

Даний винахід належить до методів генної інженерії рослин і одержання інсуліну. Більш конкретно, даний винахід належить до способів одержання інсуліну в насінні рослин.

Інсулін являє собою важливий пептидний гормон, необхідний для підтримки гомеостазу глюкози в крові у ссавців, включаючи людину, і інших хребетних. У здорових індивідуумів підвищення рівня глюкози в крові стимулює β -клітини підшлункової залози до виділення інсуліну. Потім поліпептид інсуліну зв'язується зі специфічними рецепторами в м'язах, печінці і жировій тканині, що веде до збільшення споживання глюкози цими цільовими тканинами, підвищення метаболізму, і зниження вироблення глюкози печінкою. Сумарний ефект цих реакцій дозволяє зберегти концентрації глюкози в крові на постійному рівні.

У індивідуумів, які страждають на цукровий діабет, аномально низькі концентрації інсуліну присутні самі по собі у вигляді хронічної гіперглікемії. Клінічні прояви хронічної гіперглікемії є різноманітними і включають сліпоту, ниркову недостатність і, якщо їх не лікувати, зрештою призводять до смерті. За деякими оцінками, цукровий діабет знаходиться на третьому місці серед самих частих причин смерті в промислово розвинених країнах, після серцево-судинних захворювань і раку (Barfoed, H.C., 1987, Chem. Eng. Prog. 83: 49-54). Для досягнення ефективного споживання і метаболізму глюкози крові клітинами, пацієнтів, які страждають на діабет, можна лікувати звичайним введенням інсуліну. Приблизно 0,7% населення світу страждає інсулінозалежним діабетом (цукровий діабет I типу) (Winter, J. et al., 2000, J. of Biotechnol. 84: 175-185). Крім того, за оцінками, кількість пацієнтів, у яких діагностовано діабет, подвоїться приблизно до 300 мільйонів в найближчі 25 років (Kjeldsen, T. et al., 2001, Biotechnol. Gen. Eng. Rev. 18: 89-121). Отже, можливість економічного рентабельного виробництва людського інсуліну в кількостях, що задовольняють очікуваній зростаючій світовій потребі в інсуліні, надто бажана.

In vivo поліпептид людського інсуліну утворюється в β -клітинах підшлункової залози у вигляді попередника одинарного поліпептидного ланцюга із 110 амінокислот, препроінсуліну, який включає розташовані на N-кінці 24 амінокислоти препослідовності, які відщеплюються відразу ж після завершення біосинтезу ланцюга (Steiner, D.F., 2000. J. Fed. Endocrinol. Metab. 13: 229-239). Проінсулін складається з ланцюгів B і A, зв'язаних зв'язуючим пептидом (C-пептид). Під час упакування гормону для секреції C-пептид відщеплюється і видаляється конвертазами прогормону, PC2 і PC1/PC3 (Steiner, D.F., 2000. J. Fed. Endocrinol. Metab. 13: 229-239). Залишається зрілий людський інсулін, білок з 51 амінокислоти, що містить два поліпептидні ланцюги, A (21 амінокислота) і B (30 амінокислот), зв'язані двома міжланцюговими дисульфідними зв'язками. Крім того, ланцюг A містить один внутрішньоланцюговий дисульфідний зв'язок.

Людський інсулін одержують за допомогою множини різних методів. Мікроорганізми, такі як *Escherichia coli* (Frank et al., 1981, in *Peptides: Proceedings of the 7th American Peptide Chemistry Symposium* (Rich & Gross, eds.), Pierce Chemical Co, Rockford, 111 pp 729-739; Chan et al., 1981, *Proc. Natl. Acad. Sci. USA* 78: 5401-5404), *Saccharomyces cerevisiae* (Thim et al., 1986, *Proc. Natl. Acad. Sci. USA* 83: 6766-6770) звичайно застосовують для рекомбінантного одержання інсуліну. У Wang et al. (*Biotechnol. Bioeng.*, 2001, 73: 74-79) показано, що гриби, такі як *Pichia pastoris*, також придатні для одержання інсуліну. Альтернативні виробничі процеси включають продукування в клітинних лініях ссавців, відмінних від людини (Yanagita, M, et al, 1992, *FEBS Lett* 311: 55-59), виділення з підшлункової залози людини, пептидний синтез або напівсинтетичне перетворення в людський інсулін свинячого і бичачого інсуліну. Однак всі ці методи дають низький вихід і вимагають більш високих витрат, ніж це бажано.

Застосування рослин як біореакторів для промислового одержання рекомбінантних білків добре відоме, і була одержана множина білків, включаючи людські терапевтичні білки. Наприклад, в патентах США 4956282, 5550038 і 5629175 описане одержання γ -інтерферону в рослинах; в патентах США 5650307, 5716802 і 5763748 детально описане одержання сироваткового альбуміну людини в рослинах, і в патентах США 5202422, 5639947 і 5959177 описане одержання антитіл в рослинах. Однією із значних переваг систем одержання рекомбінантного білка на основі рослин, є те, що при збільшенні площі вирощування рослин продукція білка може бути недорого масштабована для одержання значно більших кількостей білка. Навпаки, системи із застосуванням ферментації і клітинних культур займають багато місця, вимагають великої кількості обладнання і споживають багато енергії, що робить збільшення виробничих потужностей дуже дорогим. Однак, незважаючи на те, що застосування рослин як біореакторів описане у множині джерел, і незважаючи на вказане вище очікуване збільшення потреби у великих кількостях інсуліну, у відомому рівні техніки представлена тільки обмежена кількість методів, які наочно показують можливість одержання інсуліну в рослинах (див.: Arakawa et al. *Nature Biotech.*, 1998, 16: 934-938; PCT 01/72959).

У Arakawa et al. описане одержання злитого білка, що містить інсулін, у бульбах трансгенної картоплі. Однак інсулін присутній тільки в кількостях не більше 0,05% від загальної кількості розчинного білка, присутнього в трансгенних бульбах. При рівні 0,05% від загального розчинного білка великі кількості біомаси повинні бути оброблені для екстрагування білка, що не вигідно з точки зору економічного застосування картопляних бульб. Більш того у Arakawa et al. не описане виділення інсуліну з тканини картопляної бульби, але замість цього запропонований підхід до профілактики настання діабету I типу шляхом одержання імунотолерантності, який включає введення інсуліну при поїданні трансгенних картопляних бульб.

У заявці на патент PCT WO 01/72959 описане одержання злитого білка, що містить інсулін, в хлоропластах трансгенного тютюну. Однак, по суті відносячи недоліки на рахунок рівнів акумуляції людських білків в тканинах рослин, винахід, описаний в WO 01/72959, обмежений в тому, що одержання в хлоропластах дає акумуляцію інсуліну в зелених тканинах, насамперед, в листі тютюну. Через відносно високий вміст води в зелених тканинах має бути оброблена велика кількість біомаси. Подальше одержання інсуліну потребує негайного екстрагування з біомаси відразу ж після збору, оскільки листя швидко висихає при зберіганні.

Таким чином, з точки зору недоліків, що є в методах рекомбінантного одержання інсуліну в рослинах відомого рівня техніки, в цей час не ясно коли і як синтетичні можливості рослин можуть бути використані для досягнення комерційного одержання інсуліну в рослинах. У даній галузі техніки існує необхідність поліпшення способів комерційного одержання інсуліну в рослинах.

Даний винахід належить до поліпшених способів одержання інсуліну в рослинах. Зокрема, даний винахід

належить до способів одержання інсуліну в насінні.

Отже, в даному винаході представлений спосіб експресії інсуліну в рослинах, що включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин; і
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини; і
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін.

У переважному варіанті даного винаходу послідовністю нуклеїнової кислоти, здатною контролювати експресію в насінні рослин, є переважний для насіння промотор, такий як промотор фазеоліну.

У переважному варіанті даного винаходу інсулін експресується таким чином, який дозволяє акумулювати поліпептид інсуліну в мембрані, яка оточує внутрішньоклітинний компартмент в клітинах насіння. Отже, в даному винаході представлений спосіб експресії інсуліну в рослинах, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин;
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну; і
- (iii) послідовність нуклеїнової кислоти, яка кодує поліпептид, здатний утримувати поліпептид інсуліну в мембрані, яка оточує внутрішньоклітинний компартмент;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини; і
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін.

У іншому переважному варіанті даного винаходу мембраною, яка оточує внутрішньоклітинний компартмент, є ендоплазматична сітка (ER) або ER-похідна накопичувальна везикула. Отже, в даному винаході представлений спосіб експресії інсуліну в рослинах, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин;
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну; і
- (iii) послідовність нуклеїнової кислоти, яка кодує поліпептид, здатний утримувати поліпептид інсуліну в ER або ER-похідній накопичувальній везикулі;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини; і
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін.

У іншому переважному варіанті конструкцію химерної нуклеїнової кислоти вводять в клітину рослин в умовах ядерної геномної інтеграції. При таких умовах послідовність химерної нуклеїнової кислоти стабільно інтегрується в геном рослини.

У ще одному переважному варіанті послідовність нуклеїнової кислоти, що кодує інсулін, оптимізована для застосування кодону рослин, і послідовність нуклеїнової кислоти, що кодує зв'язувальний пептид (С-пептид), укорочена. Переважні послідовності нуклеїнових кислот відповідно до даного винаходу кодують людський, бичачий або свинячий інсулін. Відповідно до даного винаходу, послідовність нуклеїнової кислоти, яка кодує послідовність проінсуліну, застосовують, якщо проінсулін модифікований шляхом укорочення довжини С-пептиду.

У іншому варіанті, в даному винаході представлений спосіб одержання насіння, що містить інсулін, з рослин. Отже, в даному винаході представлений спосіб одержання насіння рослин, що містить інсулін, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин; і
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини;
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння; і
- (d) одержання насіння з вказаних рослин, де насіння містить інсулін.

Переважно, принаймні 0,1% від загального білка насіння, присутнього в насінні, складає інсулін.

Насіння може застосовуватися для одержання популяції рослин, кожна з яких містить множину насіння, яке експресує інсулін.

У даному винаході також представлені рослини, здатні давати насіння, яке експресує інсулін. У переважному варіанті даного винаходу рослини здатні дати насіння, яке містить послідовність химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції:

(а) першу послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин, з яким вона функціонально зв'язана; і

(b) другу послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну, де насіння містить інсулін.

Переважно, принаймні 0,1% від загального білка насіння, присутнього в насінні, складає інсулін.

У переважному варіанті послідовність химерної нуклеїнової кислоти інтегрують в ядерний геном рослини.

У ще одному переважному варіанті даного винаходу рослиною, що застосовується є сафлор, льон або *Arabidopsis*.

У ще одному аспекті даного винаходу представлене насіння рослин, яке експресує інсулін. У переважному варіанті даного винаходу насіння рослин містить послідовність химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції:

(а) першу послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин, з яким вона функціонально зв'язана; і

(b) другу послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну.

Переважно, принаймні 0,1% від загального білка насіння, присутнього в насінні, складає інсулін. Насіння є джерелом, з якого бажаний поліпептид інсуліну, який синтезується в клітинах насіння, може бути екстрагований, і інсулін може застосовуватися для лікування пацієнтів, які страждають на діабет.

Інші характеристики і переваги даного винаходу стають зрозумілі з представленого нижче докладного опису. Повинно бути зрозуміло, що докладний опис і конкретні приклади, які показує переважні варіанти, дані тільки для ілюстрації, оскільки множина змін і модифікацій в межах об'єму даного винаходу стають очевидні фахівцям в даній галузі техніки при прочитанні докладного опису.

Далі описані креслення, що належать до даного винаходу, в яких:

На Фіг.1 зображена нуклеотидна послідовність (SEQ ID NO:1) і виведена послідовність амінокислот (SEQ ID NO:2) злитого білка інсуліну (PRS-D9scFv-Klip27-MI-KDEL) pSBS4404. Розрахункова послідовність амінокислот представлена однобуквеним кодом. Виведена послідовність амінокислот PRS сигнального пептиду показана курсивом, виведена послідовність амінокислот D9scFv показана жирним шрифтом, виведена послідовність амінокислот KLIP 27 підкреслена, виведена послідовність амінокислот міні-інсуліну показана жирним курсивом і послідовність KDEL показана жирним підкресленим шрифтом.

На Фіг.2 зображена нуклеотидна послідовність (SEQ ID NO:3) і виведена послідовність амінокислот (SEQ ID NO:4) злитого білка інсуліну (OLEO-KLIP8-KLIP7-MI) pSBS4405. Розрахункова послідовність амінокислот представлена однобуквеним кодом. Виведена послідовність амінокислот *Arabidopsis thaliana* 18кДа олеозину показана курсивом, виведена послідовність амінокислот KLIP8 показана жирним шрифтом, виведена послідовність амінокислот KLIP27 підкреслена і виведена послідовність амінокислот міні-інсуліну показана жирним курсивом.

На Фіг.3 зображена нуклеотидна послідовність (SEQ ID NO:5) і виведена послідовність амінокислот (SEQ ID NO:6) злитого білка інсуліну (PRS-MI-тетраосновний лінкер-D9Scfv-KDEL) 4414. Розрахункова послідовність амінокислот показана однобуквеним кодом. Виведена послідовність амінокислот PRS сигнального пептиду показана курсивом, виведена послідовність амінокислот міні-інсуліну (B30 тетраосновна) показана жирним шрифтом, виведена послідовність амінокислот тетраосновного лінкера підкреслена, виведена послідовність амінокислот D9scFv показана жирним курсивом і послідовність KDEL показана жирним підкресленим шрифтом.

На Фіг.4 (A-D) показана рекомбінантна експресія злитих білків інсуліну в трансформованих лініях *Arabidopsis thaliana* (4404-2, -17, -20 і 4405-4) на основі забарвленого кумасі SDS-PAGE і вестерн-блот аналізу. Стрілки вказують на положення міграції 38,5кДа і 34,2кДа злитих поліпептидів, PRS-D9(scfv)-KLIP27-MIw/KDEL і OLEO-KLIP8-KLIP27-MI, відповідно, у відновних умовах. На Фіг.4A (забарвлений кумасі гель) і 4B (відповідний вестерн-блот з антиінсуліном E2E3) зображений загальний білок насіння дикого типу (wt) і трансгенного насіння, яке експресує 4404 і 4405 конструкції. На Фіг.4B (забарвлений кумасі гель) і 4D (відповідний вестерн-блот з антиінсуліном E2E3) показаний білок масляного тільця, одержаний з насіння дикого типу і трансгенного насіння, яке експресує 4404 і 4405 конструкції. На Фіг.4 (E-F) показана рекомбінантна експресія злитих білків інсуліну в трансформованих лініях *Arabidopsis thaliana* (4419-9 і 4414-20) на основі забарвленого кумасі SDS-PAGE і вестерн-блот аналізу. Молекулярні маси маркерів (M) дорівнюють 10, 15, 25, 37, 50, 75, 100, 150кДа. Контроль включає hIN (стандарт рекомбінантного людського інсуліну) і hProlIN (стандарт рекомбінантного людського проінсуліну), виділених в невідновних умовах.

На Фіг.5 зображені визначені рівні експресії в доступних лініях насіння T3 (4404-2, -17, -20, 4405-4, -13, -19) і лініях насіння T2 (4414-9 і -20). Рівні трансгену і % молярної експресії MI визначають на основі денситометрії.

На Фіг.6 зображений аналіз забарвленого кумасі SDS-PAGE (15%) препаратів масляних тілець до елюювання (-OB), препаратів OB після елюювання мурашиною кислотою (-OB') і концентрованого елююваного матеріалу (-E). Стрілки вказують положення мігруючого злитого поліпептиду. Контрольна група дикого типу практично не містить які-небудь основні білки після елюювання, в той час як концентрований 4404 матеріал містить злитий білок, деяку кількість укорочених продуктів (можливо, гідролізований злитий білок) і, можливо, деяку кількість альбумінів, елюйованих спільно.

На Фіг.7 зображені хроматограми, які показують характерний час утримування стандарту людського інсуліну в порівнянні з елююваним розщепленим трипсином 4404 злитим білком на колонці C18. Стандарт hIN являє собою стандарт рекомбінантного людського інсуліну (0,5мкг).

На Фіг.8 зображений мас-спектральний аналіз стандарту людського інсуліну (A) в порівнянні з розщепленими трипсином і очищеними BERPX 4404 (B) фракціями, зібраними через 17,0-17,5 хвилин.

На Фіг.9 зображений аналіз забарвленого кумасі SDS-PAGE (15%) загального екстрагованого білка насіння і білка, одержаного з масляного тільця (OB) з колоній, експресуючих 4405 в порівнянні з диким типом (не рекомбінантним) насіння. Стрілка вказує положення мігруючого злитого білка.

На Фіг.10 зображена хроматограма характерного часу утримування стандарту людського інсуліну в порівнянні з розщепленим трипсином препаратом 4405 OB, одержана обернено-фазовою BERPX на колонці C18. Стандарт hIN являє собою стандарт рекомбінантного людського інсуліну (0,5мкг).

На Фіг.11 зображений мас-спектральний аналіз стандарту людського інсуліну (A) в порівнянні з розщепленими трипсином очищеними BERPX 4405 (B) фракціями, зібраними через 17,0-17,5 хвилин.

На Фіг.12 зображена хроматограма розщепленого трипсином препарату 4405 масляного тільця (пунктирна лінія) в порівнянні зі стандартом людського інсуліну (суцільна лінія). Фракції елююваного розщепленого інсуліну збирають між 7-35мC/см і концентрують ліофілізацією для біологічного дослідження інсуліну.

На Фіг.13 зображені зміни рівнів глюкози в сироватці у самців мишей B6 після ін'єкцій негативного контролю (пусті кола= плацебо з фізіологічним розчином, суцільні кола= розщеплені трипсином масляні тільця дикого типу), позитивного контролю (пусті квадрати= Humulin R, пусті трикутники= Roche hIN) в порівнянні з інсуліном, одержаним з рослин з 4405 масляних тілець (суцільні ромби= SDS hIN DesB₃₀).

На Фіг.14 зображене дослідження із забарвленим кумасі гелем білків масляних тілець від двох характерних колоній (4409-6 і 4409-8) в порівнянні з міграцією злитого білка Oleosin-hPIN (показано чорними стрілками) для нетрансформованого (wt) *Arabidopsis*. Рівні експресії визначають денситометрією з

одержанням середніх значень близько 0,10% від загального білка насіння. Цей рівень оцінюють понад і за межами співміграції ендogenous білка з тією ж молекулярною масою з нетрансформованого насіння (wt), який складає приблизно 0,04% від загального білка насіння.

Як зазначено вище, даний винахід належить до поліпшеного способу одержання інсуліну в трансгенних рослинах. Автори даного винаходу несподівано виявили, що рівні акумуляції інсуліну, які перевищують 0,1% від загального білка клітини, можуть бути одержані в рослинах шляхом рекомбінантного одержання інсуліну в насінні рослин. Такі рівні експресії, які принаймні в десять разів більше тих, які були одержані раніше, дозволяють комерційне виробництво інсуліну в рослинах. Одержання в насінні забезпечує гнучкість при зберіганні і транспортуванні інсуліну у вигляді сировини, оскільки інсулін зберігає свою активність при екстрагуванні з насіння, що зберігалось. Більш того кількість біомаси, яку необхідно піддати екстрагуванню, обмежена через відносно низький вміст води в насінні рослин.

Отже, в даному винаході представлений спосіб експресії інсуліну в рослинах, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

(i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин; і

(ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;

(b) введення конструкції химерної нуклеїнової кислоти в клітину рослини; і

(c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін.

Відповідно до даного винаходу було несподівано виявлено, що інсулін акумулюється в насінні рослин в кількостях, які раніше не досягалися, якщо інсулін експресується в насінні таким чином, який дозволяє ізоляцію поліпептиду інсуліну всередині клітин насіння в мембрані, яка оточує внутрішньоклітинний компартмент. Отже, в даному винаході представлений спосіб експресії інсуліну в рослинах, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

(i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин;

(ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну; і

(iii) послідовність нуклеїнової кислоти, яка кодує поліпептид, здатний утримувати поліпептид інсуліну в мембрані, яка оточує внутрішньоклітинний компартмент;

(b) введення конструкції химерної нуклеїнової кислоти в клітину рослини; і

(c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін.

Терміни і визначення

Якщо не визначено інакше, всі технічні і наукові терміни, що застосовуються в даному описі, мають значення, загальноприйняті серед фахівців в галузі техніки, до якої належить даний винахід. Де це можливо, всі патенти, заявки, опубліковані заявки і інші публікації, включаючи послідовності нуклеїнових кислот і поліпептидів з GenBank, SwissPro і інших баз даних, згаданих в даному описі, представлені тут як посилання у всій повноті.

Термін «послідовність нуклеїнової кислоти» в даному описі належить до послідовності нуклеозидних або нуклеотидних мономерів, яка складається з основ, цукрів і міжцукрових (скелетних) зв'язків. Термін також включає модифіковані або заміщені послідовності, які включають не існуючі в природі мономери або їх частини. Послідовності нуклеїнових кислот відповідно до даного винаходу можуть бути послідовностями дезоксирибонуклеїнової кислоти (ДНК) або послідовностями рибонуклеїнової кислоти (РНК), і можуть включати існуючі в природі основи, включаючи аденін, гуанін, цитозин, тимідин і урацил. Послідовності також можуть містити модифіковані основи. Приклади таких модифікованих основ включають аза- і деаза- аденін, гуанін, цитозин, тимідин і урацил; і ксантин і гіпоксантин.

Термін «послідовність нуклеїнової кислоти, яка кодує інсулін» і «послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну», які можуть бути взаємозамінними, стосуються будь-якої і всіх послідовностей нуклеїнових кислот, які кодують поліпептид інсуліну, включаючи поліпептиди інсуліну, перераховані в таблиці (SEQ ID NO:7-145), а також будь-який поліпептид інсуліну ссавця, і будь-яких послідовностей нуклеїнових кислот, які кодують проінсулін і препроінсулін. У даному описі «проінсулін» належить до поліпептиду інсуліну, який включає зв'язуючий пептид або «С-пептид», який зв'язує А і В ланцюги поліпептиду інсуліну. У природному людському інсуліні С-пептид являє собою ланцюг із 31 залишку амінокислоти, яка зв'язує залишок В30 із залишком А1. Термін «препроінсулін» належить до молекули проінсуліну, що додатково містить N-кінцеву сигнальну послідовність, яка направляє виникнення трансляції в ER рибосомах. Послідовності нуклеїнових кислот, які кодують поліпептид інсуліну, також включають будь-яку і всі послідовності нуклеїнових кислот, які (i) кодують поліпептиди, які практично ідентичні послідовності поліпептиду інсуліну, представлений вище; або (ii) гібридизуються з будь-якою послідовністю нуклеїнової кислоти, вказаною вище, в умовах принаймні помірно жорсткої гібридизації, або які будуть гібридизуватися в принаймні помірно жорстких умовах, але для застосування синонімічних кодонів.

Під терміном «практично ідентичний» розуміють те, що дві поліпептидні послідовності переважно ідентичні на принаймні 75% і, більш переважно на принаймні 85%, і найбільш переважно ідентичні на принаймні 95%, наприклад, ідентичні на 96%, 97%, 98% або 99%. Для визначення відсотка ідентичності між двома поліпептидними послідовностями, послідовності амінокислот двох таких послідовностей сполучають, переважно із застосуванням алгоритму Clustal W (Thompson, JD, Higgins DG, Gibson TJ, 1994, Nucleic Acids Res. 22(22): 4673-4680, разом з матрицею кількісної оцінки BLOSUM 62 (Henikoff S, and Henikoff J.G., 1992, Proc. Natl. Acad. Sci. USA 89: 10915-10919) і штрафом розкриття інтервалу 10 і штрафом протяжності інтервалу 0,1, таким чином, що збіг вищого порядку досягається між двома послідовностями, де принаймні 50% від загальної довжини однієї з послідовностей включене у вирівнювання. Інші способи, які можуть застосовуватися для вирівнювання послідовностей, включають спосіб вирівнювання Нідльмана і Вунша (J. Mol. Biol., 1970, 48: 443) виправлений Смітом і Ватерманом (Adv. Appl. Math., 1981, 2: 482) таким чином, що збіг вищого порядку одержують між двома послідовностями, і між двома послідовностями визначають кількість

ідентичних амінокислот. Інші методи розрахунку відсотка збігу між двома амінокислотними послідовностями відомі в даній галузі техніки і включають, наприклад, описані Carlslo і Lipton (SIAM J. Applied Math., 1998, 48: 1073) і описані в Computational Molecular Biology, Lesk, e.d. Oxford University Press, New York, 1988, Bioscomputing: Informatics and Genomics Projects). В основному, для таких розрахунків застосовують програмне забезпечення. Програмне забезпечення, яке може застосовуватися для цієї мети, включає, але не обмежене ними, GCG (Devereux et al. Nucleic Acids Res., 1984, 12: 387), BLASTP, BLASTN і FASTA (Altschul et al., J. Molec. Biol, 1990: 215: 403).

Під «принаймні, помірно жорсткими умовами гібридизації» розуміють умови, які сприяють селективній гібридизації між двома комплементарними молекулами нуклеїнової кислоти в розчині. Гібридизація може відбуватися зі всією або з частиною послідовності молекули нуклеїнової кислоти Гібридизована частина звичайно містить принаймні 15 (наприклад, 20, 25, 30, 40 або 50) нуклеотидів. Фахівець в даній галузі техніки зрозуміє, що стабільність дуплексу або гібриду нуклеїнової кислоти визначається T_m , яка в буферах, що містять натрій, є функцією концентрації іона натрію і температури ($T_m = 81,5^\circ\text{C} - 16,6 (\log_{10}[\text{Na}^+]) + 0,41(\% (\text{G}+\text{C}) - 600/1)$) або подібне рівняння). Отже, параметри умов промивання, які визначають стабільність гібриду, включають концентрацію іона натрію і температуру. Для ідентифікації молекул, які схожі, але не ідентичні, відомій молекулі нуклеїнової кислоти, прийнятна 1% розбіжність, що дає зниження близько 1°C в T_m , наприклад, якщо вважається, що молекули нуклеїнової кислоти ідентичні на $>95\%$, температура кінцевої промивки буде знижена на близько 5°C . Базуючись на цих міркуваннях фахівець в даній галузі техніки легко підбере відповідні умови гібридизації. У переважних варіантах вибирають жорсткі умови гібридизації. Наприклад, наступні умови можуть застосовуватися для одержання жорсткої гібридизації: гібридизація в 5 x хлориді натрію/цитраті натрію (SSC)/5 x розчину Денхардта/0,1% SDS при $T_m - 5^\circ\text{C}$ на основі представленого вище рівняння, з подальшою промивкою 0,2xSSC/0,1% SDS при температурі 60°C . Помірно жорсткі умови гібридизації включають промивання 3xSSC при температурі 42°C . Однак зрозуміло, що еквівалентні обмеження можуть бути досягнуті із застосуванням альтернативних буферів, солей і температур. Додаткові інструкції по проведенню гібридизації можна знайти в Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1-6.3.6 і в Sambrook et al., Molecular Cloning, a Laboratory manual, Cold Spring Harbor Laboratory Press, 1989, Vol.3.

У даному описі терміни «інсулін» і «поліпептид інсуліну», які можуть бути взаємозамінюваними, відносяться до будь-якого і всіх поліпептидів інсуліну, включаючи поліпептиди інсуліну, перераховані в таблиці (SEQ ID NO:7-145), а також до молекули поліпептиду, що містить послідовність залишків амінокислот, які (i) практично ідентичні послідовностям амінокислот, що складають будь-який з представлених тут поліпептидів інсуліну, або (ii) кодованою послідовністю нуклеїнової кислоти, здатної гібридуватися в принаймні помірно жорстких умовах з будь-якою послідовністю нуклеїнової кислоти, що кодує представлений тут інсулін, або здатної гібридуватися в принаймні помірно жорстких умовах з будь-якою послідовністю нуклеїнової кислоти, що кодує представлений тут інсулін, але для застосування синонімічних кодонів. Терміни інсулін і поліпептид інсуліну включають поліпептиди проінсуліну і поліпептиди міні-інсуліну. Поліпептид інсуліну переважно має людське, бичаче або свиняче походження.

