru, je genomického nebo syntetického původu.

3. Způsob podle nároku 1 až 2, v y z n a č u j í c í s e

t í m, že statistický DNA fragment má délku 6 až přibližně 600 párů bází.

4. Způsob podle nároku 1 až 3, v y z n a č u j í c í s e t í m, že statistický DNA fragment koduje vysoký podíl polárních aminokyselin.

5. Způsob podle nároku 1 až 4, v y z n a č u j í c í s e t í m, že statistický DNA fragment koduje alespoň jeden prolin.

6. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že a/nebo m a/nebo q znamená 1 nebo číslo větší než 1.

7. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že p znamená 1.

8. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že SP znamená DNA sekvenci kodující α -faktorový signální peptid, signální peptid myších slinných žláz, karboxypeptidasový signální peptid, signální peptid kvasinek BAR1 nebo lipasový signální peptid *Humicola lanuginosa* nebo jeho derivát.

9. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že PS je DNA sekvence kodující Lys-Arg, Arg-Lys, Lys-Lys, Arg-Arg nebo Ile-Glu-Gly-Arg.

10. Způsob podle nároku 1, v y z n a č u j í c í s e t í m, že heterologický peptid je volen ze souboru zahrnujícího aprotinin, tkáňový faktorový inhibitor nebo jiné proteasové inhibitory, insulin nebo insulinové prekursory, lidský nebo hovězí růstový faktor, interleukin, glukagon, tkáňový plasminogenový aktivátor, transformační růstový faktor α nebo β , růstový faktor odvozený od krevních destiček, enzymy nebo jejich funkční analogy.

11. Kvasinkový expresní clonující vektor, v y z n a č u j í c í s e t í m, že zahrnuje sekvenci

5'-SP-X_n-3'-RS-5'-X_m-(NZT)_p-X_q-PS-*gen*-3'

kde znamená

- SP DNA sekvenci kodující signální peptid
- X_n DNA sekvenci kodující n aminokyselin, přičemž n znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
- RS restriční endonukleasové rekognizační místo pro včlenění statistických DNA fragmentů, kteréžto místo je na spojení X_n a X_m,
- X_m DNA sekvenci kodující m aminokyselin, přičemž m znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
- (NZT)_p DNA sekvenci kodující Asn-Xaa-Thr, přičemž m znamená 0 nebo číslo 1,
- X_q DNA sekvenci kodující q aminokyselin, přičemž q znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,
- PS DNA sekvenci kodující peptid definující kvasinkové zpracovatelské místo a
- *gen* sekvenci kodující heterologický polypeptid.

12. Kvasinkový expresní clonující vektor podle nároku 11, v y z n a č u j í c í s e t í m , že n a/nebo m a/nebo q znamená 1 nebo číslo větší než 1.

13. Kvasinkový expresní clonující vektor podle nároku 11, v y z n a č u j í c í s e t í m , že p znamená 1.

14. Způsob podle nároku 11, v y z n a č u j í c í s e t í m , že SP znamená DNA sekvenci kodující α-faktorový signální peptid, signální peptid myších slinných žláz, karboxypeptidasový signální peptid nebo signální peptid kvasinek BAR1.

15. Způsob podle nároku 11, v y z n a č u j í c í s e t í m , že PS je DNA sekvence kodující Lys-Arg, Arg-Lys, Lys-Lys, Arg-Arg nebo Ile-Glu-Gly-Arg.

16. Způsob podle nároku 11, v y z n a č u j í c í s e t í m , že heterologický peptid je volen ze souboru zahrnujícího

aprotinin, tkáňový faktorový inhibitor nebo jiné proteasové inhibitory, insulin nebo insulinové prekursory, lidský nebo hovězí růstový faktor, interleukin, glukagon, tkáňový plasminogenový aktivátor, transformační růstový faktor α nebo β , růstový faktor odvozený od krevních destiček, enzymy nebo jejich funkční analogy.