Термін «поліпептид, здатний зберігати поліпептид інсуліну в мембрані, яка оточує внутрішньоклітинний компартмент» в даному описі належить до будь-якого поліпептиду, який, будучи зв'язаним з поліпептидом інсуліну, здатний ізолювати поліпептид інсуліну у внутрішньоклітинній структурі, оточеній мембраною, і розташовуватися у внутрішньоклітинному просторі клітини рослини, як визначено плазматичною мембраною клітини рослини.

Термін «поліпептид, здатний утримувати поліпептид інсуліну в ER або ER-похідній накопичувальній везикулі» належить до будь-якого поліпептиду, який, будучи зв'язаним з поліпептидом інсуліну, здатний ізолювати поліпептид інсуліну або в ендоплазматичній сітці, або в накопичувальному компартменті, який є похідним з ендоплазматичної сітки, такому як, наприклад, масляне тільце, в клітині рослини.

Термін «масляне тільце» або «масляні тільця» належить до будь-якої органіли, що накопичує масло або жир, в клітині насінної рослини (описані, наприклад, в: Huang (1992) Ann. Rev. Plant Mol. Biol., 43: 177-200).

Термін «химерна», що застосовується в даному описі в контексті з послідовністю нуклеїнової кислоти, належить до принаймні двох зв'язаних послідовностей нуклеїнових кислот, які не зв'язані в природі. Химерні послідовності нуклеїнових кислот включають зв'язані послідовності нуклеїнових кислот різного природного походження. Наприклад, послідовність нуклеїнової кислоти, що міститься в промоторі рослини, зв'язана з послідовністю нуклеїнової кислоти, що кодує людський інсулін, вважається химерною. Химерні послідовності нуклеїнових кислот також можуть містити послідовності нуклеїнової кислоти одного і того ж природного походження, за умови, що вони не зв'язані в природі. Наприклад, послідовність нуклеїнової кислоти, що міститься в промоторі, одержаному з певного типу клітин, може бути зв'язана з послідовністю нуклеїнової кислоти, що кодує поліпептид, одержаний з того ж типу клітин, але який в природі не зв'язаний з послідовністю нуклеїнової кислоти, що міститься в промоторі. Химерні послідовності нуклеїнових кислот також включають послідовності нуклеїнових кислот, які містять будь-які природні послідовності нуклеїнових кислот, зв'язані з будь-якими не існуючими в природі послідовностями нуклеїнових кислот.

Одержання рекомбінантних векторів експресії, які містять химерні послідовності нуклеїнових кислот, що кодують інсулін, і промотор, здатний контролювати експресію в клітинах насінної рослини

Послідовності нуклеїнових кислот, що кодують інсулін, які можуть застосовуватися у відповідності зі способами і композиціями, представленими в даному описі, можуть включати будь-які послідовності нуклеїнових кислот, які кодують поліпептид інсуліну, включаючи будь-який проінсулін і препрінсулін.

Приклади послідовностей нуклеїнових кислот, що кодують інсулін, добре відомі в даній галузі техніки і звичайно легко доступні з множини різних джерел ссавців, включаючи людину (Bell, G.I. et al., 1980, Nature 284: 26-32), свиню (Chance, R.E. et al., 1968, Science 161: 165-167), бика (D'Agostino, J. et al., 1987, Mol. Endocrinol. 1: 327-331), вівцю (Peterson, J.D. et al., 1972, Biol. Chem. 247: 4866-4671) і подібні, а також з рослинних джерел (Oliveira, A.E.A. et al., 1999, Protein Pept. Lett. 6: 15-21). Послідовності, що кодують інсулін,

які можуть застосовуватися, включають кодуєчі поліпептидні ланцюги, представлені як SEQ ID NO:7-SEQ ID NO:145. Відповідні послідовності нуклеїнових кислот, які кодують поліпептидні ланцюги інсуліну, можуть бути легко ідентифіковані за допомогою номерів ідентифікатора Swiss Protein, представлених в таблиці. Застосовуючи ці послідовності нуклеїнових кислот, додаткові нові послідовності нуклеїнових кислот, які кодують інсулін, можуть бути легко ідентифіковані із застосуванням методик, відомих фахівцям в даній галузі техніки. Наприклад, бібліотеки, такі як експресійні бібліотеки, кДНК і геномні бібліотеки, можуть бути досліджені, і бази даних, які містять інформацію про послідовності з проектів, що встановлюють послідовність, можуть бути перевірені на предмет схожих послідовностей. Можуть застосовуватися альтернативні методи виділення додаткових послідовностей нуклеїнових кислот, які кодують поліпептид інсуліну, і нові послідовності можуть бути виявлені і використані відповідно до даного винаходу. У переважних варіантах послідовності нуклеїнових кислот, які кодують інсулін, включають людський, свинячий і бичачий інсулін.

Множина аналогів інсуліну відома в даній галузі техніки (див., наприклад, патенти США 5461031, 5474987, 5164366, і 5008241) і може застосовуватися відповідно до даного винаходу. Аналоги, які можуть застосовуватися, включають молекули людського інсуліну, в яких залишок амінокислоти 28 ланцюга В (В28) змінений з природного пролінового залишку на аспартат, лізин або ізолейцин. У іншому варіанті лізиновий залишок В29 модифікований в пролін. Далі, аспарагін А21 може бути замінений на аланін, глутамін, глутамат, гліцин, гістидин, ізолейцин, лейцин, метіонін, серин, треонін, триптофан, тирозин або валін. Також аспарагін В3 може бути модифікований в лізин. Інші приклади аналогів інсуліну, які можуть застосовуватися у відповідності з даним винаходом, включають людський інсулін, що не містить залишку В30, також часто званий «desВ30» або «В(1-29)»; інсулін, що не містить останні 3 амінокислотних залишки «В(1-27)»; інсулін, що не містить фенілаланіновий залишок В1; і аналоги, в яких А-ланцюг або В-ланцюг має N-кінцеве або C-кінцеве подовження, наприклад, В-ланцюг може бути подовжений на N-кінці шляхом додавання двох аргінінових залишків.

У переважних варіантах послідовність нуклеїнової кислоти, що кодує інсулін, являє собою проінсулін. У інших переважних варіантах застосовують послідовність нуклеїнової кислоти, що кодує інсулін, в якій С-пептид модифікований відносно його природної форми. Залишки амінокислот в С-пептиді можуть бути замінені, і С-пептид може бути подовжений або укорочений. У цьому випадку термін «міні-інсулін» належить до поліпептиду інсуліну, який був модифікований таким чином, що С-пептид був укорочений відносно його природної форми. У переважних варіантах застосовують міні-інсулін. Переважно, С-пептид молекули міні-інсуліну коротше, ніж 20 залишків амінокислот, більш переважно коротше, ніж 15 залишків амінокислот, і найбільш переважно коротше, ніж 9 залишків амінокислот, наприклад, 7, 5 або 3 залишки. Переважно, як і у випадку з природними молекулами інсуліну, С-пептид міні-інсуліну включає сайт розщеплення на його С- і N-кінцях. Такі сайти розщеплення можуть являти собою будь-які зручні сайти розщеплення, відомі в даній галузі техніки, наприклад, метіонін, що розщеплюється бромідом ціаногену, одиничний основний залишок або пара основних залишків, що розщеплюються трипсином або протеазами, подібними трипсину, або карбоксипептидазою. Наприклад, С-пептид може містити лізин на С-кінці, наприклад, Ala-Ala-Lys (SEQ ID NO:146) або двоосновний сайт обробки безпосередньо перед залишком GlyAl, наприклад, Asn-Lys-Arg (SEQ ID NO:147) або Arg-Arg-Lys-Gln-Lys-Arg (SEQ ID NO:148), або тетраосновний сайт обробки безпосередньо перед залишком Gly A1, наприклад, Arg-Arg-Lys-Arg (SEQ ID NO:149). Отже, молекули міні-інсуліну, які можуть застосовуватися відповідно до даного винаходу, включають:

В(1-29/30)-X₁-X₂-X₃-Y₁A(1-21),

де

X₁ є амінокислотою;

X₂ є амінокислотою;

X₃ є Lys або Arg;

Y₁ є пептидним зв'язком або 1-17 залишками амінокислот;

В(1-29/30) є В-ланцюгом людського інсуліну, що містить залишки амінокислот 1-29 або 1-30; і

A(1-21) є А-ланцюгом людського інсуліну, що містить залишки амінокислот 1-21.

У переважному варіанті X₁ є залишком основної амінокислоти (Lys або Arg) і Y₁ є або пептидним зв'язком, або 1-17 залишками амінокислот, де С-кінцевий залишок є залишком основної амінокислоти (Lys або Arg).

Далі, молекули міні-інсуліну, які можуть застосовуватися відповідно до даного винаходу, включають ті, які представлені формулою:

В(1-27)-X₂-X₃-X₁-Y-A(1-21),

де

X₁ є пептидом із 1-8 залишків амінокислот, що включає принаймні один залишок ароматичної амінокислоти;

X₂ є одним з Pro, Asp, Lys або He в положенні 28 В-ланцюга;

X₃ є одним з Pro, Lys, Ala, Arg або Pro-Thr в положенні 29 В-ланцюга;

Y є Lys або Arg;

В(1-27) є В-ланцюгом людського інсуліну, що містить залишки амінокислот 1-27; і

A(1-21) є А-ланцюгом людського інсуліну, що містить залишки амінокислот 1-21.

Інші приклади молекул нуклеїнових кислот, що кодують поліпептиди міні-інсуліну, які можуть застосовуватися відповідно до даного винаходу, включають описані в Markussen et al., Walter de Gruyter & Co. 1987, in: Peptides pp 189-194; Thim et al., 1989, in: Genetics and molecular biology of industrial microorganisms, American Society for Microbiology pp 322-328; і ті, які представлені в патентах США 4916212, 5324641 і 6521738. Зміни послідовності нуклеїнової кислоти, яка кодує інсулін, для одержання аналогів інсуліну можуть бути зроблені із застосуванням множини методик модифікації нуклеїнових кислот, відомих фахівцям в даній галузі техніки, включаючи, наприклад, сайтспрямований мутагенез, цільовий мутагенез, довільний мутагенез, додавання органічних розчинників, перестановку генів або поєднання цих методів, і інші методики, відомі фахівцям в даній галузі техніки (Shraishi et al., 1988, Arch. Biochem. Biophys, 358: 104-115; Galkin et al., 1997,

Protein Eng. 10: 687-690; Carugo et al., 1997, Proteins 28: 10-28; Hurley et al., 1996, Biochem, 35: 5670-5678; Holmberg et al., 1999, Protein Eng. 12: 851-856).

Відповідно до даного винаходу, несподівано було виявлено, що інсулін акумулюється в насінні рослин в кількостях, які досі не були одержані, якщо інсулін експресується в насінні, переважно таким чином, що поліпептид інсуліну всередині клітини насіння ізольований в мембрані, яка оточує внутрішньоклітинний компартмент. У переважних варіантах даного винаходу поліпептид інсуліну ізольований в ER або ER-похідній накопичувальній везикулі. Для досягнення такого накопичення інсуліну в ER або ER-похідній накопичувальній везикулі, відповідно до даного винаходу, поліпептид, який кодує інсулін, зв'язаний з поліпептидом, який утримує поліпептид інсуліну в ER або ER-похідній накопичувальній везикулі, замість транспортування з ER в, наприклад, апопласт. Поліпептиди, які можуть застосовуватися відповідно до даного винаходу для утримування поліпептиду інсуліну в ER, включають будь-які поліпептиди, здатні ізолювати інсулін в ER. Такі поліпептиди можуть бути синтезовані або одержані з будь-яких біологічних джерел. У переважному варіанті даного винаходу поліпептиди, які здатні утримувати інсулін, включають поліпептиди, що містять С-кінцевий мотив утримування в ER. Приклади таких С-кінцевих мотивів утримування в ER включають KDEL, HDEL, DDEL, ADEL і SDEL послідовності (SEQ ID NO:150-154, відповідно). Інші приклади включають HDEF (SEQ ID NO:155) (Lehmann et al., 2001, Plant Physiol. 127(2): 436-49) або два аргінінових залишки, близьких до N-кінця, розташовані в положеннях 2 і 3, 3 і 4 або 4 і 5 (Abstract from Plant Biology 2001 Program, ASPB, July 2001, Providence, Rhode Island, USA). Послідовності нуклеїнових кислот, які кодують С-кінцевий мотив ER утримування, переважно зв'язані з послідовністю нуклеїнової кислоти, що кодує поліпептид інсуліну таким чином, що поліпептид, здатний утримувати інсулін в ER, зв'язаний з С-кінцем поліпептиду інсуліну.

Для досягнення накопичення поліпептиду інсуліну в ER-похідній накопичувальній везикулі, поліпептид інсуліну зв'язаний з поліпептидом, який здатний утримувати поліпептид інсуліну в ER-похідній накопичувальній везикулі. Поліпептид, здатний утримувати поліпептид інсуліну в ER-похідній накопичувальній везикулі, який може застосовуватися відповідно до даного винаходу, включає будь-який поліпептид, здатний ізолювати поліпептид інсуліну в ER-похідній накопичувальній везикулі. Поліпептиди, здатні утримувати інсулін в ER-похідній накопичувальній везикулі, можуть бути синтезовані або одержані з будь-якого біологічного джерела. У переважному варіанті ER-похідна накопичувальна везикула являє собою масляне тільце, і поліпептид інсуліну зв'язаний з білком масляного тільця або його значною частиною, здатною утримувати поліпептид інсуліну в ER-похідній накопичувальній везикулі. Білки масляного тільця, які можуть застосовуватися відповідно до даного винаходу, включають будь-які білки, які природним чином зв'язані з масляним тільцем. Особливо перевалені білки масляного тільця включають олеозини, наприклад, олеозин *Arabidopsis* (van Rooijen et al., (1991) Plant Mol. Biol. 18: 1177-1179), кукурудзяний олеозин (Bowman-Vance et al., 1987, J. Biol. Chem. 262: 11275-11279; Qu et al., 1990, J. Biol. Chem. 265: 2238-2243), олеозин моркви (Hatzopoulos et al., (1990) Plant Cell 2: 457-457) або олеозин *Brassica* (Lee et al., 1991, Plant Physiol. 96: 1395-1397), калеозини (див., наприклад, Genbank, номер доступу AF067857) і стеролеозини (Lin et al., 2002 Plant Physiol. 128(4): 1200-11). В переважному варіанті білком масляного тільця є рослинний олеозин, який має схожість послідовності з іншими рослинними олеозинами, такими як олеозин, виділений з *Arabidopsis thaliana* (SEQ ID NO:156) або *Brassica napus* (SEQ ID NO:157). У іншому варіанті білком масляного тільця є калеозин або зв'язуючий кальцій білок з рослинного, грибового або іншого джерела, яке відрізняється гомологією послідовності з рослинними калеозинами, такими як калеозин, виділений з *Arabidopsis thaliana* (SEQ ID NO:158 і SEQ ID NO:159). У іншому варіанті білком масляного тільця є стерол еозин (SEQ ID NO:160), дегідрогеназа, яка зв'яже стерин (Lin L-J et al., (2002) Plant Physiol 128: 1200-1211). Поліпептид, що кодує інсулін, може бути зв'язаний з білком масляного тільця через N-кінець, також як і через С-кінець, і через фрагменти білка масляного тільця, такі як, наприклад, центральний домен олеозину. Нові білки масляного тільця можуть бути виявлені, наприклад, при одержанні масляних тілець (методики одержання масляних тілець див., наприклад, в патенті США 6650554) і ідентифікації білків в препаратах масляних тілець за допомогою, наприклад, SDS електрофорезу в гелі. Проти таких білків можуть бути одержані поліклональні антитіла, які застосовують для скринінгу бібліотек кДНК для ідентифікації послідовностей нуклеїнових кислот, що кодують білки масляних тілець. Нові білки масляних тілець також можуть бути одержані із застосуванням відомих послідовностей нуклеїнових кислот, що кодують білки масляних тілець, із застосуванням, наприклад, вказаних тут послідовностей білків масляних тілець, що кодують білки масляних тілець, для дослідження, наприклад, кДНК або геномних бібліотек на присутність білків масляних тілець.

Поліпептиди, здатні утримувати інсулін в ER або ER-похідній накопичувальній органелі, звичайно не відщеплюються, і інсулін може накопичуватися у вигляді злитого білка що, наприклад, звичайно відбувається, якщо KDEL сигнал утримування застосовується для утримування поліпептиду в ER, або якщо білок масляного тільця застосовується для утримування поліпептиду в ER-похідній накопичувальній органелі.

Химерна послідовність нуклеїнової кислоти додатково може містити нуклеотидну послідовність, яка націлює послідовність нуклеїнової кислоти на ендомембранну систему («сигнальний пептид»). У варіантах даного винаходу, в яких поліпептид інсуліну утримується в ER за допомогою послідовності, здатної утримувати поліпептид в ER, такий як KDEL, HDEL або SDEL поліпептид, особливо бажано включати послідовність нуклеїнової кислоти, що кодує сигнальний пептид. Приклади сигнальних пептидів, які можуть застосовуватися відповідно до даного винаходу, включають сигнальну послідовність білка, зв'язаного з патогенезом тютюну (PR-S) (SEQ ID NO:161) (Sijmons et al., 1990, Bio/technology, 8: 217-221), сигнальну послідовність лектину (Boehn et al., 2000, Transgenic Res, 9(6): 477-86), сигнальну послідовність насиченого гідроксипроліном глікопротеїну з *Phaseolus vulgaris* (Yan et al., 1997, Plant Physiol. 115(3): 915-24 і Corbin et al., 1987, Mol. Cell Biol 7(12): 4337-44), сигнальну послідовність пататину картоплі (Iturriaga, G et al., 1989, Plant Cell 1: 381-390 і Beyan et al., 1986, Nuc. Acids Res. 41: 4625-4638) і сигнальну послідовність альфа-амілази ячменю (Rasmussen and Johansson, 1992, Plant Mol. Biol. 18(2): 423-7). Такі цільові сигнали можуть бути *in vivo* відщеплені від послідовності інсуліну, що звичайно відбувається коли застосовується цільовий сигнал апопласту, такий як сигнальна послідовність S-білка, зв'язаного з патогенезом тютюну (PR-S) (Sijmons et al.,

1990, Bio/technology, 8: 217-221). Інші сигнальні пептиди можуть бути визначені за допомогою сервера SignalP World Wide Web (<http://www.cbs.dtu.dk/services/SignalP>), в якому можна визначити присутність і розташування сайтів відщеплення сигнального пептиду в послідовностях амінокислот різних організмів. Загалом, існує незначне збереження первинної послідовності амінокислоти, хоч загальні фізико-хімічні властивості зберігаються до деякої межі. Загальна структура сигнальних пептидів має три області, коротку аміно-кінцеву «п-область», що містить позитивно заряджені залишки, центральну гідрофобну «h-область», що змінюється в розмірі від 7 до 15 амінокислот, і карбокси-кінцеву «с-область», що містить полярні амінокислоти і сайт розщеплення, який розпізнається мембранозв'язаними сигнальними ферментами пептидази (Nakai K., 2000, *Advances in Protein Chem* 54: 277-344). Цільовий сигнал, який також може застосовуватися відповідно до даного винаходу, включає природну сигнальну послідовність інсуліну (довжина людської послідовності становить 24 амінокислоти). У переважних варіантах розташовану на N-кінці цільову послідовність апопласту, таку як вказана вище PR-S послідовність тютюну, застосовують в поєднанні з розташованою на C-кінці утримуючою в ER послідовністю, такою як послідовність KDEL.

У іншому переважному варіанті послідовність нуклеїнової кислоти, яка кодує лідерну послідовність α -фактора дріжджів, зв'язана з N-кінцем послідовності нуклеїнової кислоти, що кодує інсулін. Лідерні послідовності дріжджів або послідовності, одержані з лідерних послідовностей дріжджів, які можуть застосовуватися відповідно до даного винаходу, включають послідовності, перераховані в SEQ ID NO:162-SEQ ID NO:171 (Kjeldsen et al., 2001, *Biotechnology and Genetic Engineering Reviews* 18: 89-121). Такі лідерні послідовності можуть також містити пептид спейсеру, розташований на C-кінці нуклеїнової кислоти, що кодує лідерну послідовність, і на N-кінці послідовності, що кодує інсулін. Відповідно до даного винаходу, такі послідовності спейсерів звичайно мають від 2 до 20 амінокислот в довжину. Таким чином, наприклад, можуть застосовуватися послідовності спейсерів SEQ ID NO:172 і SEQ ID NO:173 (Kjeldsen et al., 2001, *Biotechnology and Genetic Engineering Reviews* 18: 89-121). В варіантах даного винаходу, в яких застосовується лідерна послідовність дріжджів, послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну, переважно являє собою поліпептид міні-інсуліну. Відповідно до даного винаходу, в особливо переважному варіанті застосовують послідовність нуклеїнової кислоти, яка кодує антитіло з одним ланцюгом, зв'язаним з послідовністю нуклеїнової кислоти, що кодує лідерний пептид секреції дріжджів, як далі описано в прикладі 1.

Химерна послідовність нуклеїнової кислоти також може містити поліпептиди, що дають стабілізуючі подовження білка на N- і/або C-кінцях. Такі подовження можуть застосовуватися для стабілізації і/або допомагають згортанню ланцюга поліпептиду інсуліну, і додатково можуть застосовуватися для полегшення очищення інсуліну. Поліпептидні подовження, які можуть застосовуватися для цієї мети, включають, наприклад, послідовність нуклеїнової кислоти, яка кодує антитіло з одним ланцюгом, нуклеїнову кислоту, яка кодує молекулу Affibody® (Affibody AB), послідовність нуклеїнової кислоти, яка кодує нетоксичну В субодиницю токсину холери (CTB) (Arakawa, T. et al., 1998, *Nat. Biotechnol.* 16: 938) або поєднання таких поліпептидів. У особливо переважному варіанті, поліпептид інсуліну утримується в мембрані, яка оточує компартмент, такий як ER, за допомогою, наприклад, послідовності KDEL, як описано вище, об'єднаної із стабілізуючим поліпептидом, який дозволяє об'єднання поліпептиду інсуліну з масляним тільцем при порушенні цілісності клітини рослини, що може статися, якщо поліпептид інсуліну виділяють з клітини рослини. Прикладом такого стабілізуючого поліпептиду є антитіло з одним ланцюгом зі специфічністю до масляного тільця. Послідовності нуклеїнових кислот, які кодують антитіла з одним ланцюгом зі специфічністю до масляного тільця, можуть бути одержані з колонії клітин гібридами, експресуючих моноклональні антитіла, діючих проти білка масляного тільця. У одному варіанті, антитіло з одним ланцюгом специфічно зв'язується з олеозином, як описано в Altling-Mees et al. (2000) IBC's International Conference on Antibody Engineering, Poster # 1. Цей варіант даного винаходу більш детально описаний в прикладі 1 нижче.

У іншому варіанті сайт розщеплення може бути розташований вище N-кінця і нижче C-кінця інсуліну, що дозволяє поліпептиду інсуліну відщеплюватися від партнера злиття, тим самим, одержуючи виділений інсулін. Приклади таких сайтів розщеплення можуть бути знайдені в WO 98/49326 (спосіб розщеплення злитих білків) і родинних заявках, і у LaVille et al. (1994) *Enzymatic and chemical cleavage of fusion proteins* In *Current Protocols in Molecular Biology* pp 16.4.5-16.4.17, John Wiley and Sons, Inc., New York NY. У переважному варіанті сайтом розщеплення є тетраосновний лінкер (наприклад, Arg-Arg-Lys-Arg - SEQ ID NO:149), який розщеплюється трипсином. У іншому переважному варіанті сайтом розщеплення є KLIP 8 (SEQ ID NO:174), який розщеплюється аспартат протеазою, включаючи хімозин.

У даному винаході також представлені способи відділення гетерологічних білків від компонентів клітинно-хазяїна розділенням фракції масляного тільця і подальшим виділенням гетерологічного білка шляхом специфічного розщеплення гетерологічного білка - злиттям білка масляного тільця. Необов'язково сайт розщеплення може бути розташований вище N-кінця і нижче C-кінця гетерологічного поліпептиду, що дозволяє злитому поліпептиду розщеплюватися і відділятися шляхом розділення фаз на пептиди, що їх складають.

Послідовність нуклеїнової кислоти, що кодує інсулін, може бути змінена для подальшого поліпшення рівнів експресії, наприклад, оптимізацією послідовності нуклеїнової кислоти відповідно до переважного застосування кодону для певного типу клітини рослини, яку вибирають для експресії поліпептиду інсуліну, або зміною мотивів, які відомі як дестабілізуючі мРНК (див., наприклад, заявку на патент PCT 97/02352). Порівняння застосування кодону послідовності нуклеїнової кислоти, яка кодує поліпептид інсуліну, і застосування кодону клітини рослини дозволяє ідентифікацію кодонів, які можуть бути змінені. Побудова синтетичних генів шляхом зміни застосування кодону описана, наприклад, в заявці на патент PCT 93/07278.

У переважному варіанті застосовувана послідовність нуклеїнової кислоти, яка кодує інсулін, представлена SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 або SEQ ID NO:195.

Відповідно до даного винаходу, послідовність нуклеїнової кислоти, яка кодує інсулін, зв'язана з промотором, здатним контролювати експресію поліпептиду інсуліну в клітині насіння рослини. Отже, в даному винаході також представлена послідовність нуклеїнової кислоти, яка кодує інсулін, зв'язана з промотором,

здатним контролювати експресію в клітині насіння рослини. Застосовувати тут промотори звичайно є відомими в даній галузі техніки промоторами і включають будь-який промотор рослинного походження, здатний контролювати експресію поліпептидів в рослинах. Загалом, промотори, одержані з дводольних рослин, застосовують, якщо дводольну рослину вибирають відповідно до даного винаходу, а промотор однодольної рослини використовують, якщо вибирають однодольну рослину. Основні промотори, які можуть застосовуватися, включають, наприклад, промотор вірусу мозаїки цвітної капусти 35S (CaMV) (Rothstein et al., 1987, Gene 53: 153-161), промотор актину рису (McElroy et al., 1990, Plant Cell 2: 163-171; патент США 6429357), промотор убіквітину, такий як промотор убіквітину кукурудзи (патенти США 5879903; 5273894) і промотор убіквітину петрушки (Kawalleck, P. et al., 1993, Plant Mol. Biol. 21: 673-684).

У переважних варіантах промотором, що застосовується, є промотор, який дає переважну експресію поліпептиду інсуліну в тканині насіння. «Переважними для насіння промоторами» в цьому значенні є промотори, які контролюють експресію рекомбінантного білка (а саме інсуліну) таким чином, що переважно принаймні 80% від загальної кількості рекомбінантного білка, присутнього у дорослій рослині, було б присутнім в насінні. Більш переважно принаймні 90% від загальної кількості рекомбінантного білка, присутнього у дорослій рослині, мають бути присутні в насінні. Найбільш переважно принаймні 95% від загальної кількості рекомбінантного білка, присутнього у дорослій рослині, мають бути присутні в насінні. Переважні для насіння промотори, які можуть застосовуватися в цьому значенні, включають, наприклад, промотор фазеоліну бобів (Sengupta-Gopalan et al., 1985, Proc. Natl. Acad. Sci. USA 82: 3320-3324); промотор олеозину Arabidopsis 18кДа (патент США 5792922) або промотор олеозину льону (WO 01/16340); промотор акумулюючого білка (лініну) насіння льону, подібного легуміну (WO 01/16340); промотор акумулюючого білка 2S льону (WO 01/16340); переважний промотор ендосперму, такий як промотор Amy32b (Rogers and Millman, J. Biol. Chem., 1984, 259: 12234-12240), промотор Amy6-4 (Kursheed and Rogers, J. Biol. Chem., 1988, 263: 18953-18960) або промотор Алеураїну (Whitoer et al., 1987, Nucleic Acids Res., 15: 2515-2535), або промотор арселіну бобів (Jaeger GD, et al., 2002, Nat. Biotechnol. Dec; 20: 1265-8). Постійно знаходять нові промотори, які застосовуються в різних рослинах. Множину прикладів промоторів рослин можна знайти у Ohamuro et al. (Biochem. of Plants, 1989, 15: 1-82).

У даному винаході можуть застосовуватися певні генетичні елементи, здатні поліпшувати експресію поліпептиду інсуліну. Ці елементи включають нетрансльовані лідерні послідовності певних вірусів, такі як лідерна послідовність AMV (Jobling and Gehrke, 1987, Nature, 325: 622-625) і інтрон, зв'язаний з промотором убіквітину кукурудзи (патент США 5504200). Звичайно химерну послідовність нуклеїнової кислоти одержують таким чином, щоб генетичні елементи, здатні посилювати експресію, були розташовані в положенні 5' послідовності нуклеїнової кислоти, яка кодує поліпептид інсуліну.

Відповідно до даного винаходу, химерні послідовності нуклеїнових кислот, які містять промотор, здатний контролювати експресію в насінні рослин, зв'язаний з послідовністю нуклеїнової кислоти, яка кодує поліпептид інсуліну, можуть бути інтегровані в рекомбінантний вектор експресії, який забезпечує хорошу експресію в клітині насіння. Отже, даний винахід включає рекомбінантний вектор експресії, який містить в напрямі 5'-3' транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітині насіння рослини; і
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;

де вектор експресії придатний для експресії в клітині рослини. Термін «придатний для експресії в клітині рослини» означає, що рекомбінантний вектор експресії містить химерну послідовність нуклеїнової кислоти відповідно до даного винаходу, зв'язану з генетичними елементами, необхідними для досягнення експресії в клітині насіння. Генетичні елементи, які можуть бути включені у вектор експресії в цьому значенні, включають транскрипційну область термінації, одну або більше послідовностей нуклеїнової кислоти, яка кодує маркерні гени, одне або більше начал реплікації і подібні. У переважних варіантах, вектор експресії також містить генетичні елементи, необхідні для інтеграції вектора або його частини в ядерний геном клітини рослини, наприклад, ліву і праву крайні послідовності Т-ДНК, які сприяють інтеграції в ядерний геном рослини у варіантах відповідно до даного винаходу, в яких клітини рослин трансформують з допомогою *Agrobacterium*.

Як зазначено вище, рекомбінантний вектор експресії звичайно містить транскрипційний термінатор, який, крім того, що він служить як сигнал для закінчення транскрипції, також може служити як захисний елемент, здатний продовжувати період напіврозпаду мРНК (Guarneros et al., 1982, Proc. Natl. Acad. Sci. USA, 79: 238-242). Термінатор транскрипції звичайно містить від близько 200 нуклеотидів до близько 1000 нуклеотидів, і вектор експресії одержують таким чином, що термінатор транскрипції розташований в положенні 3' послідовності нуклеїнової кислоти, яка кодує інсулін. Кінцеві послідовності, які можуть застосовуватися відповідно до даного винаходу, включають, наприклад, кінцеву область нопаліну (Bevan et al, 1983, Nucl. Acids. Res., 11: 369-385), термінатор фазеоліну (van der Geest et al., 1994, Plant J. 6: 413-423), термінатор арселіну (Jaeger GD, et al., 2002, Nat. Biotechnol. Dec; 20: 1265-8), термінатор для генів октопінсинтази *Agrobacterium tumefaciens* або інших подібних функціональних елементів. Транскрипційні термінатори можуть бути одержані як описано у An (An, 1987, Methods in Enzym. 153: 292).

Відповідно до даного винаходу, вектор експресії також може містити маркерний ген. Маркерні гени, які можуть застосовуватися відповідно до даного винаходу, включають всі гени, які дозволяють відрізнити трансформовані клітини від нетрансформованих клітин, включаючи всі селектовані і скринабельні маркерні гени. Маркерний ген може бути маркером стійкості, таким як маркер стійкості до антибіотиків, наприклад, канаміцину (патент США 6174724), ампіциліну, G418, блеоміцину, гігроміцину, що дозволяє селекцію характерних особливостей хімічними методами, або маркери, толерантні до хімічного агента, такого як цукор маноза із звичайною фітотоксичністю (Negrotto et al., 2000, Plant Cell Rep. 19: 798-803). Інші відповідні маркери, які можуть застосовуватися відповідно до даного винаходу, включають маркери, здатні передавати стійкість до гербіцидів, таких як гліфосат (патенти США 4940935; 5188642), фосфінотрицин (патент США 5879903) або сульфонілсечовини (патент США 5633437). Маркери стійкості, будучи тісно зв'язаними з послідовністю нуклеїнової кислоти, яка кодує поліпептид інсуліну, можуть застосовуватися для збереження

тиску відбору на популяцію клітин рослини або рослини, які не втратили послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну. Скринабельні маркери, які можуть застосовуватися для ідентифікації трансформантів шляхом візуальної перевірки, включають β -глюкуронідазу (GUS) (патенти США 5268463 і 5599670) і зелений флуоресцентний білок (GFP) (Niedz et al., 1995, Plant Cell Rep., 14: 403).

Рекомбінантні вектори, відповідні для введення послідовностей нуклеїнових кислот в рослини, включають вектори на основі *Agrobacterium* і *Rhizobium*, такі як Ti і Ri плазмиди, включаючи, наприклад, pBIN19 (Bevan, Nuci. Acid. Res., 1984, 22: 8711-8721), pGKB5 (Bouchez et al., 1993, C R Acad. Sci. Paris, Life Sciences, 316: 1188-1193), ряд бінарних векторів pCGN (McBride and Summerfelt, 1990, Plant Mol. Biol., 14: 269-276) і інші бінарні вектори (наприклад, патент США 4940838).

Рекомбінантні вектори експресії, послідовності нуклеїнових кислот і химерні послідовності нуклеїнових кислот відповідно до даного винаходу можуть бути одержані відповідно до методик, добре відомих фахівцям в галузі молекулярної біології. Таке одержання звичайно включає види бактерій *Escherichia coli* як проміжний клонуєчий хазяїн. Одержання векторів *E. coli*, а також векторів трансформації рослин може бути здійснено із застосуванням відомих методик, таких як рестрикційне розщеплення, лігування, гелевий електрофорез, секвентування ДНК, полімеразна ланцюгова реакція (PCR) і інші методики. Ці методики дозволяють зв'язувати послідовності нуклеїнових кислот і поліпептиди, що належать до даного винаходу. Великий вибір векторів клонування доступний для здійснення необхідних стадій, необхідних для одержання рекомбінантного вектора експресії. Серед векторів з системою реплікації, що функціонує в *E. coli*, є вектори, такі як pBR322, ряд векторів pUC, ряд векторів M13mp, pBluescript і т.д. Звичайно такі вектори клонування містять маркер, що дозволяє відбір трансформованих клітин. Послідовності нуклеїнових кислот можуть бути введені в ці вектори, і вектори можуть бути введені в *E. coli*, вирощену у придатному середовищі. Рекомбінантні вектори експресії можуть бути легко виділені з клітин при зборі і лізисі клітин. Загальні інструкції по одержанню рекомбінантних векторів представлені в, наприклад: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, том 3.

Одержання рослин, що містять насіння, здатне експресувати інсулін

Відповідно до даного винаходу, химерну послідовність нуклеїнової кислоти вводять в клітину рослини, і клітини вирощують у дорослих рослинах, здатних давати насіння, де насіння експресує поліпептид інсуліну

Відповідно до даного винаходу, може бути вибраний будь-який вид рослини або клітин рослини. Конкретні клітини, що застосовуються відповідно до даного винаходу, включають клітини, одержані з *Arabidopsis thaliana*, бразильського горіха (*Betholettia excelsa*); рицини звичайної (*Ricinus communis*), кокоса (*Cocos nucifera*); коріандру (*Coriandrum sativum*); бавовни (під *Gossypium*); арахісу (*Arachis hypogaea*); жожоба (*Simmondsia chinensis*); льняного насіння/льону (*Linum usitatissimum*); кукурудзи (*Zea mays*); гірчиці (під *Brassica* і *Sinapis alba*); гвінейської олійної пальми (*Elaeis guineensis*); оливи (*Olea europaea*); рапсу (вигляд *Brassica*); рису (*Oryza sativa*); сафлору (*Carthamus tinctorius*); сої (*Glycine max*); гарбуза (*Cucurbita maxima*); ячменю (*Hordeum vulgare*); пшениці (*Triticum aestivum*) і соняшника (*Helianthus annuus*).

Відповідно до даного винаходу, в переважному варіанті застосовують рослини або клітини рослин з насіння олійних рослин. Олійні рослини, насіння яких застосовується відповідно до даного винаходу, включає арахіс (*Arachis hypogaea*); гірчицю (під *Brassica* і *Sinapis alba*); рапс (вигляд *Brassica*); нут (*Cicer arietinum*), сою (*Glycine max*); бавовну (*Gossypium hirsutum*); соняшник (*Helianthus annuus*); (*Lentil Lens culinaris*); льняне насіння/льон (*Linum usitatissimum*); білу конюшину (*Trifolium repens*); оливу (*Olea europaea*); гвінейську олійну пальму (*Elaeis guineensis*); сафлор (*Carthamus tinctorius*) і боби нарбону (*Vicia narbonensis*).

Відповідно до даного винаходу, в особливо переважному варіанті застосовують сафлор, *Arabidopsis* або льон.

Методики введення рослинних рекомбінантних векторів експресії в клітину рослини, також позначеної далі як «трансформація», добре відомі фахівцям в даній галузі техніки, і звичайно залежать від вибраної клітини рослини. Загальні методики введення рекомбінантних векторів експресії включають електропорацію; методики, опосередковані хімічними сполуками, наприклад, поглинання нуклеїнової кислоти, опосередковане CaCl_2 ; бомбардування частинками (біолістику); застосування природних інфекційних послідовностей нуклеїнових кислот, наприклад, послідовності нуклеїнових кислот, одержані з вірусів, або послідовності, одержані з *Agrobacterium* або *Rhizobium*, поглинання нуклеїнової кислоти, опосередковане поліетиленгліколем (ПЕГ), мікроін'єкції і застосування ниткоподібних кристалів карбиду кремнію.

У переважних варіантах вибирають таку методику трансформації, яка дозволяє інтеграцію химерної послідовності нуклеїнової кислоти в геном клітини рослини, і, переважно, ядерний геном клітини рослини. Відповідно до даного винаходу, це вважається особливо бажаним, оскільки застосування такої методики дозволяє перенесення химерної послідовності нуклеїнової кислоти в рослини потомства при статевій репродукції. Методи трансформації, які можуть застосовуватися відповідно до даного винаходу, включають методи, опосередковані біолістиками і *Agrobacterium*.

Методики трансформації для дводольних видів рослин добре відомі. Звичайно застосовують трансформацію, опосередковану *Agrobacterium*, завдяки її високій ефективності, а також загальну сприйнятливості більшістю, якщо не всіма, дводольними видами рослин. Трансформація з *Agrobacterium* звичайно включає перенесення бінарного вектора, такого, як будь-якого із згаданих вище бінарних векторів, що включає химерну послідовність нуклеїнової кислоти відповідно до даного винаходу, з *E. coli* у відповідний штам *Agrobacterium* (наприклад, EHA101 і LBA4404) за допомогою, наприклад, схрещування трьох батьків зі штамом *E. coli*, що містить рекомбінантний бінарний вектор, і штамом *E. coli*, що містить допоміжну плазмиду, яка мобілізує бінарний вектор по відношенню до цільового штаму *Agrobacterium*, або трансформації ДНК штаму *Agrobacterium* (Hofgen et al., Nucl. Acids Res., 1988, 16: 9877). Інші методики, які можуть застосовуватися для трансформації клітин дводольних рослин, включають біолістики (Sanford, 1988, Trends in Biotechn. 6: 299-302); електропорацію (Fromm et al., 1985, Proc. Natl. Acad. Sci. USA., 82: 5824-5828); опосередковане ПЕГ поглинання ДНК (Potrykus et al., 1985, Mol. Gen. Genetics, 199: 169-177); мікроін'єкції (Reich et al., Bio/Techn., 1986, 4: 1001-1004); і ниткоподібні кристали карбиду кремнію (Kaepler et al., 1990,

Plant Cell Rep., 9: 415-418) або трансформацію in planta із застосуванням, наприклад, методики занурення квітки (Clough and Bent, 1998, Plant J., 16: 735-743).

Однодольні види рослин можуть бути трансформовані із застосуванням різних методик, включаючи бомбардування частинками (Christou et al., 1991, Biotechn. 9: 957-962; Weeks et al., Plant Physiol, 1993, 102: 1077-1084; Gordon-Kamm et al., Plant Cell, 1990, 2: 5603-618); опосередковане ПЕГ поглинання ДНК (європейські патенти 0292 435; 0392 225) або трансформацію, опосередковану *Agrobacterium* (Goto-Fumiyuki et al., 1999, Nature-Biotech. 17: 282-286).

Конкретна методика трансформації рослини може варіюватися в залежності від вигляду рослини і типу клітини рослини (наприклад, одержані з розсади типи клітин, такі як гіпокотилі і котилідони або тканина ембріона), які вибирають як клітини-мішені для трансформації. Як зазначено вище, в особливо переважному варіанті застосовують сафлор, *Arabidopsis* або льон. Методика одержання трансформатів сафлору описана у Baker and Dyer (Plant Cell Rep., 1996, 16: 106-110). Інші протоколи трансформації певних видів рослин можуть бути знайдені в: *Biotechnology in Agriculture and Forestry 46: Transgenic Crops I* (Y.P.S. Bajaj ed.), Springer-Verlag, New York (1999) і *Biotechnology in Agriculture and Forestry 47: Transgenic Crops II* (Y.P.S. Bajaj ed.), Springer-Verlag, New York (2001).

Для трансформації клітини рослин вирощують, і при появі диференціюючої тканини, такої як паростки і коріння, дорослі рослини регенерують. Звичайно регенерують множину рослин. Методики регенерації рослин залежать від вигляду рослин і типу клітин, і відомі фахівцям в даній галузі техніки. Подальші вказівки по культивуванню тканин рослин можна знайти в, наприклад, *Plant Cell and Tissue Culture*, 1994, Vasil and Thorpe Eds., Kluwer Academic Publishers; і в: *Plant Cell Culture Protocols (Methods in Molecular Biology 111)*, 1999, Hall Eds., Humana Press.

У одному аспекті даний винахід належить до способу виділення насіння рослин, яке містить інсулін. Отже, даний винахід належить до способу одержання насіння рослин, яке містить інсулін, який включає

(а) одержання конструкції химерної нуклеїнової кислоти, що містить в 5'-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин; і
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини;
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння; і
- (d) одержання насіння з вказаних рослин, де насіння містить інсулін.

У переважних варіантах множину трансформованих рослин одержують, вирощують і досліджують на присутність бажаної химерної послідовності нуклеїнової кислоти, присутність якої в передбачуваних трансформатах може бути протестована, наприклад, вирощуванням в селективному середовищі, в якому застосовують маркери, стійкі до гербіцидів, шляхом безпосереднього нанесення гербіциду на рослину, або саузерн-блотингом. Якщо визначається присутність химерної послідовності нуклеїнової кислоти, трансформовані рослини можуть бути відібрані для вирощування потомства і, зрештою, дорослих рослин, що містять множину насіння, яке містить бажану послідовність химерної нуклеїнової кислоти. Таке насіння може застосовуватися для виділення інсуліну, або воно може бути висаджене для одержання двох або більше наступних поколінь. Звичайно бажано висаджувати множину трансгенного насіння для одержання трансгенних рослин, кожна з яких містить насіння, яке містить химерну послідовність нуклеїнової кислоти, яка кодує інсулін. Більш того звичайно бажано пересвідчитися в гомозиготності рослин для того, щоб пересвідчитися в безперервному успадкуванні рекомбінантного поліпептиду. Методи відбору гомозиготних рослин добре відомі фахівцям в даній галузі техніки. Методи одержання гомозиготних рослин, які можуть застосовуватися відповідно до даного винаходу, включають одержання і трансформацію гаплоїдних клітин або тканин з подальшою регенерацією гаплоїдних рослин і подальшим перетворенням в диплоїдні рослини, наприклад, обробкою коліном або іншими агентами, які ушкоджують мікротрубочки. Рослини можна вирощувати відповідно до звичайної сільськогосподарської практики.

У іншому аспекті даний винахід належить до рослин, здатних давати насіння, яке експресує інсулін. У переважному варіанті даного винаходу рослини, здатні давати насіння, містять послідовність химерної нуклеїнової кислоти, яка містить в 5'-3' напрямку транскрипції:

(а) першу послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин, з якими вона функціонально зв'язана; і

(b) другу послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну.

У переважному варіанті химерна послідовність нуклеїнової кислоти стабільно інтегрована в ядерний геном рослини.

У ще одному варіанті даний винахід належить до насіння рослин, яке експресує інсулін. У переважному варіанті даного винаходу насіння рослин містить химерну послідовність нуклеїнової кислоти, яка містить в 5'-3' напрямку транскрипції:

(а) першу послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин, з якими вона функціонально зв'язана; і

(b) другу послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну.

Відповідно до даного винаходу одержують насіння, в якому переважно принаймні 0,1% від загального розчинного білка, присутнього в насінні, складає інсулін. У іншому переважному варіанті даного винаходу одержують насіння, в якому принаймні 0,2%, 0,3%, 0,5% або 1,0% від загального розчинного білка, присутнього в насінні, складає інсулін. Поліпептид інсуліну може бути присутнім в різних типах клітин, включаючи, наприклад, гіпокотилі і вісь ембріона, включаючи коріння ембріона і листя ембріона, де однодольні види рослин, включаючи злаки і кукурудзу, застосовують в тканині ендосперму.

Одержання інсуліну з насіння рослин

Після одержання насіння рослини білок інсуліну може бути виділений з насіння за допомогою будь-якої методики очищення білка, відомої в даній галузі техніки. Отже, даний винахід належить до способу виділення

інсуліну з насіння рослин, де спосіб включає

(а) одержання конструкції химерної нуклеїнової кислоти, яка містить в 5-3' напрямку транскрипції у вигляді функціонально зв'язаних компонентів:

- (i) послідовність нуклеїнової кислоти, здатну контролювати експресію в клітинах насіння рослин; і
- (ii) послідовність нуклеїнової кислоти, яка кодує поліпептид інсуліну;
- (b) введення конструкції химерної нуклеїнової кислоти в клітину рослини;
- (c) вирощування клітини рослин в зрілій рослині, здатній давати насіння, де насіння експресує інсулін;
- (d) одержання насіння, яке експресує інсулін; і
- (e) виділення інсуліну з насіння.

Насіння рослин може бути подрібнене із застосуванням будь-якого загальноприйнятого способу, що дає значне руйнування мембрани клітини і стінок клітини. Можуть застосовуватися умови сухого і вологого подрібнення (патент США 3971856; Lawhon et al., 1977, J. Am. Oil Chem. Soc., 63: 533-534). Придатне обладнання для подрібнення включає колоїдні млини, дискові млини, ІКА млини, гомогенізатори промислового масштабу і подібні. Вибір обладнання для подрібнення залежить від типу насіння і загальних вимог. Тверді домішки, такі як оболонка насіння, волокнистий матеріал, нерозчинні вуглеводи, білки і інші нерозчинні у воді домішки, можуть бути видалені з фракції насіння за допомогою, наприклад, методик на основі відбору за розміром, таких як процеси на основі фільтрації або гравітації, такі як центрифугування. У переважних варіантах унікають застосування органічних розчинників, які звичайно застосовуються при витяганні масла, таких як гексан, оскільки такі розчинники можуть пошкодити поліпептид інсуліну. Практично чистий інсулін може бути виділений з насіння за допомогою множини додаткових методик очищення, таких як методики на основі центрифугування; методики на основі відбору за розміром, включаючи, наприклад, ультрафільтрацію через мембрану і ультрафільтрацію в поперечному потоку; і методики хроматографії, включаючи, наприклад, іонообмінну хроматографію, хроматографію з відбором за розміром, афінну хроматографію, високоефективну рідинну хроматографію (ВЕРХ), рідинну експрес-хроматографію білком, хроматографію з гідрофобною взаємодією, і подібні. Звичайно для одержання практично чистого інсуліну застосовують сукупність таких методик.

У особливо переважному варіанті даного винаходу поліпептид інсуліну виділяють з домішок насіння взаємодією поліпептиду інсуліну з масляними тільцями. Цей метод вважається особливо переважним, оскільки він дозволяє видаляти домішки насіння, включаючи білок насіння, особливо ефективним і недорогим способом. Як зазначено вище, взаємодія поліпептиду інсуліну з масляними тільцями може бути досягнута зв'язуванням поліпептиду інсуліну з білком масляного тільця, або зв'язуванням поліпептиду інсуліну з поліпептидом, що володіє спорідненістю з масляним тільцем, таким як антитіло з одним ланцюгом, яке володіє спорідненістю до масляного тільця. У першому варіанті поліпептид інсуліну буде ізолюваний в клітинах масляних тілець і, отже, буде очищатися спільно з масляними тільцями. У другому варіанті, будучи експресованим в мембрані, яка оточує внутрішньоклітинний компартмент, такий як ER, поліпептид інсуліну буде зв'язуватися з масляним тільцем при руйнуванні клітин насіння під час процесу подрібнення. Спосіб виділення масляних тілець описаний в патенті США 5650554.

Фармацевтичні композиції інсуліну можуть бути одержані з очищеного інсуліну, і такі композиції можуть застосовуватися для лікування діабету. Звичайно очищений інсулін змішують з фармацевтично прийнятним носієм або розріджувачем в кількостях, достатніх для надання терапевтично корисної дії при відсутності небажаних побічних ефектів на пацієнта. Для одержання композиції інсуліну масову частину інсуліну розчиняють, суспендують, диспергують або змішують іншим способом у вибраному носії або розріджувачі до ефективної концентрації, що дозволяє поліпшити стан, що лікується. Фармацевтичні композиції інсуліну переважно мають вигляд одиничних дозованих форм. Терапевтично ефективні дози для парентерального введення людського інсуліну відомі в даній галузі техніки. При застосуванні аналогів інсуліну або інших способів введення, терапевтично ефективні дози, що застосовуються, можуть бути емпірично визначені фахівцем в даній галузі техніки із застосуванням відомих протоколів тестування, або екстраполяцією даних тестування *in vivo* або *in vitro*. Однак зрозуміло, що концентрації і дози можуть варіюватися відповідно до важкості стану, який полегшують. Також зрозуміло, що для будь-якого конкретного пацієнта можуть застосовуватися певні режими дозування у часі, які визначаються лікуючим або спостерігаючим лікування фахівцем.

Фармацевтичні розчини або суспензії можуть включати, наприклад, стерильні розріджувачі, такі як, наприклад, вода, лактоза, сахароза, дикальційфосфат або карбоксиметилцелюлоза. Застосовувані носії включають воду, фізіологічний розчин, водну декстрозу, гліцерин, гліколі, етанол і подібні, які можуть утворювати розчин або суспензію. При бажанні, фармацевтичні композиції також можуть містити нетоксичні допоміжні речовини, такі як змочувальні агенти; емульгуючі агенти; солюбілізатори; протимікробні агенти, такі як бензиловий спирт і метилпарабени; антиокиснювачі, такі як аскорбінова кислота і бісульфіт натрію; хелатуючі агенти, такі як етилендіамінтетраоцтова кислота (EDTA); рН буферні агенти, такі як ацетатні, цитратні або фосфатні буфери; і їх поєднання.

Кінцева композиція препаративної форми інсуліну в основному залежить від способу введення інсуліну. Інсулін, одержаний відповідно до даного винаходу, може вводитися будь-яким бажаним способом; однак парентеральний, пероральний, легеневий, букальний і назальний способи введення вважаються найчастіше застосовуваними способами введення. Парентеральні препаративні форми можуть бути в ампулах, одноразових шприцах або флаконах, які містять одну або декілька доз, зроблених з скла, пластику або інших придатних матеріалів.

Приклади

Представлені нижче приклади наведені тільки для ілюстрації і не обмежують даний винахід.

Приклад 1 Одержання білка інсуліну, експресованого у вигляді злитого білка міні-інсуліну (MI) з розщеплюваним трипсином про-пептидом

Одержання pSBS4404: PRS-D9scFv-Klip27-MI-KDEL злитого білка

Один з досліджуваних злитих білків починається з послідовності, зв'язаної з патогеном тютюну (PRS) (Sijmons et al., 1990, Bio/technology, 8: 217-221), яка служить як сигнальний пептид для цільової експресії в ER співтрансляційним методом. Відразу ж після неї йде послідовність, що кодує Fv антитіло з одним ланцюгом (scFv) зі специфічною до видів спорідненістю убівітину з *Arabidopsis thaliana*, позначена D9scFV, далі йде розщеплюваний трипсином пропептид (KLIP27), одержаний з TA57 пропептиду дріжджів (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18: 89-121). За нею йде міні-інсулін (MI), описаний у Kjeldsen et al. (2001), з додаванням сигналу ER-утримання KDEL на C-кінці поліпептиду.

Основний ланцюг цієї плазмиди, pSBS4055, оснований на бінарному векторі рослини, pPZP200, описаному у Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25: 989-994). Замість описаного сайту множинного клонування, *pat* ген, що додає рослині-хазяїну стійкості до фосфінотрицину (Wohleben et al., 1988, Gene 70: 25-37), керований промотором/термінатором убівітину з *Petroselinum crispum* (Kawalleck et al., 1993, Plant. Mol. Bio., 21: 673-684), був введений між лівою і правою пограничними послідовностями. У доповнення до цієї касети, промотор/термінатор β -фазеоліну з *Phaseolus vulgaris* (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80: 1897-1901), керуючий PRS, був субклонований. Стандартну PCR (Horton et al., 1989, Gene 77: 61-68) застосовували для злиття синтетичної PRS-кодуючої послідовності з приєднаними сайтами рестрикції SphI/HindIII ендонуклеази з 3'-кінця промотору фазеоліну з одержанням pSBS4011. SphI-D9scFv-XhoI, SwaI, HindIII введених послідовностей генерували PCR ампліфікацію D9scFv кДНК клону (Sean Hemmingsen lab, неопубліковано) з праймерами 1325 (GCATGCTGACATTG TGATGACACAGTC) - SEQ ID NO:175 і 1326 (AAGCTTGCATTTAAATACTCGAGACTGTGAGAGTGGTGCCTTG) - SEQ ID NO:176. Подальше лігування цього фрагмента в сайтах SphI/HindIII pSBS4011 дає плазмиду pSBS4055.

Послідовність Klip27-MI синтезують з чотирьох олігонуклеотидів, що частково перекриваються, в які введений кодон *Arabidopsis thaliana*, що застосовується для підвищення ефективності трансляції в системах експресії, основаних на рослинах. Олігонуклеотиди 1324 (GAAGAAGGAGAGCCTAAGTTTGTTAATCAACATCTTTGTGGATCTCATCTTGTGAGGCTCT STACCTTG) - SEQ ID NO:177 і 1323 (CCTTAGGAGTGTAGAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) SEQ ID NO:178 гібридизують в їх компліментарному накладенні 20 нуклеотидів, і подовжують з одержанням 5'-кінця Klip27-MI злиття, те ж саме проводять з олігонуклеотидами 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID NO:179 і 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGC AACATTGTTC AACAATTCCCTTAGCAGCCTT) - SEQ ID NO:180 з одержанням 3'-кінця. Дві половини лігують з подальшим рестрикційним розщепленням з Bsu361 з одержанням повної Klip27-MI кодуючої послідовності. PCR такого злиття гена із застосуванням праймерів 1364 (CTCGAGTCAACCAATTGATGACACTGAATC) -SEQ ID NO:181 і 1334 (AAGCTTCAAAGTTTCATCCTTGTGCAATAGTTCTCCAATTG) - SEQ ID NO:182 приєднують 5' сайт розщеплення XhoI рестрикційної ендонуклеази і 3' KDEL ДНК послідовність плюс HindIII сайт розщеплення для подальшого лігування в XhoI/HindIII-розрив pSBS4055. Результатом є плазміда pSBS4404: послідовність ДНК, що кодує злитий білок PRS-D9scFv-Klip27-MI-KDEL, поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Промотор фазеоліну контролює специфічно-тимчасову і специфічну до тканин експресію трансгену під час розвитку насіння. Повна послідовність нуклеїнової кислоти (SEQ ID NO:1) і послідовність амінокислоти (SEQ ID NO:2) злитого білка інсуліну 4404 (PRS-D9scFv-Klip27-MI-KDEL) показана на Фіг.1.