17. Kvasinkový expresní vektor, v y z n a ě u j í c í s e t í m , že zahrnuje sekvenci

5'-SP-X_n-statistická DNA-X_m-(NZT)_p-X_q-PS-*gen*-3'

kde znamená

SP DNA sekvenci kodující signální peptid

X_n DNA sekvenci kodující n aminokyselin, přičemž n znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

statistická DNA fragment včleněný do restričního endonukleasového rekognizačního místa ve spojení X_n a X_m a

RS restriční endonukleasové rekognizační místo pro včlenění statistických DNA fragmentů, kteréžto místo je na spojení X_n a X_m,

X_m DNA sekvenci kodující m aminokyselin, přičemž m znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

(NZT)_p DNA sekvenci kodující Asn-Xaa-Thr, přičemž m znamená 0 nebo číslo 1,

X_q DNA sekvenci kodující q aminokyselin, přičemž q znamená 0 nebo celé číslo 1 až přibližně 10 aminokyselin,

PS DNA sekvenci kodující peptid definující kvasinkové zpracovatelské místo a

gen sekvenci kodující heterologický polypeptid, přičemž sekvence X-X koduje vedoucí peptidovou sekvenci.

18. Způsob podle nároku 17, v y z n a ě u j í c í s e t í m , že statistický DNA fragment, začleňovaný do vektoru, je genomického nebo syntetického původu.

19. Způsob podle nároku 17 nebo 18, v y z n a ě u j í c í s e t í m , že statistický DNA fragment má délku 6 až přibližně 600 párů bází.

20. Způsob podle nároku 17 až 19, v y z n a č u j í c í s e t í m, že statistický DNA fragment koduje vysoký podíl polárních aminokyselin.

21. Způsob podle nároku 17 až 20, v y z n a č u j í c í s e t í m, že statistický DNA fragment koduje alespoň jeden prolin.

22. Způsob podle nároku 17, v y z n a č u j í c í s e t í m, že a/nebo m a/nebo q znamená 1 nebo číslo větší než 1.

23. Způsob podle nároku 17, v y z n a č u j í c í s e p znamená 1.

24. Způsob podle nároku 17, v y z n a č u j í c í s e t í m, že SP znamená DNA sekvenci kodující α -faktorový signální peptid, signální peptid myších slinných žláz, karboxypeptidasový signální peptid, signální peptid kvasinek BAR1 nebo lipasový signální peptid *Humicola lanuginosa* nebo jeho derivát.

25. Způsob podle nároku 17, v y z n a č u j í c í s e t í m, že PS je DNA sekvence kodující Lys-Arg, Arg-Lys, Lys-Lys, Arg-Arg nebo Ile-Glu-Gly-Arg.

26. Způsob podle nároku 17, v y z n a č u j í c í s e t í m, že heterologický peptid je volen ze souboru zahrnujícího aprotinin, tkáňový faktorový inhibitor nebo jiné proteasové inhibitory, insulin nebo insulinové prekursory, lidský nebo hovězí růstový faktor, interleukin, glukagon, tkáňový plasminogenový aktivátor, transformační růstový faktor α nebo β , růstový faktor odvozený od krevních destiček, enzymy nebo jejich funkční analogy.

27. Kvasinková buňka schopná exprese heterologického polypeptidu a transformovaná kvasinkovým expresním vektorem podle nároku 17 až 26.

28. Způsob produkce heterologického polypeptidu v kvasinkách,

v y z n a č u j í c í s e t í m , že se kultivuje kvasničná buňka, schopná exprese heterologického polypeptidu a transformovaná kvasinkovým expresním vektorem podle nároku 17 až 26 za včlenění vedoucí peptidové sekvence konstruované způsobem podle nároku 1, ve vhodném prostředí k expresi a sekreci heterologického polypeptidu, načež se heterologický peptid izoluje z prostředí.



SPOLEČNÁ ADVOKÁTNÍ KANCELAR
VŠETEČKA A PARTNERI
Hájkova 2
120 00 Praha 2