Одержання pSBS44CL5. OLEO-Klip8-KHp27-MI злитого білка

Другий досліджуваний злитий білок починається з 18кДа олеозину з *Arabidopsis thaliana*, далі продовжується розщеплюваним хімозином пропептидом (KLip8) - SEQ ID NO:175. Відразу ж за ним йде послідовність, яка кодує розщеплюваний трипсином пропептид (Klip27), одержана з TA57 пропептиду дріжджів, як описано вище (за Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18: 89-121). Вона злита з міні-інсуліном (MI), описаним вище (Kjeldsen et al. 2001). Експресія цього злитого білка націлена на нові масляні тілця, що утворюються в процесі розвитку ембріона.

Основний ланцюг цієї плазмиди, pSBS4055, оснований на бінарному векторі рослини, pPZP200, описаному у Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25: 989-994). Замість описаного сайту множинного клонування, *pat* ген, що додає рослині-хазяїну стійкості до фосфінотрицину (Wohleben et al., 1988, Gene 70: 25-37), керований промотором/термінатором убівітину з *Petroselinum crispum* (Kawalleck et al., 1993, Plant. Mol. Bio., 21: 673-684), був введений між лівою і правою пограничними послідовностями. У доповнення до цієї касети, промотор/термінатор β -фазеоліну з *Phaseolus vulgaris* (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80: 1897-1901), який керує злиттям геномної послідовності-Klip8 *Arabidopsis* 18кДа олеозину, був субклонований. Стандартну PCR (Horton et al., 1989, Gene 77: 61-68) застосовували для злиття послідовності генів олеозину-Klip8 з сайтами XhoI/HindIII рестрикційної ендонуклеази з 3'-кінця промотору фазеоліну з одержанням pSBS4010.

Послідовність Klip27-MI синтезують з чотирьох олігонуклеотидів, що частково перекриваються, в які введений кодон *Arabidopsis thaliana*, що застосовується для підвищення ефективності трансляції в системах експресії, основаних на рослинах. Олігонуклеотиди 1324 (GAAGAAGGAGAGCCTAAGTTTGTTAATCAACATCTTTGTGGATCTCATCTTGTGAGGCTCT STACCTTG) - SEQ ID NO:177 і 1323 (CCTTAGGAGTGTAGAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) SEQ ID NO:178 гібридизують в їх компліментарному накладенні 20 нуклеотидів, і подовжують з одержанням 5'-кінця Klip27-MI злиття, те ж саме проводять з олігонуклеотидами 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID NO:179 і 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGC AACATTGTTC AACAATTCCCTTAGCAGCCTT) - SEQ ID NO:180 з одержанням 3'-кінця. Дві половини лігують з подальшим рестрикційним розщепленням з Bsu361 з одержанням повної Klip27-MI кодуючої послідовності. PCR такого злиття гена із застосуванням праймерів 1364 (CTCGAGTCAACCAATTGATGACACTGAATC) -SEQ ID NO:181 і 1329 (AAGCTTCAAGTTGCAATAGTTT) - SEQ ID NO:183 приєднують 5' сайт розщеплення XhoI рестрикційної ендонуклеази і 3' HindIII сайт розщеплення, відповідно, для подальшого лігування в XhoI/HindIII-розрив pSBS4010. Результатом є плазміда pSBS4405: послідовність ДНК, що кодує злитий білок OLEO-Klip8-Klip27-

MI, поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Промотор фазеоліну контролює специфічно-тимчасову і специфічну до тканин експресію трансгену під час розвитку насіння. Повна послідовність нуклеїнової кислоти (SEQ ID NO:3) і послідовність амінокислоти (SEQ ID NO:4) злитого білка інсуліну 4405 (OLEO-Klip8-Klip27-MI) показана на Фіг.2.

Одержання рSBS4414: PRS-MI-тетраосновний лінкерp-D9Scfv-KDEL злитого білка

Інший з досліджуваних злитих білків починається з послідовності, зв'язаної з патогеном тютюну (PRS) (Sijmons et al., 1990, Bio/technology, 8: 217-221), яка служить як сигнальний пептид для цільової експресії в ER співтрансляційним методом. Відразу ж після неї йде послідовність, що кодує міні-інсулін (MI) описаний у Kjeldsen et al. (2001) за винятком того, що область міні-С пептиду (AAK-SEQ ID NO:146) замінюють проміжним В₃₀ треонінтетраосновним сайтом (В₃₀-T-RRKR) (SEQ ID NO:149), послідовністю між В₍₁₋₂₉₎- і А₍₁₋₂₁₎ - ланцюгами людського інсуліну. Потім відразу ж йде послідовність, що кодує другий тетраосновний лінкер, а потім Fv антитіло з одним ланцюгом (scFv) з видоспецифічною спорідненістю проти 18кДа олеозину з *Arabidopsis thaliana*, позначена D9scFV. На С-кінці поліпептиду доданий сигнал ER-утримування KDEL.

Основний ланцюг цієї плазмиди, рSBS4055, оснований на бінарному векторі рослини, рPZP200, описаному у Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25: 989-994). Замість описаного сайту множинного клонування, pat ген, що додає рослині-хазяїну стійкості до фосфінотрицину (Wohlleben et al., 1988, Gene 70: 25-37), керований промотором/термінатором убіквітину з *Petroselinum crispum* (Kawalleck et al., 1993, Plant. Mol. Bio., 21: 673-684), був введений між лівою і правою пограничними послідовностями. У доповнення до цієї касети, промотор/термінатор β-фазеоліну з *Phaseolus vulgaris* (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80: 1897-1901), керуючий PRS, був субклонований. Стандартну PCR (Horton et al., 1989, Gene 77: 61-68) застосовували для злиття синтетичної PRS-кодуєчої послідовності з приєднаними сайтами SphI/HindIII рестрикційної ендонуклеази з 3'-кінця промотору фазеоліну з одержанням PSBS4011.

Послідовність Klip27-MI синтезують з чотирьох олігонуклеотидів, що частково перекриваються, в які введений кодон *Arabidopsis thaliana*, що застосовується для підвищення ефективності трансляції в системах експресії, основаних на рослинах. Олігонуклеотиди 1324 (GAAGAAGGAGAGCCTAAGTTTGTAAATCAACATCTTTGTGGATCTCATCTTGTGAGGCTCT CTACCTTG) - SEQ ID NO:177 і 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) SEQ ID NO:178 гібридизують в їх компліментарному накладенні 20 нуклеотидів, і подовжують з одержанням 5'-кінця Klip27-MI злиття, те ж саме проводять з олігонуклеотидами 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID NO:179 і 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACATTGTTC AACAATTCCCTTAGCAGCCTT) - SEQ ID NO:180 з одержанням 3'-кінця. Дві половини лігують з подальшим рестрикційним розщепленням з Bsu36I з одержанням повної Klip27-MI кодуєчої послідовності. PCR такого злиття гена із застосуванням праймерів 1363 (GCATGCCCAACCAATTGATGACACTG) - SEQ ID NO:184 і 1334 (AAGCTTCAAAGTTTCATCCTTGTGCAATAGTTCTCCAATTG) - SEQ ID NO:182 приєднують 5' сайт розщеплення SphI рестрикційної ендонуклеази і 3' KDEL ДНК послідовність плюс HindIII сайт розщеплення для подальшого лігування в SphI/HindIII-розрив рSBS4011. Результатом є плазміда рSBS4402: послідовність ДНК, що кодує злитий білок PRS-Klip27-MI-KDEL, поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Промотор фазеоліну контролює специфічно-тимчасову і специфічну до тканин експресію трансгену під час розвитку насіння. Вектор експресії рослини, рSBS4402, служить як зразок введення сайту між В і А ланцюгами інсуліну і між MI і D9Scfv.

Проміжний тетраосновний сайт (В₃₀-T-RRKR) поміщують між В₍₁₋₂₉₎- і А₍₁₋₂₁₎-ланцюгами людського інсуліну за допомогою PCR із застосуванням праймерів 1515 (GCATGCATGCCTTTGTAAATCAACATCTTTGTGG) SEQ ID NO:185 і 1518 (ACATTGTTCAACAATTCCTCTCTTTCTTCTAGTCTTAGGAGTGTAGAAAAATCC) SEQ ID NO:186 із застосуванням рSBS4402 як зразка. Одержаний 124п.н. фрагмент застосовують в поєднанні з праймером 1517 (GCATAAGCTTCAAAGCTCATCCTTTGAGC) SEQ ID NO:187 із застосуванням рSBS3400 як зразка. Необхідно зазначити, що рSBS3400 являє собою плазмиду, яка містить D9scFv-KDEL фрагмент з сайтом рестрикції HindIII. Реакція PCR дає 955п.н. продукт, який вводить тетраосновний (RRKR)-D9Scfv-KDEL-HindIII в 124п.н. SphI-MI фрагмент. Фрагмент 955п.н. потім лігують і субклонують в рGEM-T (Promega) з одержанням рSBS3403. Весь SphI-MI (з В₃₀-T-RRKR модифікованим С пропептидом)-RRKR-D9Scfv-KDEL-HindIII фрагмент вводять в заздалегідь розрізаний (SphI/HindIII) рSBS4402 з одержанням рSBS4414. Повна послідовність нуклеїнової кислоти (SEQ ID NO:5) і послідовність амінокислоти (SEQ ID NO:6) злитого білка інсуліну 4414 (PRS-MI-тетраосновний лінкер-D9Scfv-KDEL) показана на Фіг.3.

Трансформація і вирощування рекомбінантної *E. coli* і *Agrobacterium* з рSBS4404, рSBS4405 або рSBS4414

Після підтвердження цілісності кДНК кодування для злитого білка аналізом послідовності, плазмиди рSBS4404, рSBS4405 і рSBS4414 трансформують в штам *E. coli* DH5α для одержання високого рівня експресії. Виділену плазмиду ДНК (100нг) змішують на льоду зі 100мкл компетентних клітин DH5α протягом 20хв. Потім клітини піддають тепловому удару при температурі 42°C протягом 45 секунд, і повертають на лід на 2хв. Потім додають 1мл середовища SOC, і клітини інкубують при температурі 37°C на енвіро-шейкері при 225об./хв., протягом 1 години перед поміщенням трансформованих клітин на планшети з LB-спектиноміцином (10г/л триптон, 5г/л екстракту дріжджів, 5г/л NaCl, 15г/л агару) і інкубуванням протягом ночі при температурі 37°C. Одну колонію використовують для інокулювання 5мл LB-спектиноміцинового бульйону. Ці культури вирощують протягом ночі при температурі 37°C. Рекомбінантну плазмиду виділяють з 1мл середовища, вибудовуваного протягом ночі, із застосуванням набору QIAprep®Spin Miniprep (Qiagen). Виділену плазмиду потім застосовують для трансформації компетентного штаму *Agrobacterium* EH101 (Hood et al., 1986; J. Bacteriol. 144: 732-743) електропорацією (25мкф, 2,5кВ, 200Ом). Рекомбінантні *Agrobacterium* помішують на планшети з АВ-спектиноміцином/канаміцином (20хАВ солі, 2М глюкози, 0,25мг/мл FeSO₄·7H₂O, 1М MgSO₄, 1М CaCl₂) і одну колонію застосовують для інокулювання 5мл АВ-спектиноміцинового/канаміцинового бульйону. Ці культури вирощують протягом ночі при температурі 28°C. Рекомбінантні *Agrobacterium* потім застосовують для трансформації рослин *Arabidopsis thaliana* методом занурення квіток (Clough et al., 1998, Plant J., 16: 735-

743). Вигляд *Arabidopsis thaliana* (C24) застосовують для всіх експериментів. Насіння висівають на поверхні суміші ґрунтів (дві третини землі Redi і одна третина перліту з pH=6,7) або суміші ґрунтів для *Arabidopsis* від Lehle Seeds (перліт, вермикуліт, торф, терра-грін, з pH=5,5) в 4-х дюймові горщики. Саджанці вирощують до стадії розеток з 6-8 листів діаметром приблизно 2,5см. Горщики поміщують всередину купола при температурі 4°C на чотири дні для холодної обробки, і потім переносять в кімнату для вирощування з температурою 24°C з постійним світлом близько 150мкЕ і відносною вологістю 60-70%. Рослини зрошують з інтервалом 2-3 дні і щотижня підгодовують добривом 1% Peters 20-19-18. Кожний горщик містить від п'яти до шести рослин. Коли рослини досягають близько 2см у висоту, перші стрілки зрізують для стимулювання росту других і третіх стрілок. Через 4-5 днів після зрізу первинних стрілок рослини готові до зараження *Agrobacterium*. Горщики з рослинами *Arabidopsis* перекидають для зараження рослин *Arabidopsis* в 500мл ре-суспезії культури *Agrobacterium*, вирощеної протягом ночі, яка містить цільовий вектор трансформації рослини, протягом 20 секунд. Важливо, щоб культура *Agrobacterium* містила 5% сахарози і 0,05% поверхнево-активної речовини Silwet L-77 (Lehle Seeds). Далі горщики накривають прозорими пластиковими кришками на 24 години для підтримки більш високої вологості. Рослини вирощують до дорослого стану і збирають суміш насіння, трансформованого і нетрансформованого. Для вибору трансгенних ліній передбачуване трансформоване насіння стерилізують швидким промиванням 70% етанолом, потім 20% комерційним відбілювачем протягом 15хв., і потім обполіскують принаймні чотири рази ddH₂O. Близько 1000 стерилізованих насінин змішують з 0,6% розплавленим агаром і рівномірно розподіляють на MS планшеті з половинною силою (Murashige and Skoog, 1962, *Physiologia Plantarum* 15: 473-497), що містить 0,3% сахарози і 80мкМ гербіциду фосфінотрицину (PPT) DL. Потім планшети поміщують в кімнату для вирощування з режимом освітлення 8 годин темряви і 16 годин світла при температурі 24°C. Через 7-10 днів передбачувані трансгенні саджанці залишаються зеленими і ростуть, в той час як нетрансформовані саджанці бліднуть. Після появи коріння передбачувані трансгенні саджанці індивідуально переносять в горщики (окремі рослини зрошують з інтервалом 3 дні і підгодовують добривами 1% Peters 20-19-18 з інтервалом 7 днів) і вирощують до дорослого стану. Горщики накривають прозорими пластиковими кришками на три дні для захисту чутливих саджанців. Через 7 днів саджанці накривають системою збору насіння від Lehle Seeds для запобігання втраті насіння через розсіювання. Насіння трансгенних рослин збирають окремо і підготовлюють для аналізу.

Приклад 2

Рівні експресії інсуліну у *Arabidopsis thaliana*

У другому прикладі визначають рівні експресії злитого білка D9scfv-KLIP27-MI KDEL (4404), OLEO-KLIP8-KLIP27-MI (4405) і PRS-MI-RRKR-D9Scfv-KDEL (4414) в трансгенному зрілому насінні *Arabidopsis thaliana*. Показано, що трансгенний продукт присутній в клітинних екстрактах зрілого насіння. Приблизно 40 трансгенних насінин *Arabidopsis thaliana* подрібнюють в ступі товчачиком в 50мкл 50мМ Тріс-НСІ, pH8,0. Потім до суспензії додають відновлювальний SDS-PAGE зразковий буфер (6х SDS зразковий буфер, 0,35М Тріс-НСІ, pH6,8, 30% гліцерин, 10% SDS, 0,012% блакитного бромфенолу, 5% β-меркаптоетанолу), і перемішують короткими завихреннями. Потім зразок швидко центрифугують і витримують при температурі 99°C протягом 10 хвилин. Після охолодження на льоду протягом 2 хвилин зразок швидко центрифугують. Зразки поміщують (10мкл - еквівалент приблизно 7 насінинам) в умови відновлення.

Для підготовлених зразків масляних тілець трансгенне і дике насіння (20мг) подрібнюють в 250мкл буфера екстрагування масляних тілець (0,4М сахарози, 0,5М NaCl, 50мМ Тріс-НСІ, pH8,0). Зразки мікроцентрифугують при 10000g протягом 10хв. Розчинні водні фракції видаляють за допомогою 26 G 5/8 1мл шприца, і жирний шар повторно суспендують в 100мкл фосфатного буфера з додаванням солі (20мМ Na₂HPO₄, pH8,0, 0,5М NaCl). Повторно суспендований жирний шар переносять в чисту пробірку мікроцентрифуги і знов центрифугують при 10000g протягом 10хв. Процедура повторюють ще 3 рази, з останнім повторним суспендуванням жирного шару в 100мкл фосфатному буфері без солі (20мМ Na₂HPO₄, pH8,0). Два додаткових промивання в фосфатному буфері без солі проводять по чергово з центрифугуванням, як описано вище. Кінцевий жирний коржик повторно суспендують в 10мкл фосфатного буфера (20мМ Na₂HPO₄, pH8,0). Беруть 5мкл аліквоту, і білок масляного тільца солюбілізують кип'ятінням в 1/10 (об./об.) 50мМ Тріс-НСІ, pH8,0 з 2% SDS. Зразок охолоджують на льоду протягом 2хв. і центрифугують при 10000g протягом 5хв. Вміст білка в підосадовій рідині визначають ІСФ аналізом білка (Pierce, Rockford, IL). Для аналізу забарвлених кумасі гелів і вестерн-блотингу 20мкг загального білка відділяють на 15% SDS-PAGE гелях в умовах відновлення із застосуванням SDS-PAGE зразкового буфера.

Потім зразки завантажують в дискретні 15% SDS-PAGE гелі і розділяють при 150 вольт протягом приблизно 1,5 години. Потім гелі або забарвлюють кумасі, або наносять на PVDF мембрану (Immobilon-P, Millipore Corporation, Bedford, MA) для вестерн-блотингу. Нанесені зразки досліджують за допомогою моноклональних антитіл, направлених проти інсуліну (Clone E2-E3; Roth et al., 1992) від Abeam (Cambridge, UK). Смуги інсуліну визначають із застосуванням повторного Sheep X IgG миші F(ab')₂ AP-кон'югата (Chemicon International, Temecula, CA) і виявляють за допомогою NBT-BCIP в GARAP буфері (Тріс-НСІ, pH9,5, 100мМ NaCl, 5мМ MgCl₂). Імунореакційноздатна смуга відповідає поліпептидній смузі, мігруючи при розрахунковій молекулярній вазі злитого білка, як показано на Фіг.4А-4F. На Фіг.4 (А-F) показана рекомбінантна експресія злитих білків інсуліну в трансформованих лініях *Arabidopsis thaliana* (4404-2, -17, -20, 4405-4, 4414-19 і 4414-20) на основі забарвленого кумасі SDS-PAGE і вестерн-блотингу. Стрілки вказують на положення мігруючих 38,5кДа, 34,2кДа і 34,2кДа злитих білків, PRS-D9(scfv)-KLIP27-MIw/KDEL (4404), OLEO-KLIP8-KLIP27-MI (4405) і PRS-MI-RRKR-D9Scfv-KDEL (4414), відповідно, в умовах відновлення. Необхідно зазначити, що злитий білок 4414 має очікувану молекулярну вагу 34,2кДа, але має більш високу реальну молекулярну вагу на SDS-PAGE гелі. На Фіг.4А (забарвлений кумасі гель) і 4В (відповідний вестерн-блотинг, досліджений з антиінсуліном E2E3) показаний загальний білок насіння з диких (wt) і трансгенних ліній насіння, яке експресує 4404 і 4405 конструкції. На Фіг.4С (забарвлений кумасі гель) і 4D (відповідний вестерн-блотинг, досліджений з антиінсуліном E2E3) показаний білок масляних тілець, одержаний з дикого і трансгенного насіння, яке експресує ті ж самі 4404 і 4405 конструкції. На Фіг.4D (забарвлений кумасі гель) і 4Е (відповідний вестерн-

блотинг, досліджений з антиінсуліном E2E3) показаний білок масляних тілець, одержаний з дикого і трансгенного насіння, яке експресує ту ж саму 4414 конструкцію. Маркери молекулярної ваги (М) дорівнюють 10, 15, 20, 25, 37, 50, 75, 100, 150, 250кДа. Контрольні групи включають hIN (стандарт рекомбінантного людського інсуліну) і hProIN (стандарт рекомбінантного людського проінсуліну), виділені в невідновних умовах. Відмінності в рівнях експресії є результатом клональних варіацій між трансформантами. Приблизні рівні білка трансгенів в експресії MI показані на Фіг.5. Рівні експресії визначають із застосуванням стрічки 18кДа олеозину як внутрішнього стандарту (еквівалентно 1,5% загальних білка насіння) денситометрією стрічки трансгену. Середній рівень експресії для PRS-D9(scfv)-KLIP27-MIw/KDEL (4404), OLEO-KLIP8-KLIP27-MI (4405) і PRS-MI-RRKR-D9Scfv-KDEL (4414) конструкцій становить 0,21% від загального білка насіння, 0,12% від загального білка насіння і 0,79% від загального білка насіння, відповідно.

Приклад 3

Розщеплення pSBS4404 і очищення BEPX

Елюювання з масляного тільця

У третьому прикладі 1г трансгенного насіння гомогенізують в 12мл буфера екстрагування (0,4М сахарози, 0,5М NaCl, 50мМ Tris-HCl, pH8,0) і центрифугують при 10000g протягом 10хв., жирні шари видаляють і вміщують в 1мл 20мМ Na₂HPO₄, 0,5М NaCl і повторно центрифугують, як описано вище. Процедuru повторюють двічі, перед промиванням і центрифугуванням жирного шару двічі в 750мкл 20мМ Na₂HPO₄. Злитий білок 4404 елюють з масляного тільця в підсадову рідину промиванням кінцевого жирного шару 5 разів в 750мкл 20мМ мурашиної кислоти, pH4,1, при центрифугуванні при 10000g між кожною промивкою. Зібрані елюйовані фракції (підсадові рідини) об'єднують і нейтралізували 2н NaOH до pH8,0. Весь розчин охолоджують до температури -80°C до замерзання і ліофілізують протягом ночі для концентрації злитого білка. Ліофілізований зразок повторно суспендують в 500мкл 50мМ Tris-HCl, pH8,0. Повторно суспендований злитий білок 4404 потім опрісняють на колонці NAP-5 (Amersham Pharmacia Biotech Ab, Uppsala, Sweden) і повторно замінують буфером (50мМ Tris-HCl, pH8,0). Опріснену фракцію потім знов заморожують і ліофілізують протягом ночі для концентрації. Кінцевий концентрований зразок повторно суспендують в кінцевому об'ємі 105мкл двічі дистильованою H₂O. Результати елюювання подані на Фіг.6. На Фіг.6 зображений аналіз забарвленого кумасі SDS-PAGE (15%) препаратів масляних тілець до елюювання (-OB), препаратів OB після елюювання мурашиною кислотою (-OB¹) і концентрованого елюйованого матеріалу (-E). Стрілки вказують на положення міграції злитого поліпептиду. Контрольна група дикого типу практично не містить які-небудь основні білки після елюювання, в той час як концентрований матеріал 4404 містить злитий білок, деяку кількість укорочених продуктів (можливо, гідролізований злитий білок) і, можливо, деякі співелюйовані альбуміни.

Розщеплення і BEPX аналіз 4404, експресованого насінням Arabidopsis

Концентрований зразок повторно суспендують в 105мкл двічі дистильованої води, і вміст білка оцінюють за допомогою тесту для білка BCA згідно з інструкцією виробника (Pierce, Rockford, IL, USA). Потім зразки розщеплюють трипсином (1:300 трипсин загальний білок, в 50мМ Tris-HCl, pH8,0, на льоду протягом 90хв.). Реакції зупиняють додаванням 10-кратного молярного надлишку TLCK (N-п-тозил-L-лізин хлорметилкетону). Потім всі реакційні суміші фільтрують через 0,2мкм фільтри (Aerodisc®, 13мМ фільтр для шприца з 0,2мкм мембраною Supof®, Pall Corporation, Ann Arbor, MI, USA) і аналізують BEPX з оборотною фазою (OF) із застосуванням C18 колонки (Zorbax 300SB-C18, Agilent Technologies, Waldbronn, Germany). Зразки завантажують в колонку і елюють зі швидкістю 1,0мл/хв. із застосуванням 19-хв. лінійного градієнта 5-50% (об./об.) ацетонітрилу в 0,1% (об./об.) TFA. Хроматограма цього аналізу подана на Фіг.7. Сліди показують розщеплений трипсином продукт із злитого білка 4404, з властивостями на колонці, практично ідентичними стандарту людського інсуліну (час утримування 17,011хв. і 17,179хв., відповідно). BEPX фракцію збирають на 17,0-17,5хв. і аналізують мас-спектрометрією PSD MALDI/TOF із застосуванням мас-спектрометра Voyager-DE STR (Applied Biosystems). MS аналіз, проведений за допомогою послуги БіоАналітичної Спектроскопії, що пропонується NRC-Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada. Повторне розчинення розщепленого продукту 4404, очищеного BEPX, як описано вище, показане на Фіг.8В в порівнянні зі стандартом людського інсуліну, показаного на Фіг.8А. Одержана маса розщепленого злитого білка 4404 з трипсином становить 6191,51Да. Розбіжність між стандартом людського інсуліну (Фіг.8А) і розщепленим продуктом 4404 (Фіг.8В) відповідає Des-B₃₀ інсуліну з KDEL сигналом, який утримується на ланцюзі А розщепленого продукту (des-B₃₀інсулін-KDEL).

Приклад 4

Розщеплення pSBS4405 і очищення BEPX

Елюювання з масляного тільця

Злитий білок (OLEO-KLIP8-KLIP27-MI) може бути частково очищений одержанням препаратів масляних тілець, як описано вище. Приблизно 1г трансгенного насіння гомогенізують в 12мл буфера екстрагування (0,4М сахарози, 0,5М NaCl, 50мМ Tris-HCl, pH8,0) і центрифугують при 10000g протягом 10хв., жирні шари видаляють і поміщують в 1мл 50мМ Tris-HCl, pH8,0, 0,5М NaCl і повторно центрифугують, як описано вище. Процедuru повторюють двічі, перед промиванням і центрифугуванням жирного шару двічі в 750мкл 50мМ Tris-HCl, pH8,0. Препарування масляних тілець в результаті дає видалення більшості фонових білків. Типовий профіль білків препарату масляних тілець з трансгенного насіння Arabidopsis, яке експресує 4405 конструкцію, показаний на Фіг.9.

Розщеплення і BEPX аналіз 4405, експресованого насінням Arabidopsis

Загальний вміст білка з повторно суспендованих масляних тілець оцінюють шляхом солюбілізації фракції препаратів (5мкл), розбавленої в 10 раз в 2% SDS, 50мМ Tris-HCl, pH8,0, кип'яченої протягом 5 хвилин і центрифугованої протягом 3хв. при 10000g. Потім вміст білка оцінюють за допомогою тесту для білка BCA згідно з інструкцією виробника (Pierce, Rockford, IL, USA). Потім зразки розщеплюють трипсином (1:300 трипсин: загальний білок, в 50мМ Tris-HCl, pH8,0, на льоду протягом 90хв.) для виділення Klip27-MI фрагмента із злитого білка. Реакції зупиняють додаванням 10-кратного молярного надлишку TLCK (N-п-тозил-

L-лізін хлорметилкетону). Зразки центрифугують при 10000g протягом 10хв., і підосадові рідини всіх реакційних сумішей фільтрують через 0,2мкм фільтри (Aerodisc®, 13мм фільтр для шприца з 0,2мкм мембраною Supro®, Pall Corporation, Ann Arbor, MI, USA). На Фіг.9 зображений аналіз забарвленого кумасі SDS-PAGE (15%) загального екстрагованого білка насіння і білка, одержаного з масляних тілець (OB) з колоній клітин, які експресують 4405, у порівнянні з диким типом (нерекомбінантним) насіння. Стрілки вказують на положення міграції злитого поліпептиду. Підосадові рідини потім аналізують ВЕРХ з оборотною фазою (ОФ) із застосуванням CI 8 колонки (Zorbax 300 SB-C18, Agilent Technologies, Waldbronn, Germany). Зразки завантажують в колонку і елюють зі швидкістю 1,0мл/хв. із застосуванням 19-хв. лінійного градієнта 5-50% (об./об.) ацетонітрилу в 0,1% (об./об.) TFA. Хроматограма цього аналізу подана на Фіг.10. Сліди показують розщеплений трипсином продукт із злитого білка 4405, з властивостями на колонці, практично ідентичними стандарту людського інсуліну (час утримування 17,220хв. і 17,179хв., відповідно). ВЕРХ фракцію збирають на 17,0-17,5хв. і аналізують мас-спектрометриєю PSD MALDI/TOF із застосуванням мас-спектрометра Voyager-DE STR (Applied Biosystems). МС-аналіз, проведений за допомогою послуги БіоАналітичної Спектроскопії, що пропонується NRC-Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada. Як показано на Фіг.11, одержана маса розщепленого злитого білка 4405 з трипсином становить 5706,30 Да. Розбіжність між стандартом людського інсуліну (Фіг.8А) і розщепленим продуктом 4405 (Фіг.11) відповідає Des-B₃₀ інсуліну (Des-Взоінсулін). Des-Взоінсулін являє собою продукт, очікуваний при корекції дозрівання трипсину злиття 4405.

Приклад 5

Очищення розщепленого трипсином MI із застосуванням АКТА Explorer (FPLC)

Очищення відщепленого MI від 4405 також забезпечує часткове очищення від реакцій розщеплення більш високого рівня аніонним обміном (Mono Q FF 1мл, Amersham Pharmacia) на АКТА Explorer (Amersham Pharmacia). Реакції розщеплення проводять з масляними тільцями 4405, як описано вище, одержаними з 30г трансгенного насіння. Підосадові рідини з реакції розщеплення або фільтрують через 0,2мкм фільтри, або концентрують ліофілізацією на Savant Speed Vac. Фільтровані реакційні суміші зразків можуть бути нанесені на колонку безпосередньо, але концентровані зразки вимагають видалення солей для ефективного зв'язування колонки. Концентровані зразки можуть бути опріснені пропусканням розщепленого матеріалу через колонку PD-10 (Amersham Pharmacia), діалізом або розбавленням до концентрації солі, еквівалентної або менше ніж 5мС/см. Опріснені зразки врівноважують 20мМ Тріс-НСІ, рН6,5. Зразки відділяють із застосуванням покрокового градієнта з NaCl 0-40% NaCl зі швидкістю потоку 1мл/хв. Визначення проводять при 214нм (визначення при 280 нм відносно слабе через низький вміст ароматичних амінокислот в інсуліні). Розчинник А складається з 20мМ Тріс-НСІ, рН6,5, а розчинник В складається з 20мМ Тріс-НСІ, рН6,5, 1,0М NaCl. Фракції (1мл), елюйовані при тій же провідності, що і стандарт інсуліну Roche, між 7-35мС/см, збирають (див. Фіг.12). На Фіг.12 представлена хроматограма розщеплених трипсином препаратів масляних тілець 4405 (пунктирна лінія) в порівнянні зі стандартом людського інсуліну (суцільна лінія). Присутність інсуліну підтверджується в зібраних фракціях за допомогою ВЕРХ, ELISA або вестерн-блотингу (дані не показані). Зібрані зразки концентрують ліофілізацією і застосовують для біотестування інсуліну, описаного в прикладі 6.

Приклад 6

Тестування переносимості інсуліну: біотестування на самцях мишей C57Bl/6(B6)

Це біотестування проводять для визначення *in vivo* дії рекомбінантного, одержаного в рослинах (DesB₃₀ IN) з розщепленого трипсином 4405, в порівнянні з людським інсуліном. Рівні глюкози в плазмі мишей B6 визначають до і після внутрішньочеревинної ін'єкції стандартів інсуліну, негативного контролю і SBS інсуліну. П'ятнадцять мишей C57Bl/6 (B6) віком приблизно 2 місяці придбавають у Jackson Laboratories (Bar Harbor, ME). Рівні глюкози в плазмі визначають автоматичним глюкометром (OneTouch Ultra, Lifescan, Johnson and Johnson). Позитивний контроль включає HumulinR® (Eli Lilly) і дріжджовий рекомбінантний стандарт людського інсуліну від Roche. Як плацебо застосовують фізіологічний розчин. Негативний контроль включає розщеплені трипсином масляні тільця, очищені з дикого типу (нерекомбінантного) насіння Arabidopsis, яке обробляють ідентично рекомбінантним препаратам розщеплених трипсином масляних тілець 4405.

Мишей B6 поміщують в клітини і годують *ad libitum* при 12-годинному циклі день-ніч. Для проведення тестів на переносимість інсуліну мишам вводять внутрішньочеревинні ін'єкції (ВЧ) інсуліну (10д/кг маси тіла), і рівні глюкози визначають через 0, 15, 30 і 60 хвилин із застосуванням автоматичного глюкометра. Всі тести на переносимість інсуліну проводять з однаковим інтервалом кожний день (9:00). Тести на переносимість інсуліну проводять принаймні протягом 2 днів, з проміжками часу перед введенням чергового тесту. Результати тестів переносимості інсуліну подані на Фіг.13. SBS DesB₃₀інсулін, одержаний з насіння 4405 (зафарбовані ромби), діє практично ідентично (статистично не відрізняється, $p < 0,05$) з HumulinR® (пусті квадрати) і інсуліном від Roche (пусті трикутники) після введення всього курсу дослідження. Всі тестовані інсуліни значно знижують рівні глюкози в плазмі ($p < 0,05$) в порівнянні з плацебо фізіологічним розчином (пусті кола) і розщепленням трипсином масляних тілець з насіння дикого типу Arabidopsis (зафарбовані кола) (негативний контроль).

Приклад 7

Одержання pSBS4401: PRS-Klip27-MI злитого білка

Один з досліджуваних злитих білків починається з послідовності, зв'язаної з патогеном тютюну (PRS) (Sijmons et al., 1990, Bio/technology, 8: 217-221), яка служить як сигнальний пептид для цільової експресії в ER співтрансляційним методом. Відразу ж після неї йде розщеплюваний трипсином пропептид (KLIP27), одержаний з TA57 пропептиду дріжджів (Kjeldsen et al., 2001, Biotechnology and Genetic Engineering Reviews 18: 89-121). Потім йде міні-інсулін (MI), описаний у Kjeldsen et al. (2001).

Основний ланцюг цієї плазмиди, pSBS4055, оснований на бінарному векторі рослини, pPZP200, описаному у Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25: 989-994). Замість описаного сайта множинного клонування, *pat* ген, що додає рослині-хазяїну стійкості до фосфінотрицину (Wohleben et al., 1988, Gene 70: 25-37), керований промотором/термінатором убівітину з *Petroselinum crispum* (Kawalleck et al., 1993, Plant. Mol. Bio., 21: 673-684), був введений між лівою і правою пограничними послідовностями. У доповнення до цієї

касети, промотор/термінатор β -фазеоліну з *Phaseolus vulgaris* (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80: 1897-1901), керуючий PRS, був субклонований. Стандартну PCR (Horton et al, 1989, Gene 77: 61-68) застосовували для злиття синтетичної PRS-кодуючої послідовності з приєднаними сайтами SphI/HindIII рестрикційної ендонуклеази з 3'-кінця промотору фазеоліну з одержанням pSBS4011.

Послідовність Klip27-MI синтезують з чотирьох олігонуклеотидів, що частково перекриваються, в які введений кодон *Arabidopsis thaliana*, що застосовується для підвищення ефективності трансляції в системах експресії, основаних на рослинах. Олігонуклеотиди 1324 (GAAGAAGGAGAGCCTAAGTTTGTAAATCAACATCTTTGTGGATCTCATCTTGTGGAGGCTCT CTACCTTG) - SEQ ID NO:177 і 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) SEQ ID NO:178 гібридизують в їх компліментарному накладенні 20 нуклеотидів, і подовжують з одержанням 5'-кінця Klip27-MI злиття, те ж саме проводять з олігонуклеотидами 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID NO:179 і 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACATTGTTC AACAATTCCCTTAGCAGCCTT) - SEQ ID NO:180 з одержанням 3'-кінця. Дві половини лігують з подальшим рестрикційним розщепленням з Bsu36I з одержанням повної Klip27-MI кодуючої послідовності. PCR такого злиття гена із застосуванням праймерів 1363 (GCATGCCCAACCAATTGATGACACTG) - SEQ ID NO:184 і 1329 (AAGCTTCAGTTGCAATAGTTC) - SEQ ID NO:183 приєднують 5' сайт розщеплення SphI і 3' HindIII рестрикційної ендонуклеази для подальшого лігування в SphI/HindIII-розрив pSBS4011 (як описано вище). Результатом є плазмід pSBS4401: послідовність ДНК (SEQ ID NO:188), що кодує злитий білок PRS-Klip27-MI (SEQ ID NO:189), поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Промотор фазеоліну контролює специфічно-тимчасову і специфічну до тканин експресію трансгену під час розвитку насіння.

Трансформація і вирощування рекомбінантної *E. coli* і *Agrobacterium* з pSBS4401

Після підтвердження цілісності кДНК кодування для злитого білка аналізом послідовності, плазмиду pSBS4401 трансформують в штам *E. coli* DH5 α для одержання високого рівня експресії. Виділену плазмиду ДНК (100нг) змішують на льоду зі 100мкл компетентних клітин DH5 α протягом 20хв. Потім клітини піддають тепловому удару при температурі 42°C протягом 45 секунд, і повертають в лід на 2хв. Потім додають 1мл середовища SOC, і клітини інкубують при температурі 37°C на енвіро-шейкері при 225об./хв. протягом 1 години перед поміщенням трансформованих клітин на планшети з LB-спектиноміцином (10г/л триптон, 5г/л екстракту дріжджів, 5г/л NaCl, 15г/л агару) і інкубуванням протягом ночі при температурі 37°C. Одну колонію використовують для інокулювання 5мл LB-спектиноміцинового бульйону. Ці культури вирощують протягом ночі при температурі 37°C. Рекомбінантну плазмиду виділяють з 1мл середовища, вибудовуваного протягом ночі, згідно з одержанням міні-препаратів Qiagen. Виділену плазмиду потім застосовують для трансформації компетентного штаму *Agrobacterium* EH101 (Hood et al., 1986; J. Bacteriol. 144: 732-743) електропорацією (25мкФ, 2,5кВ, 200Ом). Рекомбінантні *Agrobacterium* поміщують на планшети з АВ-спектиноміцином/канаміцином (20хАВ солі, 2М глюкози, 0,25мг/мл FeSO $_4$ ·7H $_2$ O, 1М MgSO $_4$, 1М CaCl $_2$) і одну колонію застосовують для інокулювання 5мл АВ-спектиноміцинового/канаміцинового бульйону. Ці культури вирощують протягом ночі при температурі 28°C. Рекомбінантні *Agrobacterium* потім застосовують для трансформації рослин *Arabidopsis thaliana* методом занурення квіток (Clough et al., 1998, Plant J., 16: 735-743), як описано в прикладі 1.

Рівні експресії інсуліну в *Arabidopsis thaliana*

Рівні експресії злитого білка KLIP27-MI (4401) визначають в трансгенному дозрілому насінні *Arabidopsis thaliana* із застосуванням методики прикладу 2 вище. Трансгенний продукт не був знайдений в клітинних екстрактах зрілого насіння.

Приклад 8

Одержання pSBS4409: OLEO-людський проінсулін (OLEO-hPIN) злитого білка

Цей злитий білок починається з 18кДа олеозину з *Arabidopsis thaliana*, далі продовжується геном, що кодує людський проінсулін (hPIN). Експресія цього злитого білка націлена на нові масляні тільця, що утворюються в процесі розвитку ембріона.

Основний ланцюг цієї плазмиди, pSBS4008, оснований на бінарному векторі рослини, pPZP200, описаному у Hajdukiewicz et al. (Plant Molecular Biology, 1994, 25: 989-994). Замість описаного сайту множинного клонування, rat ген, що надає рослині-хазяїну стійкості до фосфінотрицину (Wohleben et al., 1988, Gene 70: 25-37), керований промотором/термінатором убіквітину з *Petroselinum crispum* (Kawalleck et al., 1993, Plant. Mol. Bio., 21: 673-684), був введений між лівою і правою пограничними послідовностями. У доповнення до цієї касети, промотор/термінатор β -фазеоліну з *Phaseolus vulgaris* (Slightom et al., 1983, Proc. Natl. Acad. Sc. USA 80: 1897-1901), керуючий геномною послідовністю *Arabidopsis* 18кДа олеозину, був субклонований. Стандартну PCR (Horton et al., 1989, Gene 77: 61-68) застосовували для злиття послідовності гена олеозину (мінус стоп кодон) з приєднаними сайтами NcoI і HindIII рестрикційної ендонуклеази з 3'-кінця промотору фазеоліну з одержанням pSBS4008.

Ген-HindIII NcoI-людського препоінсуліну синтезують у вигляді одинарного 335 п. н. куска за допомогою Аптагену з застосуванням переважного використання кодону рослини. Подальше лігування у NcoI/HindIII-розрив pSBS4008 дає плазмиду pSBS4400: послідовність ДНК, яка кодує злитий білок олеозин-людський препоінсулін, поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Плазмід pSBS4400 служить як зразок для одержання людського проінсуліну (hPIN) стандартної PCR з використанням рfu ДНК полімерази з праймерами, направленими проти 5'-кінця (1457 оліго TTCTGTGAACCAACACTTG - SEQ ID NO:190) і 3'-кінця (1458 оліго AAGCTTTCAGTTACAGTAGT - SEQ ID NO:191), включаючи сайт HindIII існуючої області проінсуліну вектора. Другий фрагмент збільшують з використанням рfu ДНК полімерази з праймером, направленим проти доступного сайту SphI (оліго 1455 GCATGCATGTGTTGAGC - SEQ ID NO:192) до 3'-кінця гена олеозину *Arabidopsis* (оліго 1456 GGTAGTGTGCTGGCCA - SEQ ID NO:193) у векторі pSBS4400. Після PCR продукти розділяють на агарозному гелі і смуги, які відповідають 267 п. н. (hPIN-HindIII) і 360п.н. (SphI-OLEO (3'-кінець)) фрагменти очищають на

гелі з використанням набору для екстрагування гелю (Qiagen). Два фрагменти зливають вразі збільшення PCR з використанням Taq ДНК полімерази з праймерами 1455 (SEQ ID NO:192) і 1458 (SEQ ID NO:193) у поєднанні з 0,01мкМ перекривного містечкового PCR праймеру (оліго 1459 GGTGGCCAGCACACTACCTTCGTGAACCAACACTTGTG - SEQ ID NO:194) для двох циклів з температурою відпау 58°C з подальшими 31 циклами при температурі 52°C для збільшення 627п.н. SphI-OLEO (3'-кінець)-hPIN-HindIII фрагмента. 627п.н. SphI-OLEO (3'-кінець)-hPIN-HindIII фрагмент потім лігують в T/A виступ pGEMT Easy Vector System™ (Promega) і застосовують для трансформації бактерій DH5α з одержанням pSBS3409 (pGEMT-SphI-OLEO (3'-кінець)-hPIN-HindIII).

Фрагмент SphI/HindIII з pSBS3409 замінюють фрагментом SphI/HindIII з pSBS4400. Стандартне обмеження гідролізації в обох pSBS3409 і pSBS4400 проводять з використанням SphI/HindIII (New England Biolabs). Фрагменти розділяють на 1,5% агарозному гелі і очищають із застосуванням набору для екстрагування гелю (Qiagen). 617п.н. SphI/HindIII фрагмент вивільняють з pSBS3409 і потім лігують в акцепторний сайт SphI/HindIII в заздалегідь розірваному основному ланцюзі вектора pSBS4400 (видалений внутрішній фрагмент SphI/HindIII) із застосуванням T4 ДНК лігази (NEB) протягом ночі при температурі 15°C.

Результатом є плазміда pSBS4409: послідовність ДНК (SEQ ID NO:195, що кодує злитий білок OLEOSIN-hPIN (SEQ ID NO:196), поміщений в бінарний вектор під контролем експресії промотором/термінатором фазеоліну. Промотор фазеоліну контролює специфічно-тимчасову і специфічну до тканин експресію трансгену під час розвитку насіння.

Трансформація і вирощування рекомбінантної *E. coli* і *Agrobacterium* з pSBS4409

Після підтвердження цілісності кДНК кодування для злитого білка аналізом послідовності, плазмиду pSBS4409 трансформують в штам *E. coli* DH5α для одержання високого рівня експресії. Виділену плазмиду ДНК (100нг) змішують на льоду зі 100мкл компетентних клітин DH5α протягом 20хв. Потім клітини піддають тепловому удару при температурі 42°C протягом 45 секунд, і повертають на лід на 2хв. Потім додають 1мл середовища SOC, і клітини інкубують при температурі 37°C на енвіро-шейкері при 225об./хв., протягом 1 години перед поміщенням трансформованих клітин на планшети з LB-спектиноміцином (10г/л триптон, 5г/л екстракту дріжджів, 5г/л NaCl, 15г/л агару) і інкубуванням протягом ночі при температурі 37°C. Одну колонію використовують для інокулювання 5мл LB-спектиноміцинового бульйону. Ці культури вирощують протягом ночі при температурі 37°C. Рекомбінантну плазмиду виділяють з 1мл середовища, вибудовуваного протягом ночі, згідно з одержанням мініпрепаратів Qiagen. Виділену плазмиду потім застосовують для трансформації компетентного штаму *Agrobacterium* EH101 (Hood et al, 1986; J. Bacteriol. 144: 732-743) електропорацією (25мкФ, 2,5кВ, 200Ом). Рекомбінантні *Agrobacterium* вміщують на планшети з АВ-спектиноміцином/канаміцином (20хАВ солі, 2М глюкози, 0,25мг/мл FeSO₄·7H₂O, 1М MgSO₄, 1М CaCl₂) і одну колонію застосовують для інокулювання 5мл АВ-спектиноміцинового/канаміцинового бульйону. Ці культури вирощують протягом ночі при температурі 28°C. Рекомбінантні *Agrobacterium* потім застосовують для трансформації рослин *Arabidopsis thaliana* методом занурення квіток (Clough et al., 1998, Plant J., 16: 735-743), як описано в прикладі 1.

Рівні експресії інсуліну в *Arabidopsis thaliana*

Рівні експресії злитого білка OLEO-hPIN (4409) визначають в трансгенному дозрілому насінні *Arabidopsis thaliana* із застосуванням методики прикладу 2 вище. Білки масляних тілець, мічені забарвленим кумасі гелем, з двох характерних колоній (4409-6 і 4409-8) порівнюють з міграцією злитого білка олеозин-hPIN (як показано чорною стрілкою) у нетрансформованих (wt) *Arabidopsis* (Fig.14). Рівень експресії визначають денситометрією з одержанням, в середньому, 0,10% від загального білка насіння. Цей рівень розраховують вище і нижче співміграції ендогенного білка з такою ж молекулярною масою в нетрансформованому насінні (wt), яке становить приблизно 0,04% від загального білка насіння.

Приклад 9

Трансформація сафлору

Представлений протокол трансформації схожий з описаним у Orlicowska T.K., et al., ((1995) Plant Cell, Tissue and Organ Culture 40: 85-91), але з модифікаціями і поліпшеннями як для трансформації S-317, так і для застосування фосфінотрицину як вибраного маркера. Дезінфікують насіння виду сафлору S-317 California, яке не пошкоджене, подряпане або хворе, в 0,1% HCl₂ протягом 12 хвилин, потім 4-5 разів обполіскують стерильною дистильованою водою. Пророщують стерильне насіння в темряві в середовищі MS (Murashige T. & Skoog F. (1962) Physiol. Plant. 15: 473-497) з 1% сахарозою і 0,25% Gerlite. Ініціюють культури *Agrobacterium* із замороженої сировини в гліцерині в 5мл АВ мінімального рідкого середовища з вибраним антибіотиком, і вирощують протягом 48 годин при температурі 28°C. Вирощують аліквоту цієї культури, вирощеної протягом ночі в 5мл бульйону Luria з вибором для трансформації. Двічі промивають 6-8мл бактерійних клітин середою АВ, і доводять до кінцевої щільності клітин 0,4-0,5 (OD600).

Видаляють дводенні сім'ядолі з пророщеного насіння, занурюють в одержані клітини *Agrobacterium* і поміщують в середовище MS з 3% сахарозою, 4мкм N6-бензиладеніну (BA) і 0,8мкм нафталіноцтової кислоти (NAA). Інкубують планшети при температурі 21 °C в темряві. Через 3 дні переносять в те ж середовище з 300мг/л тиментину. Через ще 4 дні всі культури переносять на світло. Через 3 дні поміщують експлантати в середовище вибору з додаванням 0,5мг/л фосфінотрицину. Для тривалого витягнення бруньок експлантати щотижня переносять в середовище MS без фітогормонів, але із збільшеною в два рази кількістю KNO₃. Зрізають паростки, які витяглися більш ніж на 10мм від початкового рівня, і окремо вирощують для селекції. Для утворення коріння поміщують зелені паростки, що представляють передбачувану трансгенну тканину, в середовище MS з 2% сахарозою, 10мкм індомасляної кислоти і 0,5мкм NAA. Переносять укорінені паростки в добре зволожену суміш ґрунтів і вирощують в умовах високої вологості і 12-годинного світла.

Приклад 10

Протокол трансформації льону

Вказана методика трансформації близька до методики, описаної у Dong J. і McHugen A. (Plant Cell Reports (1991) 10: 555-560), Dong J. і McHugen A. (Plant Sciences (1993) 88: 61-71) і Mlynarova et al. (Plant Cell Reports

(1994) 13: 282-285). Дезінфікують насіння льону, яке не пошкоджене, подряпане або хворе, в 70% етанолі протягом 5-7 хвилин, потім протягом 25 хвилин у 50% відбілюючому розчині з Tween 20 (3-4 краплі на 100мл) при безперервному перемішуванні. Насіння 5-7 разів обполіскують стерильною дистильованою водою. Пророщують стерильне насіння на світлі в середовищі MS (Murashige T. & Skoog F. (1962) *Physiol. Plant.* 15: 473-497) з 2% сахарозою і 0,3% Gerlite в посудинах Magenta. Для трансформації вирощують культури *Agrobacterium* протягом ночі в бульйоні АВ з додаванням відповідного антибіотика для селекції. Двічі промивають 6-8мл витриманих протягом ночі клітин, і повторно суспендують в 5мл АВ бульйони; додають 2мл цієї сировини до 98мл індукційного середовища (основне середовище MS з 3% сахарозою, 5мкм 6-бензиламінопурину (BA) і 0,25мкм альфа-нафталіноцтової кислоти (NAA) і доводять до кінцевої OD₆₀₀ 1,0.

Відсікають гіпокотильні експлантати і інокують в одержаному розчині клітин *Agrobacterium* протягом близько 4 год. (обережно перемішують планшети 1-2 рази під час цього періоду). Після періоду зараження видаляють експлантати з рідкої інокуляційного середовища і наносять на стерильний фільтрувальний папір. Поміщують 15-20 експлантатів в 0,7% отверджене агаром індукційне середовище в планшет для культивування тканин. Герметично закривають планшети пластиковими кришками і спільно культивують експлантати протягом 48 год. при освітленні (23-24°C). Через 2 дні переносять зелені меристематичні експлантати в те ж середовище, що містить 300мг/л Тиментину (середовище для попередньої селекції) і закривають пластиковими кришками. Через 3 дні переносять культури у вказане вище середовище, що містить 10мг/л DL PPT (селекція 1). Закривають планшети Parafilm® і інкубують при температурі 24°C при освітленні. Переносять культури кожні два тижні і зберігають в цьому середовищі протягом одного місяця. Для подовження паростків переносять культури кожні два тижні в середовище для селекції II (основне середовище MS, що містить 2% сахарози, 500мг/л буфера MES, 300мг/л Тиментину і 10мг/л DL PPT) в посудинах Magenta. Передбачувані трансформовані паростки, які виживають після селекції, є темно-зеленими і утворюють значну кількість коріння через 7-10 днів при окремому висаджуванні в середовище для селекції II. Переносять паростки з корінням в стерилізований тепличний ґрунт в невеликих горщиках, і накривають рослини прозорими пластиковими кришками для акліматизації. Для дозрівання переносять активно зростаючі рослини в одногалонові горщики з добре зволоженою сумішшю ґрунтів в тепличних умовах.

Хоч даний винахід описаний за допомогою чітко визначених переважних прикладів, повинно бути зрозуміло, що даний винахід не обмежено описаними прикладами. Навпаки, даний винахід включає різні модифікації і еквівалентні структури, що входять в суть і об'єм представленої формули винаходу.

Всі публікації, патенти і заявки на патенти включені сюди як посилання повністю, в тому ж об'ємі, в якому кожна окрема публікація, патент або заявка на патент визначена і індивідуально включена сюди як посилання.

Таблица

Приклади відомих послідовностей інсуліну

SEQ ID NO	Мотив інсуліну (ідентифікатор послідовності амінокислоти) {ідентифікатор послідовності нуклеїнової кислоти}
Природний людський інсулін	
7	(P01308) людський препроінсулін {містить гени V00565, M10039, J00265, X70508, L15440, BC005255 і AJ009655}
Природний не людський інсулін	
Ссавці	
8	(AAB25818) Проінсулін C-пептид <i>Equus przewalskii</i> (коні, зебри, носороги і тапір (<i>Perissodactyla</i>))
9	(P01310) Препроінсулін <i>Equus caballus</i> (кінь)
10	(P01311) Препроінсулін {включає гени U03610 і M61153} <i>Oryctolagus cuniculus</i> (домашній кролик)
11	(P01312) Інсулін <i>Balaenoptera physalus</i> (смугач)
12	(P01314) Інсулін <i>Balaenoptera borealis</i> (івасевий смугач)
13	(P01315) Препроінсулін {включає гени AF064555 і AY044828} <i>Sus scrofa</i> (свиня)
14	(P01316) Інсулін <i>Elephas maximus</i> (азіатський слон)
15	(P01317) Препроінсулін {ген M54979} <i>Bos Taurus</i> (корова)
16	(P01318) Препроінсулін {ген U00659} <i>Ovis aries</i> (вівця)
17	(P01320) Інсулін <i>Camelus dromedarius</i> (дромадер)
18	(P01321) Препроінсулін {ген V00179} види <i>Canis</i> (собака)
19	(P01328) Інсулін <i>Hystrix cristata</i> (чубатий дикобраз)
20	(P10604) Препроінсулін {ген J02989} <i>Aotus trivigratus</i> (трисмугова дурукулі)
21	(P30406) Препроінсулін {ген J00336} <i>Macaca fascicularis</i> (макак-краббід)
22	(P30407) Препроінсулін {ген X61092} <i>Cercopithecus aethiops</i> (африканська зелена мавпа)
23	(P30410) Препроінсулін {ген X61089} <i>Pan troglodytes</i> (шимпанзе)
24	(Q9TQY7) Інсулін <i>Ornithorhynchus anatinus</i> (качконіс)
25	(AAM76641) Інсулін {ген AY092024} <i>Pongo pygmaeus</i> (орангутанг)
26	(AAN06935) Препроінсулін {ген AN011815, що містить AY137498, AY137499 і AY137500} <i>Gorilla gorilla</i> (горила)
27	(INMKSQ) Інсулін <i>Saimiri sciureus</i> (білячий саймірі)
28	(P01313) Препроінсулін {ген M26328} <i>Cricetulus longicaudatus</i> (довгохвостий хом'як)
29	(P01322) Попередник інсуліну 1 {включає гени V01242, V01242 і M25584} <i>Rattus norvegicus</i> (норвезький щур)

30	(P01323) Попередник інсуліну 2 {включає гени V01243, J00748, M25583 і M25585} <i>Rattus norvegicus</i> (норвезький щур)
31	(P01324) Інсулін <i>Acomys sahirinus</i> (голчата миша)
32	(P01325) Попередник інсуліну 1 {включає гени X04725 і AK007482} <i>Mus musculus</i> (хатня миша)
33	(P01326) Попередник інсуліну 2 {ген X04724} <i>Mus musculus</i> (хатня миша)
34	(P01327) Інсулін <i>Chinchilla brevicaudata</i> (шиншила)
35	(P01329) Препроінсулін {включає гени K02233 і M1 1713} <i>Cavia porcellus</i> (домашня морська свинка)
36	(P17715) Препроінсулін {ген M57671} <i>Octodon degus</i> (дегу)
37	(P18109) Інсулін <i>Didelphis virginiana</i> (північноамериканський опосум)
38	(P21563) Препроінсулін (види <i>Rodentia</i>)
39	(Q62587) Препроінсулін {ген X98241} <i>Psammomys obesus</i> (денна піщанка)
40	(Q91X13) Препроінсулін {ген AY038604} <i>Spermophilus tridecemlineatus</i> (тринадцятисмуговий ховрах)
41	(740063A) Інсулін C пептид <i>Cavia porcellus</i> (домашня морська свинка)
Птахи	
42	(P1332) Препроінсулін {містить гени AH002454 (що містить J00872, J00873 і J00874), V00416, V00418 і X58993} <i>Gallus gallus</i> (курка)
43	(P01333) Препроінсулін <i>Anas platyrhynchos</i> (крячка)
44	(P07454) Інсулін <i>Anser anser anser</i> (сірий гусак)
45	(51463) Препроінсулін {гени AH006925 (що містить S66611 і S66612)} <i>Selasphorus rufus</i> (руда колібрі)
Риби	
46	(073727) Препроінсулін {ген AF036326} <i>Danio rerio</i> (смугастих даніо)
47	(P01335) Препроінсулін {ген X00989} <i>Cyprinus carpio</i> (звичайний карп)
48	(P01337) Інсулін види <i>Batrachoididae</i> (жабоподібні риби)
49	(P01339) Інсулін <i>Thunnus thynnus</i> (синій тунець)
50	(P01340) Інсулін <i>Katsuwonus pelamis</i> (смугастих тунець)
51	(P01341) Препроінсулін {ген V00634} <i>Lophius piscatorius</i> (морський диявол)
52	(P01342) Препроінсулін {ген V00649} <i>Muxine glutinosa</i> (атлантична міксина)
53	(P04667) Препроінсулін {включає гени X00148, J00936, K01655 і X13559} <i>Oncorhynchus keta</i> (кета)
54	(P07453) <i>Myoxocephalus scorpius</i> (гренландський бичок-крючкоріг)
55	(P09476) Інсулін <i>Lepisosteus spatula</i> (панцирна щука)
56	(P09477) Інсулін <i>Platichthys flesus</i> (європейська камбала)
57	(P09536) Інсулін <i>Hydrolagus coliei</i> (плямиста химера)
58	(P12704) Інсулін <i>Squalus acanthias</i> (звичайна колюча акула)
59	(P12705) Препроінсулін <i>Torpedo marmorata</i> (мармуровий електричний скат)
60	(P13190) Проінсулін {ген U82395} <i>Callorhynchus milii</i> (химера)
61	(P14806) Інсулін <i>Petromyzon marinus</i> (морська мінога)
62	(P23187) Інсулін <i>Oncorhynchus gorbuscha</i> (горбуша)
63	(P29335) Інсулін <i>Amia calva</i> (амія)
64	(P42633) Інсулін <i>Anguilla rostrata</i> (американський вугор)
65	(P81025) Препроінсулін {ген AF038123} <i>Oreochromis niloticus</i> (нільська тилапія)
66	(P81423) Інсулін <i>Acipenser gueldenstaedtii</i> (осетер)
67	(P81881) Інсулін <i>Piaractus raesopotamicus</i> (паку)
68	(Q9W7R2) Препроінсулін {ген AB029318} <i>Verasper moseri</i> (камбала)
69	(1603264A) Інсулін C пептид <i>Anguilla anguilla</i> (європейський вугор)
Амфібії	
70	(P12706) Попередник інсуліну 1 {ген M24443} <i>Xenopus laevis</i> (африканська шпорцева жаба)
71	(P12707) Попередник інсуліну 2 {ген M24442} <i>Xenopus laevis</i> (африканська шпорцева жаба)
Рептилії	
72	(P31887) Інсулін <i>Trachemys scripta</i> (червоновуха черепаха)
73	(P12703) Інсулін <i>Alligator mississippiensis</i> (американський алігатор)
74	(P12708) Інсулін <i>Zaocys dumnades</i> (змія)
75	(P01334) Інсулін <i>Crotalus atrox</i> (техаський гримучник)
Генно-інженерний інсулін	
76	(AAA72172) Синтетичний препроінсулін {ген J02547}
77	(AAA72916) Альфа ланцюг 3'-кінця синтетичного інсуліну {гени AH003171 або M38610}
78	(AAA72917) Бета ланцюг 3'-кінця синтетичного інсуліну {гени AH003171 або M38611}
79	(CAA00712) Синтетичний інсулін {ген A07755}
80	(CAA00713) Синтетичний інсулін {ген A07758}
81	(CAA00714) Синтетичний інсулін {ген A07761}
82	(CAA00715) Білковий продукт без назви {ген A07764}
83	(CAA00736) Синтетичний проінсулін {ген A08012} (EP 0367163-A)
84	(CAA00783) Синтетичний інсулін {ген A08468} (EP 0376156-A)
85	(CAA01581) Модифікований попередник інсуліну {ген A21951} (WO 9011299)

86	(CAA01254) Синтетичний інсулін {ген A15938} (EP 0214826-A)
87	(CAA01799) Синтетична конструкція інсуліну Asp (B1), Asp (B4), Asp (B10), Asp (B16), Glu (B27) {ген A26317}
88	(CAA01798) Синтетична конструкція попередника інсуліну Glu (B9), Glu (A12) {ген A26314}
89	(CAA23424) Синтетичний проінсулін {ген V00082}
90	(CAA24707) С-ланцюг синтетичного інсуліну {ген V01461}
91	(CAA25151) В-ланцюг синтетичного інсуліну {ген X00462}
92	(CAD60056) Синтетичний білковий продукт без назви {ген AX573757} (пат. WO 02/079250)
93	(ген M31026) В- і міні-С-ланцюги синтетичного людського інсуліну із застосуванням хроматографії на дезактивованому силікагелі
94	(1BZVA) Ланцюг А [d-A1ab26]-Des(B27-B30)-інсулін-B26-амід А суперпотужного аналога інсуліну з одним заміщенням
95	(1BZVB) Ланцюг В [d-A1ab26]-Des(B27-B30)-інсулін-B26-амід А суперпотужного аналога інсуліну з одним заміщенням
96	(1HUIA) Ланцюг А мутанта інсуліну (B1, B10, B16, B27) glu, Des-B30, Nmr
97	(1HUIB) Ланцюг В мутанта інсуліну (B1, B10, B16, B27) glu, Des-B30, Nmr
98	(1HLSA) Ланцюг А мутанта людського інсуліну - His (B16)
99	(1HLSB) Ланцюг В мутанта людського інсуліну - His (B16)
100	(1JCAA) Ланцюг А нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
101	(1JCAB) Ланцюг В нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
102	(1JCAC) Ланцюг С нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
103	(1JCAD) Ланцюг D нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
104	(1J73A) Ланцюг А нестабільного аналога інсуліну з природною активністю
105	(1J73B) Ланцюг В нестабільного аналога інсуліну з природною активністю
106	(1J73C) Ланцюг С нестабільного аналога інсуліну з природною активністю
107	(1J73D) Ланцюг D нестабільного аналога інсуліну з природною активністю
108	(1KMFA) Ланцюг А мутанта людського інсуліну Ile-A2-Allo-Ile, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
109	(1KMFB) Ланцюг В мутанта людського інсуліну Ile-A2-Allo-Ile, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
110	(1K3MA) Ланцюг А мутанта людського інсуліну Ile-A2-Ala, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
111	(1K3MB) Ланцюг В мутанта людського інсуліну Ile-A2-Ala, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
112	(1LW8A) Ланцюг А Allo-Ilea2-інсуліну, неактивного хірального аналога
113	(1LW8B) Ланцюг В Allo-Ilea2-інсуліну, неактивного хірального аналога
114	(1LW8C) Ланцюг С Allo-Ilea2-інсуліну, неактивного хірального аналога
115	(1LW8D) Ланцюг D Allo-Ilea2-інсуліну, неактивного хірального аналога
116	(1LKQA) Ланцюг А мутанта людського інсуліну Ile-A2-Gly, Val-A3-Gly, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
117	(1LKQB) Ланцюг В мутанта людського інсуліну Ile-A2-Gly, Val-A3-Gly, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
118	(1MHIA) Ланцюг А B9 (Asp) мутанта
119	(1MHIB) Ланцюг В B9 (Asp) мутанта
120	(1MHJA) Ланцюг А B25 (phe) мутанта
121	(1MHJB) Ланцюг В B25 (phe) мутанта
122	(1VKTA) Ланцюг А, людський інсулін, модель з двома дисульфідами
123	(1VKTB) Ланцюг В, людський інсулін, модель з двома дисульфідами
124	(AAG59607) Синтетичний альбебетин інсулін {ген AY017185}
125	(AAG59606) Синтетичний альбеферон інсулін {ген AY017184}
Злиті білки інсуліну	
Людські	
126	(AAB27046) N-кінець злитого білка інтерлейкін 2-інсуліну
127	(AAB27047) N-кінець злитого білка бета-галактозидаза-інсулін
128	(PC7082) Злитий білок фактора росту епідермісу/Des-B30 попередника людського інсуліну з одним ланцюгом
Не людські	
129	(AAA72177) Злитий білок E.coli пеніцилінази/інсулін I щура 5'-кінця {ген AH003149 або J02553}
130	(AAA72178) Злитий білок E.coli пеніцилінази/інсулін I щура 3'-кінця {ген AH003149 або J02554}
131	(AAA72179) Злитий білок сигнальної послідовності інсуліну щура/E.coli бета-галактозидази {ген J02555}
132	(AAA72181) Злитий білок мавпячого вірусу 40 (SV40)/препроінсуліну I щура {ген J02559}
Міні-інсулін	

Людський	
133	(1EFEA) Ланцюг А активного міні-проінсуліну, M2pi
134	(1JCAA) Ланцюг А нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
135	(1JCAB) Ланцюг В нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
136	(1JCAC) Ланцюг С нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
137	(1JCAD) Ланцюг D нестандартного дизайну нестабільних аналогів інсуліну з поліпшеною активністю
138	(1JK8C) Ланцюг С пептиду людського інсуліну - Hla-Dq8 комплекс
139	(1J73A) Ланцюг А нестабільного аналога інсуліну з природною активністю
140	(1J73B) Ланцюг В нестабільного аналога інсуліну з природною активністю
141	(1J73C) Ланцюг С нестабільного аналога інсуліну з природною активністю
142	(1J73D) Ланцюг D нестабільного аналога інсуліну з природною активністю
143	(1SJTA) Ланцюг А міні-проінсуліну, мутант аналога інсуліну з двома ланцюгами: Des B30, His (B10) asp, Pro (B28) asp
144	(1SJTB) Ланцюг В міні-проінсуліну, мутант аналога інсуліну з двома ланцюгами: Des B30, His (B10) asp, Pro (B28) asp
145	(1SJU) Міні-проінсулін, мутант аналога інсуліну з одним ланцюгом: Des B30, His (B10) asp, Pro (B28) asp і пептидний зв'язок між Lys B29 і Gly A1

Список послідовностей

SEQ ID NO:1 і 2 є нуклеотидною послідовністю і виведеною послідовністю амінокислот, відповідно, PRS-D9scFv-KLIP27-MI-KDEL злитого білка в плазміді pSBS4404.

SEQ ID NO:3 і 4 є нуклеотидною послідовністю і виведеною послідовністю амінокислот, відповідно, OLEO-KLIP8-KLIP27-MI злитого білка в плазміді pSBS4405.

SEQ ID NO:5 і 6 є нуклеотидною послідовністю і виведеною послідовністю амінокислот, відповідно, PRS-MI-тетраосновний линкер-D9scFv-KDEL злитого білка в плазміді pSBS4414.

SEQ ID NO:7-145 є відомими послідовностями інсуліну, які описані в таблиці.

SEQ ID NO:146-148 є послідовностями амінокислот фрагментів С-пептиду інсуліну.

SEQ ID NO:149 є послідовністю амінокислот тетраосновного процесингового пептиду.

SEQ ID NO:150-155 є послідовностями амінокислот поліпептидів, здатних утримувати поліпептид інсуліну в ER.

SEQ ID NO:156-160 є послідовностями амінокислот поліпептидів, здатних утримувати поліпептид інсуліну в ER-похідній накопичувальній органелі.

SEQ ID NO:161 є послідовністю амінокислот PRS сигнальної послідовності.

SEQ ID NO:162-171 є послідовностями амінокислот лідерних послідовностей дріжджів і похідних від них послідовностей.

SEQ ID NO:172-173 є послідовностями амінокислот проміжних пептидів.

SEQ ID NO:174 є послідовністю амінокислот послідовності KLIP8.

SEQ ID NO:175 є послідовністю амінокислот прямого праймеру 1325, який є комплементарним до 5'-області D9scFv кДНК клону і існує для додавання сайту SphI до 5'-області для сприяння подальшому лігуванню.

SEQ ID NO:176 є послідовністю амінокислот зворотного праймеру 1326, який є комплементарним до 3'-області D9scFv кДНК клону і існує для додавання сайту XhoI до 3'-області для сприяння подальшому лігуванню.

SEQ ID NO:177 є послідовністю амінокислот прямого праймеру 1324, який є комплементарним до області 20 нуклеотиду зворотного праймеру 1323 і існує для утворення 5'-кінця злитого Klip27-MI.

SEQ ID NO:178 є послідовністю амінокислот зворотного праймеру 1323, який є комплементарним до області 20 нуклеотиду прямого праймеру 1324 і існує для утворення 5'-кінця злитого Klip27-MI.

SEQ ID NO:179 є послідовністю амінокислот прямого праймеру 1322, який є комплементарним до області 19 нуклеотиду зворотного праймеру 1321 і існує для утворення 3'-кінця злитого Klip27-MI.

SEQ ID NO:180 є послідовністю амінокислот зворотного праймеру 1321, який є комплементарним до області 19 нуклеотиду прямого праймеру 1322 і існує для утворення 3'-кінця злитого Klip27-MI.

SEQ ID NO:181 є послідовністю амінокислот прямого праймеру 1364, який є комплементарним до 5'-області Klip27-MI послідовності і існує для додавання сайту XhoI до 5'-області для сприяння подальшому лігуванню.

SEQ ID NO:182 є послідовністю амінокислот прямого праймеру 1334, який є комплементарним до 3'-області Klip27-MI послідовності і існує для додавання сайту HindI до 3'-області для сприяння подальшому лігуванню.

SEQ ID NO:183 є послідовністю амінокислот прямого праймеру 1329, який є комплементарним до 3'-області Klip27-MI послідовності і існує для додавання сайту HindIII до 3'-області для сприяння подальшому лігуванню.

SEQ ID NO:184 є послідовністю амінокислот прямого праймеру 1363, який є комплементарним до 5'-області Klip27-MI послідовності і існує для додавання сайту SphI до 5'-області для сприяння подальшому лігуванню.

SEQ ID NO:185 є послідовністю амінокислот прямого праймеру 1515, який є комплементарним до 5'-області ланцюга В послідовності інсуліну і існує для введення проміжного тетраосновного сайту між вихідними

A і В ланцюгами людського інсуліну в поєднанні із зворотним праймером 1518.

SEQ ID NO:186 є послідовністю амінокислот зворотного праймеру 1518, який є комплементарним до 3'-області ланцюга В інсуліну і 5'-області ланцюга А з проміжною послідовністю тетраосновного міні-с-пептиду і існує для введення проміжного тетраосновного сайту між вихідними А і В ланцюгами людського інсуліну.

SEQ ID NO:187 є послідовністю амінокислот зворотного праймеру 1517, який є комплементарним до 3'-області D9scFv/KDEL послідовності і існує для збільшення всього MI-тетраосновний лінкер-D9Scfv-KDEL з одержанням вставки pSBS4414.

SEQ ID NO:188 і 189 є нуклеотидною послідовністю і виведеною послідовністю амінокислот, відповідно, PRS-KLIP27-MI злитого білка в плазміді PSBS4401.

SEQ ID NO:190 є послідовністю амінокислот прямого праймеру 1457, який є комплементарним до 5'-області В ланцюга послідовності інсуліну і існує для одержання фрагмента людського проінсуліну (hPIN) в поєднанні із зворотним праймером 1591.

SEQ ID NO:191 є послідовністю амінокислот зворотного праймеру 1458, який є комплементарним до 3'-області людського проінсуліну (hPIN) і існує для одержання людського про (hPIN) і додання 3' HindIII сайту клонування.

SEQ ID NO:192 є послідовністю амінокислот прямого праймеру 1455, який є комплементарним до 5'-області сайту SphI pSBS4404 і існує для збільшення гена олеозину Arabidopsis в поєднанні із зворотним праймером 1456.

SEQ ID NO:193 є послідовністю амінокислот зворотного праймеру 1456, який є комплементарним до 3'-області сайту гена олеозину Arabidopsis і існує для збільшення гена олеозину Arabidopsis в поєднанні з прямим праймером 1455.

SEQ ID NO:194 є послідовністю амінокислот перекривного містечкового PCR праймеру, який є комплементарним до 3'-області сайту гена олеозину Arabidopsis і 5'-кінця гена людського проінсуліну і існує для одержання вставки pSBS4409 в поєднанні з прямим праймером 1455 і зворотним праймером 1456.

SEQ ID NO:195 і 196 є нуклеотидною послідовністю і виведеною послідовністю амінокислот, відповідно, OLEO-hPIN злитого білка в плазміді pSBS4409.

SEQ ID NO:1

atgaacttccttaagtctttccctttctacgctttcctttgtttcggtcaatacttcggttgc
gttacgcattgctgacattgtgatgacacagctctccatcctccctggctatgtcagtgaggacag
cgggtcactatgcgctgcaagtccagtcagagccttttaaaaagtaccaatcaaaagaactat
ttggcctggtagcagcagaaaccaggacagctctcctaaacttctgggtatactttgcatccact
aggggaatctggggctccctgatcgcttcattagggcagtggtatctgggacagatttcactcttacc
atcagcagtggtgcaggctgaagacctggcagattacttctgtcagcaacattataaacactcct
cccacgttcggtgctgggaccaagctggagcttaagcggctctccgaacgggtgcttctcatagc
ggttctgcaccaggcactagctctgcatctggatctcaggtgcacctgcagcagctctggagct
gagctgatgaagcctggggcctcaatgaagatatcctgcaaggctactggctacacattcagt
agctactggatagagtgggtaaagcagaggcctggacatggccttgagtggattggagagatt
ttacctggcagtggttagtactacctacaatgagaagttcaagggcaaggccacattcactgca
gatacatcctccaacacagcctacatgcaactcagcagcctgacatctgaggactctgccgtc
tattactgtgcaagattggatggtgactcctggggccaaggcaccactctcacagctctcgagt
caaccaattgatgacactgaatcccagaccacgtcagtgaaacctcatggccgatgatactgag
agcgcggttgctacacaaaacaaattcgggaggtcttgacgttgctcggttgatctccatggct
aagagagaagaaggagagcctaagtttggttaatacaacatctttgtggatctcatcttgttgag
gctctctaccttggtgtgtggagaaagaggatttttctacactcctaaggctgctaagggaatt
gttgaacaatggtgcacttctatttgcctcactttaccaattggagaactattgcaacaaggat
gaactttga

SEQ ID NO:2

MNFLKSFPFYAFLCFGQYFVAVTHADIVMTQSPSSLAMSVGQRVTMRCKSSQSLLKSTNQKNY
LAWYQQKPGQSPKLLVYPFASTRESGVPRFIGSGSGTDFLTITSSVQAEDLADYFCQOHYNTP
PTFGAGTKLELKRSPNGASHSGSAPGTSSASGSQVHLQSGAELMKPGASMKISCKATGYTFS
SYWIEWVKQRPGHGLEWIGEILPGSGSTTYNEKFKGKATFTADTSSNTAYMQLSSLTSEDSAV
YYCARLDVDSWGQGTTLTVSSQPIDDTESQTTSVNLMADDTESAFATQTNSSGLDVVGLISMA
KREEGEPKFVNQHLGSHLVEALYLVCGERGFFYTPKAAGIVEQCCTSIICSLYQLENYCNKD
EL

SEQ ID NO:3

atggcgggatacagctagaggaacccatcacgatatcatcggcagagaccagtagccgatg
atgggcccagagaccagagaccagtagccgatgtccggacgaggatctgactactccaagtct
aggcagattgctaaagctgcaactgctgtcacagctgggtgggtccctccttgttctctcc
agccttacccttgttggaaactgtcatagctttgactggtgcaaacacctctgctcggtatc
ttcagcccaatccttgtcccggctctcatcacagttgcactcctcatcacagggttttctt
tctctggagggtttggcattgcccgtataaccgttttctcttggatttacgcaacggga
gagcaccacagggatcagacaagttggacagtgcaaggatgaagttgggaagcaaagct
caggatctgaaagacagagctcagtagctacggacagcaacatactgggtggggaacatgac
cgtgaccgtactcgtgggtggccagcacactaccatggctgagatcacccgcatctcctc
tacaaaggtaagtctctccgtaaggcgctgaaggaacatggacttctagaagacttcttg
cagaaacaacagtaggcatctcgagcaagttccaaccaattgatgacactgaatcccag
accacgtcagtgaaacctcatggccgatgatactgagagcggtttgctacacaaaacaaat
tcgggagggtcttgacgttgctggattgatctccatggctaagagagaagaaggagagcct
aagtttgtaatacaacatcttctgtggatctcatcttgttgaggctctctaccttggtgtg
ggagaaagaggatttttctacactcctaaggctgctaagggaattggtgaacaatggtgc
acttctatttgcctcactttaccaattggagaactattgcaactga

SEQ ID NO:4

MADTARGTHHDIIGRDQYPMGRDRDQYQMSGRGSDYSKSRQIAKAATAVTTAGGSLVLSSLT
LVGTVIALTVATPLLVIFSPILVPALITVALLITGFLSSGGFGIAAITVFSWIYATG
EHPQGSCLKDSARMKLGSKAQDLKDRAQYQGQHTTGGEHDRDRTRGGQHTTMAEITRIPLYKG
KSLRKALKEHGLLEDFLQKQYGISSKFQPIDDTESQTTSVNLMADDTESAFATQTN

SGGLDVVGLISMAKREEGEPKFVNQHLGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI
CSLYQLENYCN

SEQ ID NO:5

atgaacttccttaagtctttccctttctacgctttcctttgtttcggtcaatacttcggtgct
gttacgcatgcctttgttaatacaacatctttgtggatctcatcttgaggctctctacctt
gtgtgtggagaaagaggatttttctacactcctaagactagaagaaagagaggaattgttgaa
caatgttgcaacttctatttgctcactttaccaattggagaactattgcaacagaagaaagaga
gacattgtgatgacacagtcctcatcctcctggctatgtcagtgaggacagcgggtcactatg
cgctgcaagtcagtcagagccttttaaaaagtaccaatcaaaagaactatttgccctggtac
cagcagaaaccaggacagtcctctaaacttctggtatactttgcatccactaggggaatctggg
gtccctgatcgcttcatagggcagtggtatctgggacagatttcactcttaccatcagcagtggtg
caggtgaagacctggcagattacttctgtcagcaacattataacactcctcccacgttcggt
gctgggaccaagtggagcttaagcggctctccgaacgggtgcttctcatagcgggttctgcacca
ggcactagctctgcatctggatctcaggtgcacctgcagcagtcctggagctgagctgatgaag
cctggggcctcaatgaagatatcctgcaaggctactggctacacattcagtagctactggata
gagtgggttaaagcagaggcctggacatggccttgagtggttgagagattttacctggcagtc
ggtagtactacctacaatgagaagttcaagggaaggccacattcactgcagatacatcctcc
aacacagcctacatgcaactcagcagcctgacatctgaggactctgccgtctattactgtgca
agattggatgttgactcctggggccaaggcaccactctcacagtgagctcaaaggatgagctt
tga

SEQ ID NO:6

MNFLKSFFFYAFLCFGQYFVAVTHAFVNQHLGSHLVEALYLVCGERGFFYTPKTRRKR
GIVEQCCTSI
CSLYQLENYCN
NRRKRDIVMTQSPSSLAMSVGQQRVTMRCKSSQSLLKSTNQKNYLAWY
QQKPGQSPKLLVYFASTRESGVPDRFIGSGSGTDFTLTISSVQAEDLADYFCQQHYNT
PPTFGAGTKLELKRSPNGASHSGSAPGTSSASGSQVHLQQSGAELMKPGASMKISCKAT
GYTFSSYWI
EWWKQRPGHGLEWIGELPGSGSTTYNEKFKGKATFTADTSSNTAYMQLSSSLTSEDS
AVYYCARLDVDSWGQGTTLTVSSKDEL

SEQ ID NO:7

MALWMRLPLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQV
GQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI
CSLYQLENYCN

SEQ ID NO:8

EAEDPQVGEVELGGGPGLGGLQPLALAGPQQ

SEQ ID NO:9

FVNQHLGSHLVEALYLVCGERGFFYTPKAXXEADPQVGEVELGGGPGLGGLQPLALAGPQQ
XXGIVEQCCTGICSLYQLENYCN

SEQ ID NO:10

MASLAALLPLALLLVLCRLDPAQAFVNQHLGSHLVEALYLVCGERGFFYTPKSRREVEELQV
GQAEELGGGPGAGGLQPSALELALQKRGIVEQCCTSI
CSLYQLENYCN

SEQ ID NO:11

GIVEQCCTSI
CSLYQLENYCN

SEQ ID NO:12

GIVEQCCAST
CSLYQLENYCN

SEQ ID NO:13

MALWTRLLPLALLALWAPAPAQAFVNQHLGSHLVEALYLVCGERGFFYTPKARREAENPQA
GAVELGGGLGGLQALALEGPPQKRGIVEQCCTSI
CSLYQLENYCN

SEQ ID NO:14
GIVEQCCTGVCSLYQLENYCN

SEQ ID NO:15
MALWTRLRPLLALLALWPPPPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPQV
GALELAGGPGAGGLEGPQKRGIVEQCCASVCSLYQLENYCN

SEQ ID NO:16
MALWTRLVPLLALLALWAPAPAHAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPQV
GALELAGGPGAGGLEGPQKRGIVEQCCAGVCSLYQLENYCN

SEQ ID NO:17
GIVEQCCASVCSLYQLENYCN .

SEQ ID NO:18
MALWMRLLPLLALLALWAPAPTRAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEDLQV
RDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:19
GIVDQCCTGVCSLYQLQNYCN

SEQ ID NO:20
MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDLQV
GQVELGGGSITGSLPPLGPMQKRGVVDQCCTSICSLYQLQNYCN

SEQ ID NO:21
MALWMRLLPLLALLALWGPDPAFAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQV
GQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:22
MALWMRLLPLLALLALWGPDVPFAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQV
GQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:23
MALWMRLLPLLALLALWGPDPAFAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQV
GQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:24
GIVEECCKGVCSMYQLENYCN

SEQ ID NO:25
CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVE
QC

SEQ ID NO:26
MALWMRLLPLLALLALWGPDPAFAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQV
GQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:27
FVNQHLCGPHLVEALYLVCGERGFFYAPKTGVVDQCCTSICSLYQLQNYCN

SEQ ID NO:28
MTLWMRLLPLLALLLWEPNPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKSRRGVEDPQV
AQLELGGGPGADDLQTLALEVAQKRGIVDQCCTSICSLYQLENYCN

SEQ ID NO:29
MALWMRFLPLLALLVLWEPKPAQAFVKQHLGPHLVEALYLVCGERGFFYTPKSRREVEDPQV
PQLELGGGPEAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID NO:30
MALWIRFLPLLALLILWEPRPAQAFVKQHLGSHLVEALYLVCGERGFFYTPMSRREVEDPQV
AQLELGGGPGAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID NO:31
GIVDQCCTSICSLYQLENYCN

SEQ ID NO:32
MALLVHFLPLLALLALWEPKPTQAFVKQHLGPHLVEALYLVCGERGFFYTPKSRREVEDPQV
EQLELGGSPGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID NO:33
MALWMRFLPLLALLFLWESHPTQAFVKQHLGSHLVEALYLVCGERGFFYTPMSRREVEDPQV
AQLELGGGPGAGDLQTLALEVAQQKRGIVDQCCTSICSLYQLENYCN

SEQ ID NO:34
GIVDQCCTSICTLYQLENYCN

SEQ ID NO:35
MALWMHLLTVLALLALWGPNTGQAFVSRHLGCSNLVETLYSVCQDDGFFYIPKDRRELEDPQV
EQTELGMGLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHQLQSYCN

SEQ ID NO:36
MAPWMHLLTVLALLALWGPNSVQAYSSQHLGCSNLVEALYMTCGRSGFYRPHDRRELEDLQVE
QAEGLGLEAGGLQPSALEMILQKRGIVDQCCNNICTFNQLQNYCNVP

SEQ ID NO:37
GIVEQCCNSICSLYQLETYCN

SEQ ID NO:38
MALWILLPLLALLILWGPDPQAFVNQHLGSHLVEALYILVCGERGFFYTPMSRREVEDPQV
GQVELGAGPGAGSEQTLALEVARQARIVQQCTSGICSLYQLENYCN

SEQ ID NO:39
MALWMRLLPLLAFLLILWEPSPAHAFVNQHLGSHLVEALYLVCGERGFFYTPKFRRGVDDPQM
PQLELGGSPGAGDLRALALEVARQKRGIVEQCCTGICSLYQLENYCN

SEQ ID NO:40
MALWTRLLPLLALLALLGPDPAQAFVNQHLGSHLVEALYLVCGERGFFYTPKSRREVEEQQG
GQVELGGGPGAGLPQPLALEMALQKRGIVEQCCTSICSLYQLENYCN

SEQ ID NO:41
ELEDPQVEQTELGMGLGAGGLQPLQALQ

SEQ ID NO:42
MALWIRSLPLLALLVFSGPGETSYAAANQHLGSHLVEALYLVCGERGFFYSPKARRDVEQPLV
SSPLRGEAGVLFPQQEYEVKVRGIVEQCCHNTCSLYQLENYCN

SEQ ID NO:43
AANQHLGSHLVEALYLVCGERGFFYSPKTXXDVEQPLVNGPLHGEVGEPLPFQHEEYQXXGIV
EQCCENPCSLYQLENYCN

SEQ ID NO:44
GIVEQCENPCSLYQLENYCN

SEQ ID NO:45
IQSLPLLALLALSGPGTSHAAVNQHLGSHLVEALYLVCGERGFFYSPKARRDAEHPLVNGPL
HGEVGDLPFQQEEFEKVKRGIVEQCCHNTCSLYQLENYCN

SEQ ID NO:46
MAVWLQAGALLVLLVSSVSTNPGTPQHLCGSHLVDALYLVCGPTGFFYNPKRDVEPLLGLFLP
PKSAQETEVDFAFKDHAELIRKRGIVEQCCHKPCSI FELQNYCN

SEQ ID NO:47
MAVWIQAGALLFLLAVSSVNANAGAPQHLCGSHLVDALYLVCGPTGFFYNPKRDVDPPLGLFLP
PKSAQETEVDFAFKDHAEVIRKRGIVEQCCHKPCSI FELQNYCN

SEQ ID NO:48
GIVEQCCHRPCDIFDLQSYCN

SEQ ID NO:49
GIVEQCCHKPCNIFDLQNYCN

SEQ ID NO:50
GIHZZCCHKPCBIFZLZBYCN

SEQ ID NO:51
MAALWLQSFSLVLLVSWPGSQAVAPAQHLCGSHLVDALYLVCGRGFFYNPKRDVDQLLGF
LPPKSGGAAAAGADNEVAEFAFKDQMEMMVKRGIVEQCCHRPCNIFDLQNYCN

SEQ ID NO:52
MALSPFLAAVIPLVLLLSRAPPSADTRTTGHLGKDLVNALYIACGVRGFFYDPTKMKRDTGA
LAAFLPLAYAEDNESQDDESIGINEVLKSKRGIVEQCCHKRCSIYDLENYCN

SEQ ID NO:53
MAFWLQAASLLVLLALSPGVDAAAAQHLCGSHLVDALYLVCGEKGFFYTPKRDVDPLIGFLSP
KSAKENEEYPFKDQTEMMVKRGIVEQCCHKPCNIFDLQNYCN

SEQ ID NO:54
GIVEQCCHRPCNIRVLENYCN

SEQ ID NO:55
GIVEQCCHKPCTIYELENYCN

SEQ ID NO:56
GIVEQCCHKPCNIFDLQNYCN

SEQ ID NO:57
GIVEQCCHNTCSLANLEGYCN

SEQ ID NO:58
GIVEHCCHNTCSLYDLEGYCNQ

SEQ ID NO:59
GIVEHCCHNTCSLFDLEGYCN

SEQ ID NO:60
VPTQRLCGSHLVDALYFVCGERGFFYSPKQIRDVGPLSAFRDLEPPLDTEMEDRFPYRQQLAG
SKMKRGIVEQCCHNTCSLVNLEGYCN

SEQ ID NO:61
GIVEQCCHRKCSIYDMENYCN

SEQ ID NO:62
GIVEQCCHKPCNIFDLQNYCN

SEQ ID NO:63
GIVEQCCLKPCTIYEMEKYCN

SEQ ID NO:64
GIVEQCCHKPCSIFDLQNYCN

SEQ ID NO:65
MAALWLQAFSLLVLMVSWPGSQAVGGPQHLCGSHLVDALYLVCGDRGFFYNPRRDVDPLLGF
LPPKAGGAVVQGGENEVTFKDQMEMMVKRGIVEECCHKPCTIFDLQNYCN

SEQ ID NO:66
GIVEQCCHSPCSLYDLENYCN

SEQ ID NO:67
GIVEQCCHKPCSIFDLQNYCN

SEQ ID NO:68
MAALWLQSVSLLVLMVSWSGSQAVLPPQHLCGAHLVDALYLVCGERGFFYTPKRDVDPLLGF
LPAKSGGAAAGGENEVAEFAFKDQMEMMVKRGIVEQCCHKPCNIFDLQNYCN

SEQ ID NO:69
DVEPLLGLFLSPKSGQENEVDFFPYKGQGEL

SEQ ID NO:70
MALWMQCLPLVLVLFSTPNTEALVNQHLCGSHLVEALYLVCGDRGFFYYPKVVRDMEQALVS
GPQDNELDGMQLQPQEQKMKRGIVEQCCHSTCSLFQLESYCN

SEQ ID NO:71
MALWMQCLPLVLVLLFSTPNTEALANQHLCGSHLVEALYLVCGDRGFFYYPKIKRDIEQAQVN
GPQDNELDGMQFQPQEQKMKRGIVEQCCHSTCSLFQLENYCN

SEQ ID NO:72
GIVEQCCHNTCSLYQLENYCN

SEQ ID NO:73
GIVEQCCHNTCSLYQLENYCN

SEQ ID NO:74
GIVEQCCENTCSLYELENYCN

SEQ ID NO:75
GIVEQCCENTCSLYQLENYCN

SEQ ID NO:76
MGLWIRLLPLIALLLILWGPDPAAAEFRMFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAE

DLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:77
GIVEQCCTSI CSLYQLENYCN

SEQ ID NO:78
FVNQHL CGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:79
FVDQHL CGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYELEDYCN

SEQ ID NO:80
FVEQHL CGSDLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYQLEEYCN

SEQ ID NO:81
FVQQHL CGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYQLENYCG

SEQ ID NO:82
FVTQHL CGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYQLEHYCS

SEQ ID NO:83
NSNGKFVNQHL CGSHLVEALYLVCGERGFFYTPKTKGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:84
NSNGKFVNQHL CGSHLVEALYLVCGERGFFYTPKTKRGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:85
FVNQHL CGSHLVEALYLVCGERGFFYTPKGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:86
RFVNQHL CGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:87
KETLTITCAVPTWLKLWTFWFAVKEVSSTNLRLRLVLSNNAVPPSAPCTNWKTTATRRSPQA

SEQ ID NO:88
KDSL TNTCAVSTWLKLCTWFAVKEVSSTLLRLRLVLSNNAVPPSANYTNWKTTATRRSPQA

SEQ ID NO:89
MFVNQHL CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:90
RREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR

SEQ ID NO:91
GPETLCGAELVDALQFVCGDRGF

SEQ ID NO:92
MKLKT VRS AVLSSLFASQVLGQPIDDTESQTTSVNL MADDTESAFATQTNSGGLDVVGLISMA
KREEGEPKFVNQHL CGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI CSLYQLENYCN

SEQ ID NO:93
tttgtcaatcagcacctttgtggttctcacctggaggctctgtacctgggtgtgtggggaa
cgtggtttcttctacacaccaagaccgcgtgaagcttaagcgtggcattgtggagcagtgc

tgcaccagcatctgctccctctaccaactggagaactactgcaac

SEQ ID NO:94
GIVEQCCTSICSLYQLENYCN

SEQ ID NO:95
FVNQHLCGSHLVEALYLVCGERGFFX

SEQ ID NO:96
GIVEQCCTSICSLYQLENYCN

SEQ ID NO:97
EVNQHLCGSELVEALELVCGERGFFYEPK

SEQ ID NO:98
GIVEQCCTSICSLYQLENYCN

SEQ ID NO:99
FVNQHLCGSHLVEALHLVCGERGFFYTPKT

SEQ ID NO:100
GIVEQCCKSICSLYQLENYCN

SEQ ID NO:101
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:102
GIVEQCCKSICSLYQLENYCN

SEQ ID NO:103
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:104
GIVEQCCXSICSLYQLENYCN

SEQ ID NO:105
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:106
GIVEQCCXSICSLYQLENYCN

SEQ ID NO:107
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:108
GXVEQCCTSICSLYQLENYCN

SEQ ID NO:109
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID NO:110
GAVEQCCTSICSLYQLENYCN

SEQ ID NO:111
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID NO:112
GXVEQCCTSICSLYQLENYCN

SEQ ID NO:113
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:114
GXVEQCCTSICSLYQLENYCN

SEQ ID NO:115
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:116
GGGEQCCTSICSLYQLENYCN

SEQ ID NO:117
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID NO:118
XIVEQCCTSICSLYQLENYCN

SEQ ID NO:119
XVNQHLCGDHLVEALYLVCGERGFFYTPKT

SEQ ID NO:120
GIVEQCCTSICSLYQLENYCN

SEQ ID NO:121
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:122
GIVEQSCTSISSLYQLENYCN

SEQ ID NO:123
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID NO:124
MDPGDPECLEQLLRRLGGSVEVEVTGGTVHVEVSPEDPGDPECLEQLLRRLGGSVEVEVTGGT
VHVEVSPGERGFFYCN

SEQ ID NO:125
MLKEKKYSPDPGDPECLEQLLRRLGGSVEVEVTGGTVHVEVSPEDPGDPECLEQLLRRLGGSV
EVEVTGGTVHVEVSPGERGFFYCN

SEQ ID NO:126
MATSXSTKKTQLQLEHLXLDLQM

SEQ ID NO:127
TMITDSLAVVLQRXDWXPGVTQL

SEQ ID NO:128
NSVLASALALTVAPMAFANSDESPLSHDGYSLHDGVSMYIEALDKFVNQHLCGSHLVEALYL
VCGERGFFYTPKGIVEQCCTSICSLYQLENYCN

SEQ ID NO:129

DTTMPAGGGGGGQHL CGPHLVEALY

SEQ ID NO:130
LENYCN

SEQ ID NO:131
MTMITDSLEFQAWGGGGGWMRF

SEQ ID NO:132
MVLRFLLPLLALLVLWEPAQA

SEQ ID NO:133
FVNQHLCGSHLVEALYLVCGERGFFYTPKTRRYPGDVKGIVEQCCTSICSLYQLENYCN

SEQ ID NO:134
GIVEQCCKSICSLYQLENYCN

SEQ ID NO:135
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:136
GIVEQCCKSICSLYQLENYCN

SEQ ID NO:137
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:138
LVEALYLVCGERGG

SEQ ID NO:139
GIVEQCCXSICSLYQLENYCN

SEQ ID NO:140
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:141
GIVEQCCXSICSLYQLENYCN

SEQ ID NO:142
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID NO:143
GIVEQCCTSICSLYQLENYCN

SEQ ID NO:144
FVNQHLCGSDLVEALYLVCGERGFFYTDK

SEQ ID NO:145
FVNQHLCGSDLVEALYLVCGERGFFYTDKGIVEQCCTSICSLYQLENYCN

SEQ ID NO:146
AAK

SEQ ID NO:147
NKR

SEQ ID NO:148
RRKQKR

SEQ ID NO:149
RRKR

SEQ ID NO:150
KDEL

SEQ ID NO:151
HDEL

SEQ ID NO:152
DDEL

SEQ ID NO:153
ADEL

SEQ ID NO:154
SDEL

SEQ ID NO:155
HDEF

SEQ ID NO:156

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys
115

SEQ ID NO:157

Met Ala Asp Thr Ala Arg Thr His His Asp Val Thr Ser Arg Asp Gln
1 5 10 15

Tyr Pro Arg Asp Arg Asp Gln Tyr Ser Met Ile Gly Arg Asp Arg Asp
20 25 30

Gln Tyr Ser Met Met Gly Arg Asp Arg Asp Gln Tyr Asn Met Tyr Gly
 35 40 45
 Arg Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Val Thr Ala Val
 50 55 60
 Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu Val Gly
 65 70 75 80
 Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile Phe Ser
 85 90 95
 Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile Thr Gly
 100 105 110
 Phe Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala Ile Thr Val Phe Ser
 115 120 125
 Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
 130 135 140
 Leu Asp Ser Ala Arg Met Lys Leu Gly Thr Lys Ala Gln Asp Ile Lys
 145 150 155 160
 Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
 165 170 175
 Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
 180 185

SEQ ID NO:158

taccatgggg tcaaagacgg agatgatgga gagagacgca atggctacgg tggctcccta 60
 tgogccggto aattaccato gccgtgctcg tgttgacttg gatgatagac ttcctaaacc 120
 ttatatgcca agagcattgc aagcaccaga cagagaacac ccgtacggaa ctocaggcca 180
 taagaattac ggaacttagtg ttottcaaca gcatgtotcc ttottcgata togatgataa 240
 tggcatcatt tacccttggg agacotactc tggactgcca atgcttggtt tcaatatcat 300
 tgggtcgctt ataatagcgc ctgttatcaa cctgaacctt agotatgcca ctctccggg 360
 gtggttacot tcacctttct tccctatata catacacaac atacacaagt caaagcatgg 420
 aagtgattca aaaacatatg acaatgaagg aaggtttatg ccggtgaatc ttgagttgat 480
 atttagcaaa tatgcgaaaa ccttgccaga caagttgagt cttggagAAC tatgggagat 540
 gacagaagga aaccgtgacg cttgggacat ttttgatgg atgcaggca aaatagagtg 600
 gggactgttg tacttgctag caagggatga agaagggttt ttgtcaaaag aagctattag 660
 gcggtgtttc gatggaagct tgttcgagta ctgtgccaaa atctacgctg gtatcagtga 720
 agacaagaca gcatactacg ccatggat 748

SEQ ID NO:159

atgggggtcaa agacggagat gatggagaga gacgcaatgg ctacgggtggc tccctatgcg 60
 ccggtcaactt accaccgccg tgctcgtgtt gacttgcatg atagacttcc taaaccttat 120

atgccaagag cattgcaagc accagacaga gaacacccgt acggaactoc aggcataag	180
aattacggac ttagtgtttc tcaacagcat gtctccttct tcgatatcga tgataatggc	240
atcattttacc cttgggagac ctactctgga ctgcgaatgc ttggtttcaa tatcattggg	300
tcgottataa tagccgctgt tatcaacctg acccttagct atgccactot tcgggggtgg	360
ttacottcac ctttcttccc tatatacata cacaacatac acaagtcaa gcattggaagt	420
gattcaaaaa catatgacaa tgaaggaagg tttatgocgg tgaatcttga gttgatattt	480
agcaaatatg cgaataacct gccagacaag ttgagtcctg gagaactatg ggagatgaca	540
gaaggaaacc gtgacgcttg ggacattttt ggatggatcg caggcaaat agagtgggga	600
ctgttgtact tgctagcaag ggatgaagaa gggtttttgt caaagaagc tattaggcgg	660
tgtttcgatg gaagcttggt cgagtaotgt gccaaaatct acgctggtat cagtgaagac	720
aagacagcat actactaa	738

SEQ ID NO:160

ATGGATCTAATCCACACTTTCTCAACTTAATAGCTCCCCCTTTCACCTTCTTCTTCTTCTC
TTTTTCTTGCCACCCTTCCAGATTTTCAAGTTCTTCTTCTTCAATCTTGGGCACCCTTTTCAGC
GAGGATGTCGCTGGAAAAGTCGTCGTCATCACCGGCGCCTCCTCCGGCATCGGCGAAAGTCTT
GCTTACGAGTATGCTAAGAGAGGGGCGTGCTTGGTGCTTGCTGCAAGAAGGGAAAGGAGTCTT
CAAGAAGTGGCCGAAAGGGCGCGGATTTGGGGTCCCGGACGTCGTGGTGGTCCGGGCGGAT
GTTTCGAAGGCGGAGGACTGCAGGAAGGTTGTTGATCAGACTATGAATCGCTTTGGAAGATTG
GATCACCTGGTCAATAACGCTGGAATTATGTGAGTTTCAATGCTGGAAGAAGTTGAAGATATT
ACTGGTTACAGAGAACTATGGATATCAACTTCTGGGGCTATGTGTATATGACCCGATTTGCC
GCCCCATACCTTAGGAATAGCAGAGGCCGAATTGTTGTACTTTCTTCATCCAGTTCTTGGATG
CCTACTCCGAGGATGAGTTTTTACAATGCAAGCAAAGCGGCGATTTACAATTTTTTGGAGACA
CTGCGGGTGGAATTCGGCCCCGATATAGGCATAACCTTGTGACTCCAGGATTCATAGAATCT
GAACTTACCAAGGCAAATCTACAATGCTGGCGAACGTGTAATTGATCAGGACATGAGAGAT
GTACAAGTGAGCACGACTCCAATCCTGAGGGTGGAAGTGCGGCAAGGTCAATCGTGAGGAGC
GCGATCCGTGGAGAAAGATACGTGACAGAGCCGGCCTGGTTTAGGGTTACTTATTGGTGGAAG
CTATTCTGCCCTGAGGTGATGGAGTGGGTATTTAGACTGATGTACTTGGCCAGCCCGGGTGAG
CCGGAGAAGGAAACGTTTGGCAAGAAGTTTTGGATTACACAGGAGTGAAGTCCTTGCTTTAC
CCGGAACCGTGCAAGTTCCGGAGCCCAAGAATGATTAA

SEQ ID NO:161

MNFLKSFPFYAFLCFGQYFVAVTHA

SEQ ID NO:162

APVNTTEDETAQAEAVIGYS DLEGDFDVAVL PFSNSTNNGLLFIBTTIASIAAKEEGVSLMAK
R

SEQ ID NO:163

APVNTTEDETAQAEAVIGYS DLEGDFDVAVL PFSNSTNNGLLFIBTTIASIAAKEEGVSMMAK

SEQ ID NO:164

QPIDEDNDTSSMAKR

SEQ ID NO:165

QPIDDTESNTTSVNLMADDTEDRFATNTTLALDVVNLI SMAKR

SEQ ID NO:166

QPIDDTESQTTTSVNLMADDTEDRFATQTTALDVVNLI SMAKR

SEQ ID NO:167
QPIDDTESQTTSVNLMADDTEDRFATQTTLALDVVNLSMAAA

SEQ ID NO:168
QPIDDTESNTTSVNLMADDTEDRFATNTTLALDVVNLSMAAA

SEQ ID NO:169
QPIDDTESNTTSVNLMADDTEDRFATNTTLAGGLDVVNLSMAKR

SEQ ID NO:170
QPIDDTESQTTSVNLMADDTESAFATQTNSGGLDVVGLISMAKR

SEQ ID NO:171
QPIDDTESQTTSVNLMADDTESAFATQTNSGGLDVVGLISMAAA

SEQ ID NO:172
EEAEAEAEPEK

SEQ ID NO:173
EEGEPEK

SEQ ID NO:174
MAEITRIPLYKGKSLRKALKEHGLLEDFLQKQYGISSKF

SEQ ID NO:175
GCATGCTGACATTGTGATGACACAGTC

SEQ ID NO:176
AAGCTTGCATTTAAATACTCGAGACTGTGAGAGTGGTGCCTTG

SEQ ID NO:177
GAAGAAGGAGAGCCTAAGTTTGTTAATCAACATCTTTGTGGATCTCATCTTGTTGAGGCTCTC
TACCTTG

SEQ ID NO:178
CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA

SEQ ID NO:179
CTAAGGCTGCTAAGGGAATTG

SEQ ID NO:180
AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACATTGTTCA
ACAATTCCCTTAGCAGCCTT

SEQ ID NO:181
CTCGAGTCAACCAATTGATGACACTGAATC

SEQ ID NO:182
AAGCTTCAAAGTTCATCCTTGTTGCAATAGTTCTCCAATTG

SEQ ID NO:183
AAGCTTCAGTTGCAATAGTTC

SEQ ID NO:184
GCATGCCCAACCAATTGATGACACTG

SEQ ID NO:185
GCATGCATGCCTTTGTTAATCAACATCTTTGTGG

SEQ ID NO:186
ACATTGTTCAACAATTCTCTCTTTCTTCTAGTCTTAGGAGTGTAGAAAAATCC

SEQ ID NO:187
GCATAAGCTTCAAAGCTCATCCTTTGAGC

SEQ ID NO: 188
ATGAACCTTCCTTAAGTCTTTCCCTTTCTACGCTTTCTTTGTTTCGGTCAATACTTCGTTGCT
GTTACGCATGCCCAACCAATTGATGACACTGAATCCCAGACCACGTCACTGAACCTCATGGCC
GATGATACTGAGAGCGCGTTTGCTACACAAACAAATTCGGGAGGTCTTGACGTTGTCGGATTG
ATCTCCATGGCTAAGAGAGAAGAAGGAGAGCCTAAGTTTGTAAATCAACATCTTTGTGGATCT
CATCTTGTGAGGCTCTCTACCTTGTGTGTGGAGAAAGAGGATTTTTCTACACTCCTAAGGCT
GCTAAGGGAATTGTTGAACAATGTTGCACTTCTATTTGCTCACTTTACCAATTGGAGAACTAT
TGCAACTGA

SEQ ID NO: 189
MNFLKSFPFYAFLCFGQYFVAVTHAQPIDDTESQTTSVNLMADDTESAFATQTNSSGGLDVVGL
ISMAREEGEPKFVNQHLGSHLVEALYLVCGERGFFYTPKAAGKIVEQCCTSIICSLYQLENY
CN

SEQ ID NO: 190
TTCGTGAACCAACACTTG

SEQ ID NO: 191
AAGCTTTCAGTTACAGTAGT

SEQ ID NO: 192
GCATGCATGTGTTGAGC

SEQ ID NO:193
GGTAGTGTGCTGGCCA

SEQ ID NO: 194
GGTGGCCAGCACACTACCTTCGTGAACCAACACTTGTG

SEQ ID NO:195
ATGGCGGATACAGCTAGAGGAACCCATCACGATATCATCGGCAGAGACCAGTACCCGATGATG
GGCCGAGACCGAGACCAGTACCAGATGTCCGGACGAGGATCTGACTACTCCAAGTCTAGGCAG
ATTGCTAAAGCTGCAACTGCTGTCACAGCTGGTGGTTCCCTCCTTGTTCTCTCCAGCCTTACC
CTTGTTGGAAGTGTCTAGCTTTGACTGTTGCAACACCTCTGCTCGTTATCTTCAGCCCAATC
CTTGTTCCCGGCTCTCATCACAGTTGCACTCCTCATCACCGGTTTTCTTTCTCTGAGGGTTT
GGCATTGCCGCTATAACCGTTTTCTCTTGATTTACAAGTAAGCACACATTTATCATCTTACT
TCATAATTTTGTGCAATATGTGCATGCATGTGTTGAGCCAGTAGCTTTGGATCAATTTTTTTG
GTAGAATAACAAATGTAACAATAAGAAATTGCAAATTTCTAGGGAACATTTGGTTAACTAAATA
CGAAATTTGACCTAGCTAGCTTGAATGTGTCTGTGTATATCATCTATATAGGTAAATGCTTG
GTATGATACCTATTGATTGTGAATAGGTACGCAACGGGAGAGCACCCACAGGGATCAGACAAG
TTGGACAGTGCAAGGATGAAGTTGGGAAGCAAAGCTCAGGATCTGAAAGACAGAGCTCAGTAC
TACGGACAGCAACATACTGGTGGGGAACATGACCGTGACCGTACTCGTGGTGGCCAGCACACT
ACCTTCGTGAACCAACACTTGTGTGGATCTCATCTCGTTGAAGCTCTCTACTTGGTTTGTGGT
GAGAGAGGATTCTTCTACACTCCTAAGACCAGAAGGGAAGCTGAGGACTTGCAGGTGGGACAA
GTTGAGTTGGGTGGAGGTCCTGGAGCAGGATCTTTGCAACCTCTCGCTTTGGAAGGTTCTTTG

CAGAAGAGAGGAATCGTTGAACAATGTTGCACTTCAATCTGTTCTTTGTATCAGTTGGAGAAC
TACTGTAACCTGA

SEQ ID NO: 196

MADTARGTHDIIIGRDQYPMGRDRDQYQMSGRGSDYSKSRQIAKAATAVATAGGSLLVLSSLT
LVGTVIALTVATPLLIVIFSPILVPALITVALLITGFLSSGGFGIAAITVFSWIYATGEHPQGS
DKLDSARMKLGSKAQDLKDRAQYYGQOHTGGEHNRDRTRGGQHTTFVNQHLGSHLVEALYLV
CGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSI CSLYQL
ENYCN

1	atgaacttccttaagtctttccctttctacgctttcctttgtttcggtcaatacttcgtt	60
1	M N F L K S F P F Y A F L C F G Q Y F V	20
61	gctgttacgcatgctgacattgtgatgacacagctctccatcctcctggctatgtcagtg	120
21	A V T H A D I V M T Q S P S S L A M S V	40
121	ggacagcgggtcactatgcgctgcaagtcagtcagagccttttaaaagtaccaatcaa	180
41	G Q R V T M R C K S S Q S L L K S T N Q	60
181	aagaactatttggcctggtaccagcagaaaccaggacagctctcctaaacttctggtatac	240
61	K N Y L A W Y Q Q K P G Q S P K L L V Y	80
241	tttgcattccactaggggaatctggggtccctgatogcttcattagggcagtggtctgggaca	300
81	F A S T R E S G V P D R F I G S G S G T	100
301	gatttcactcttaccatcagcagtggtgcaggctgaagacctggcagattacttctgtcag	360
101	D F T L T I S S V Q A E D L A D Y F C Q	120
361	caacattataaacactcctccacgcttcggtgctgggaccaagctggagcttaagcggctct	420
121	Q H Y N T P P T F G A G T K L E L K R S	140
421	ccgaacgggtgcttctcatagcgggttctgcaccaggcactagctctgcattctggatctcag	480
141	P N G A S H S G S A P G T S S A S G S Q	160
481	gtgcacctgcagcagctctggagctgagctgatgaagcctggggcctcaatgaagatatcc	540
161	V H L Q Q S G A E L M K P G A S M K I S	180
541	tgcaaggctactggctacacattcagtagctactggatagagtggttaagcagagccct	600
181	C K A T G Y T F S S Y W I E W V K Q R P	200
601	ggacatggccttgagtggattggagagattttacctggcagtggttagtactacctacaat	660
201	G H G L E W I G E I L P G S G S T T Y N	220
661	gagaagttcaagggcaaggccacattcactgcagatacatcctccaacacagcctacatg	720
221	E K F K G K A T F T A D T S S N T A Y M	240
721	caactcagcagcctgacatctgaggactctgccgtctattactgtgcaagattggatgtt	780
241	Q L S S L T S E D S A V Y Y C A R L D V	260
781	gactcctggggccaaggcaccactctcacagctctcgagtcaccaattgatgacactgaa	840
261	D S W G Q G T T L T V S S <u>Q P I D D T E</u>	280
841	tcccagaccacgtcagtgaaacctcatggccgatgatactgagagcgcgtttgctacacaa	900
281	<u>S Q T T S V N L M A D D T E S A F A T Q</u>	300
901	acaaattcgggaggtcttgacgttgctggattgatctccatggctaagagagaagaagga	960
301	<u>T N S G G L D V V G L I S M A K R E E G</u>	320
961	gagcctaagtttgtaataacacatctttgtggatctcatcttggttgaggctctctacett	1020
321	<u>E P K F V N Q H L C G S H L V E A L Y L</u>	340

1021 gtgtgtggagaaagaggatttttctacactcctaaggctgctaaggggaattgttgaacaa 1080
 341 V C G E R G F F Y T P K A A K G I V E Q 360

1081 tgttgcacttctatattgctcactttaccaattggagaactattgcaacaaggatgaactt 1040
 361 C C T S I C S L Y Q L E N Y C N X D E L 380

1041 tga

Fig. 1-2

1 atggcgggatacagctagaggaacccatcacgatatcatcggcagagaccagtaccogatg 60
 1 M A D T A R G T H H D I I G R D Q Y P M 20

61 atgggcccagagaccgagaccagtaccagatgtccggacgaggatctgactactccaagtct 120
 21 M G R D R D Q Y Q M S G R G S D Y S K S 40

121 aggcagattgctaaagctgcaactgctgtcacagctgggtgggtccctccttgttctctcc 180
 41 R Q I A K A A T A V T A G G S L L V L S 60

181 agccttacccttgttggaaactgtcatagctttgactggttgaacacctctgctcgttatc 240
 61 S L T L V G T V I A L T V A T P L L V I 80

241 ttcagcccaatccttgtcccggtctctcatcacagttgcactcctcatcacoggttttctt 300
 81 F S P I L V P A L I T V A L L I T G F L 100

301 tctctggagggttttggcattgccgctataacgcgttttctcttggatttacgcaacggga 360
 101 S S G G F G I A A I T V F S W I Y A T G 120

361 gagcaccacagggatcagacaagttggacagtgaaggatgaagttgggaagcaaagct 420
 121 E H P Q G S D K L D S A R M K L G S K A 140

421 caggatctgaaagacagagctcagtactacggacagcaacatactggtggggaacatgac 480
 141 Q D L K D R A Q Y Y G Q Q H T G G E H D 160

481 cgtgaccgtactcgtggtggccagcacactaccatgggtgagatcacccgcattcctctc 540
 161 R D R T R G G Q H T T M A E I T R I P L 180

541 tacaaggttaagtctctccgtaaggcgctgaaggaaacatggacttctagaagacttcttg 600
 181 Y K G K S L R K A L K E H G L L E D F L 200

601 cagaaacaacagtatggcatctcgagcaagttccaaccaattgatgacactgaatcccag 660
 201 Q K Q Q Y G I S S K F Q P I D D T E S Q 220

661 accacgtcagtgaacctcatggccgatgatactgagagcgogtttgctacacaaacaaat 720
 221 T T S V N L M A D D T E S A F A T Q T N 240

721 tcgggaggtctttgacgttgtcggattgatctccatggctaagagagaagaaggagagcct 780
 241 S G G L D V V G L I S M A K R E E G E P 260

781 aagtttgttaatcaacatctttgtggatctcatcttgttgaggctctctaccttgtgtgt 840
 261 K F V N Q H L C G S H L V E A L Y L V C 280

841 ggagaaagaggatttttctacactcctaaggctgctaaggggaattgttgaacaatgttgc 900
 281 G E R G F F Y T P K A A K G I V E Q C C 300

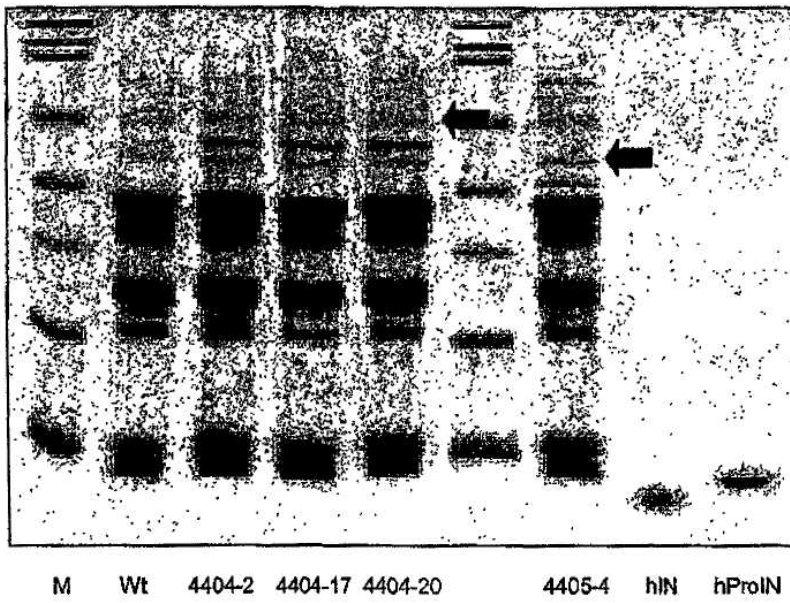
901 acttctatttgcactttaccaattggagaactattgcaactga
 301 T S I C S L Y Q L E N Y C N -

Fig. 2

1	atgaacttccttaagtctttccctttctacgctttcctttgtttcggtcaataacttcgtt	60
1	M N F L K S F P F Y A F L C F G Q Y F V	20
61	gctgttacgcatgcctttgttaatacaacatctttgtggatctcatcttgttgaggctctc	120
21	A V T H A F V N Q H L C G S H L V E A L	40
121	taccttgtgtgtggagaaaggaggtttttctacactcctaagactagaagaaagagagga	180
41	Y L V C G E R G F F Y T P K T R R K R G	60
181	attgttgaacaatgttgcaacttctatttgcctactttaccaattggagaactattgcaac	240
61	I V E Q C C T S I C S L Y Q L E N Y C N	80
241	agaagaaagagagacattgtgatgacacagctctccatcctccctggctatgtcagtgagg	300
81	<u>R R K R</u> D I V M T Q S P S S L A M S V G	100
301	cagogggtcactatgcgctgcaagtcacagtcagagccttttaaaaagtaccaatcaaaag	360
101	Q R V T M R C K S S Q S L L K S T N Q K	120
361	aactatttggcctgggtaccagcagaaaccaggacagctctcctaaacttctgggtatacttt	420
121	N Y L A N Y Q Q K P G Q S P K L L V Y F	140
421	gcattccactaggggaatctggggtccctgatcgcttcattaggcagtggtctgggacagat	480
141	A S T R E S G V P D R F I G S G S G T D	160
481	ttcactcttaccatcagcagtggtgcaggctgaagacctggcagattacttctgtcagcaa	540
161	F T L T I S S V Q A E D L A D Y F C Q Q	180
541	cattataacactcctcccacgttcgggtgctgggaccaagttggagcttaagcgggtctccg	600
181	H Y N T P P T F G A G T K L E L K R S P	200
601	aacgggtgcttctcatagcgggttctgcaccaggcactagctctgcattctggatctcaggtg	660
201	N G A S H S G S A P Q T S S A S G S Q V	220
661	cacctgcagcagctctggagctgagctgatgaagcctggggcctcaatgaagatatcctgc	720
221	H L Q Q S G A E L M K P G A S M K I S C	240
721	aaggctactggctacacattcagtagctactggatagagtggttaaagcagaggcctgga	780
241	K A T G Y T F S S Y W I E N V K Q R P G	260
781	catggccttgagtggattggagagattttacctggcagtggttagtaactacctacaatgag	840
261	H G L E W I G E I L P G S G S T T Y N E	280
841	aagttcaagggcaaggccacattcactgcagatacatcctccaacacagcctacatgcaa	900
281	K F K G K A T F T A D T S S N T A Y M Q	300
901	ctcagcagcctgacatctgaggactctgccgtctattactgtgcaagattggatgttgac	960
301	L S S L T S E D S A V Y Y C A R L D V D	320
961	tcctggggccaaggcaccactctcacagtgagctcaaaggatgagctttga	
321	S W G Q G T T L T V S S <u>K D E L</u> -	

Фиг. 3

A)



B)

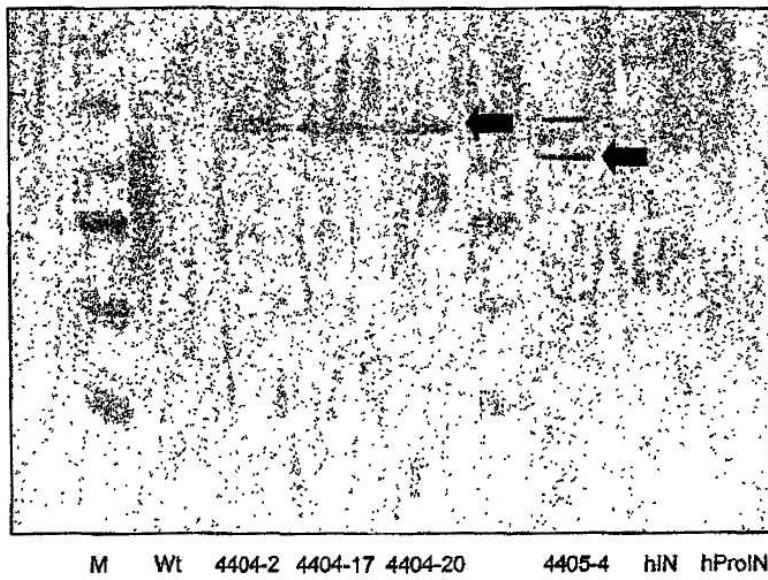
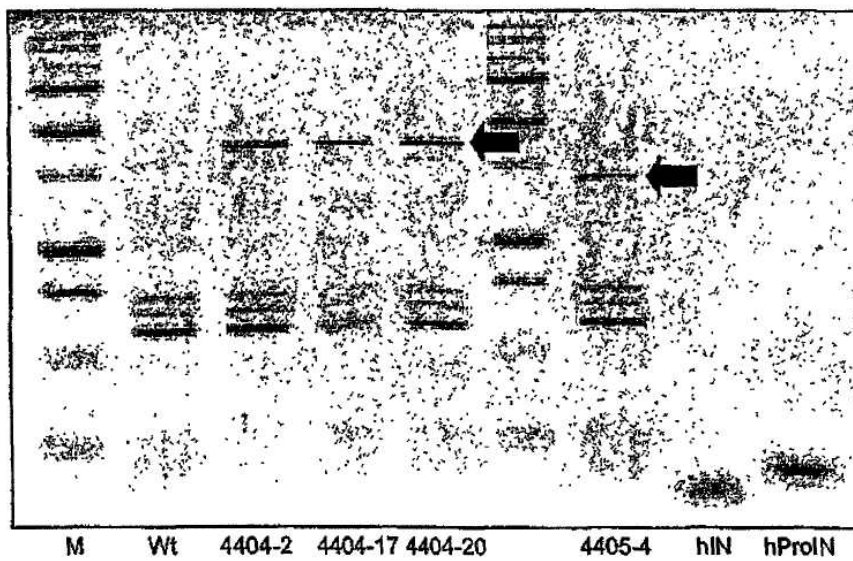


Fig. 4

C)



D)

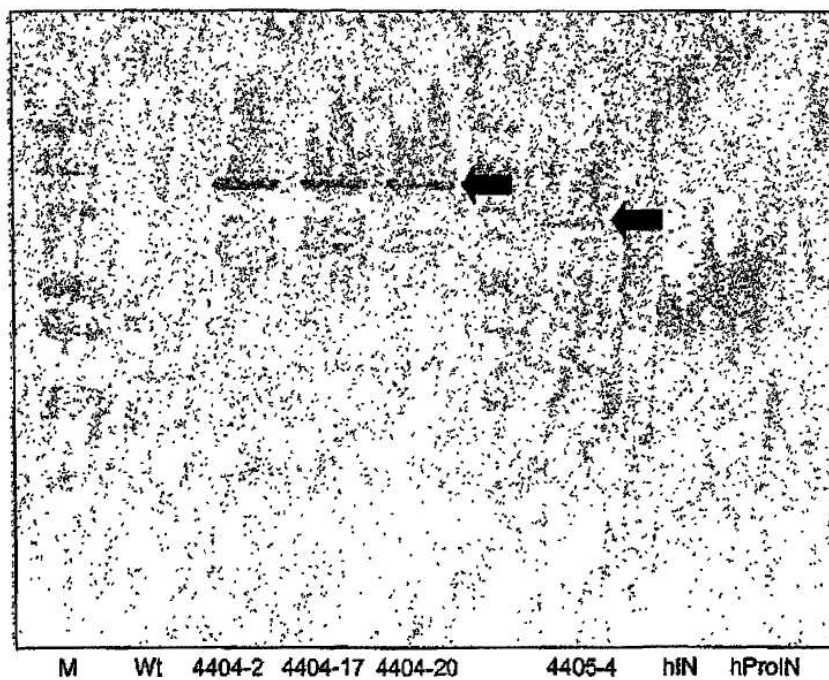
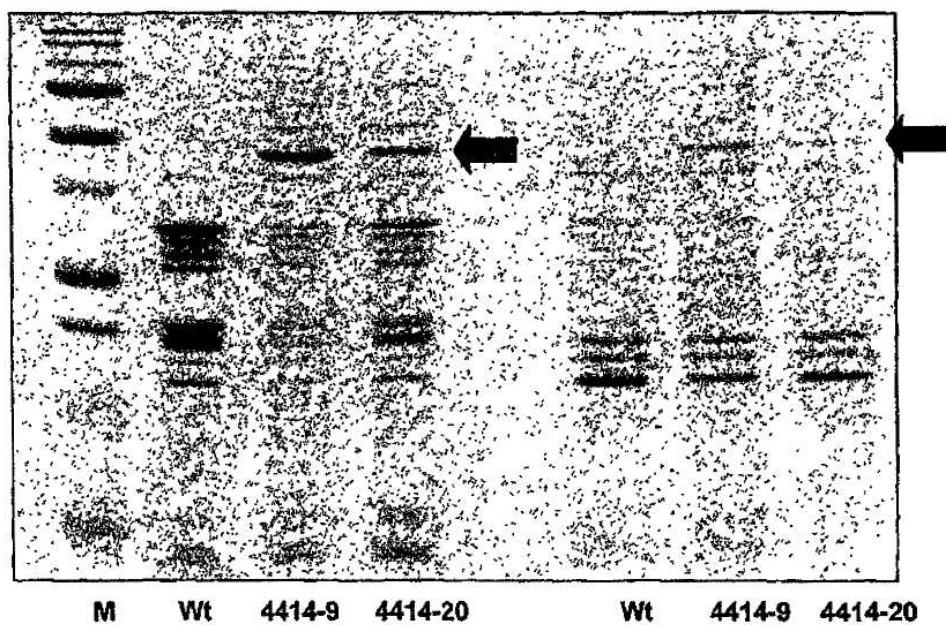


Fig. 4-2

E)



F)

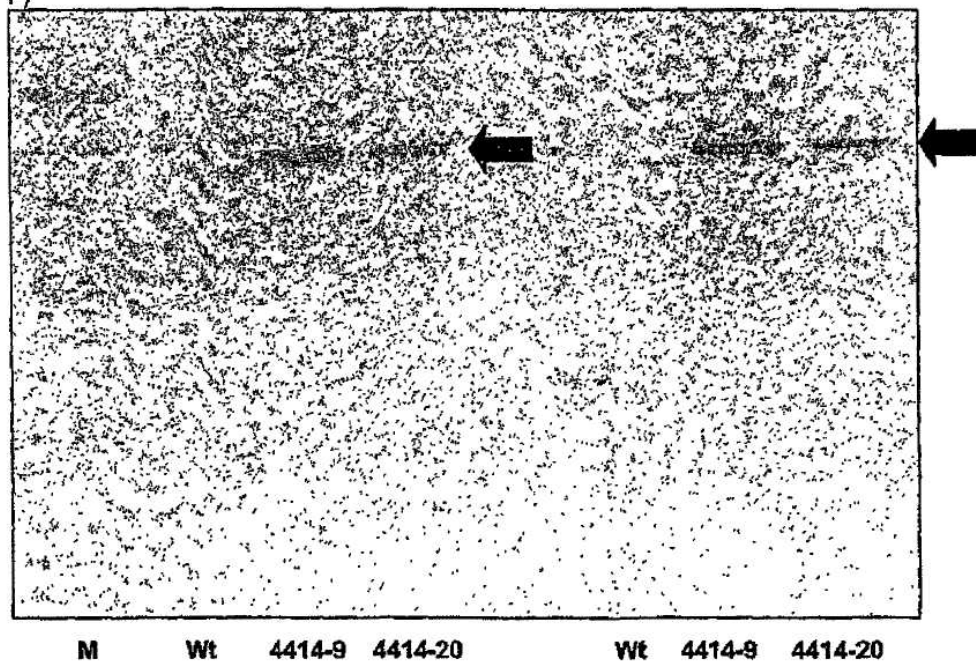
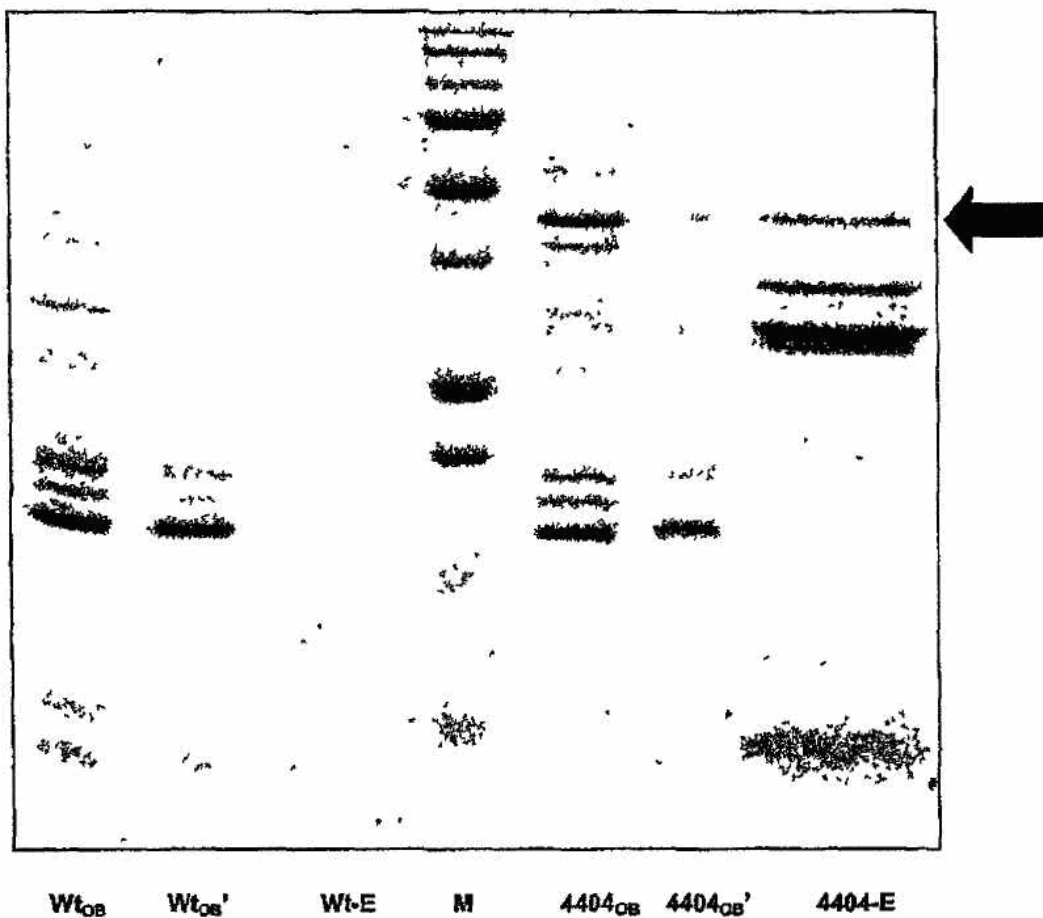


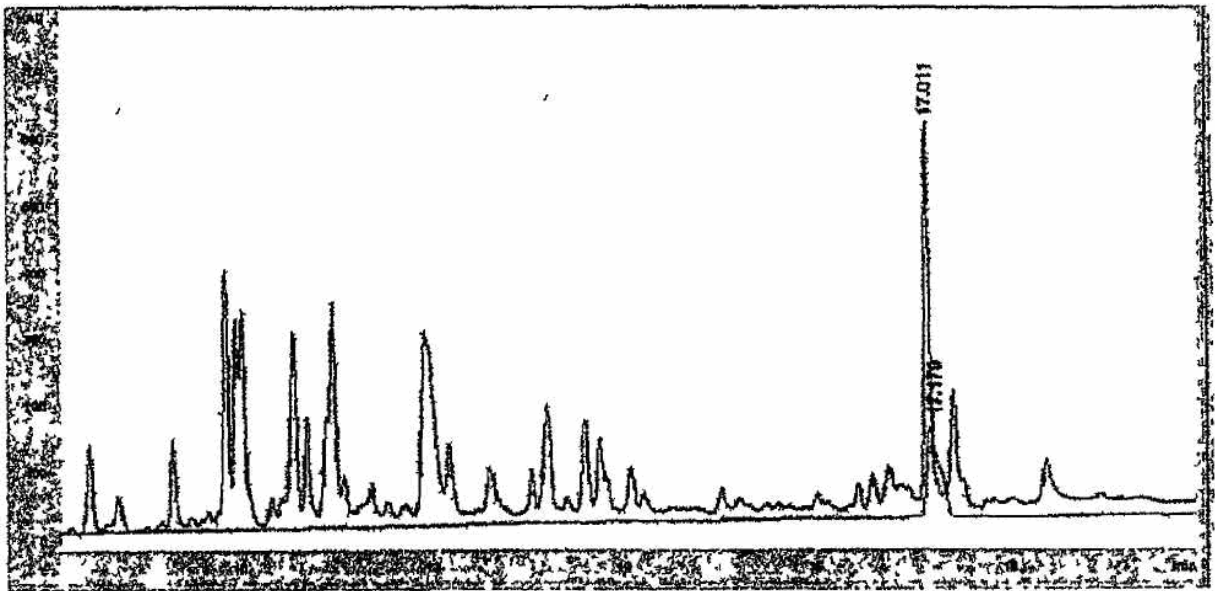
Fig. 4-3

Конструкція	Лінія	Трансген (% від загальної кількості білка насіння)	Міні-інсулін (% від загальної кількості білка насіння)
4404	2	1,20	0,20
4404	17	1,09	0,19
4404	20	1,39	0,24
4405	4	0,63	0,11
4405	13	0,62	0,11
4405	19	0,75	0,13
4414	9	6,78	1,15
4414	20	2,50	0,42
wt	C24	0	0

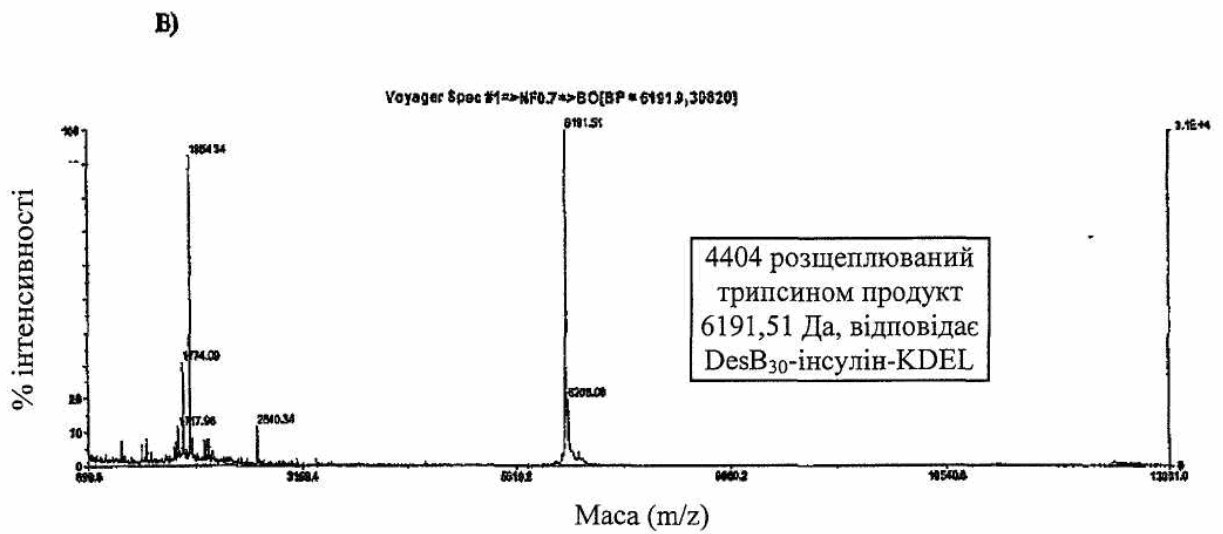
Фіг. 5



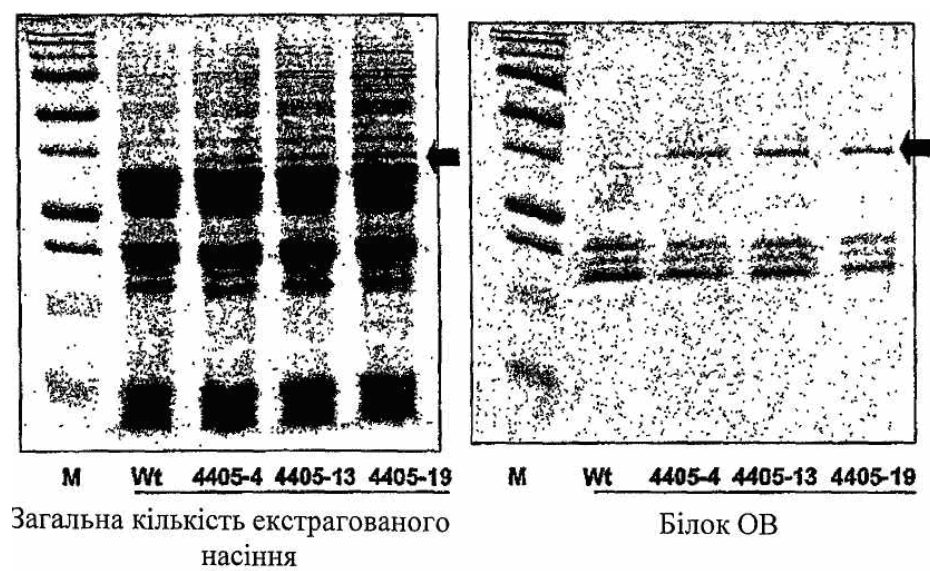
Фіг. 6



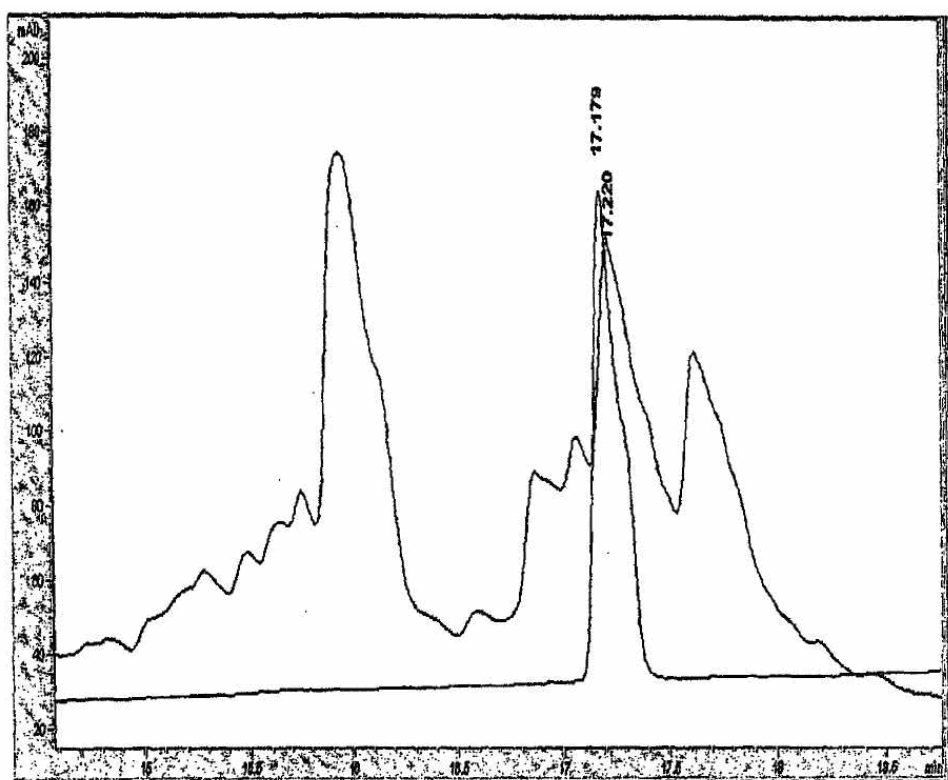
Фіг. 7



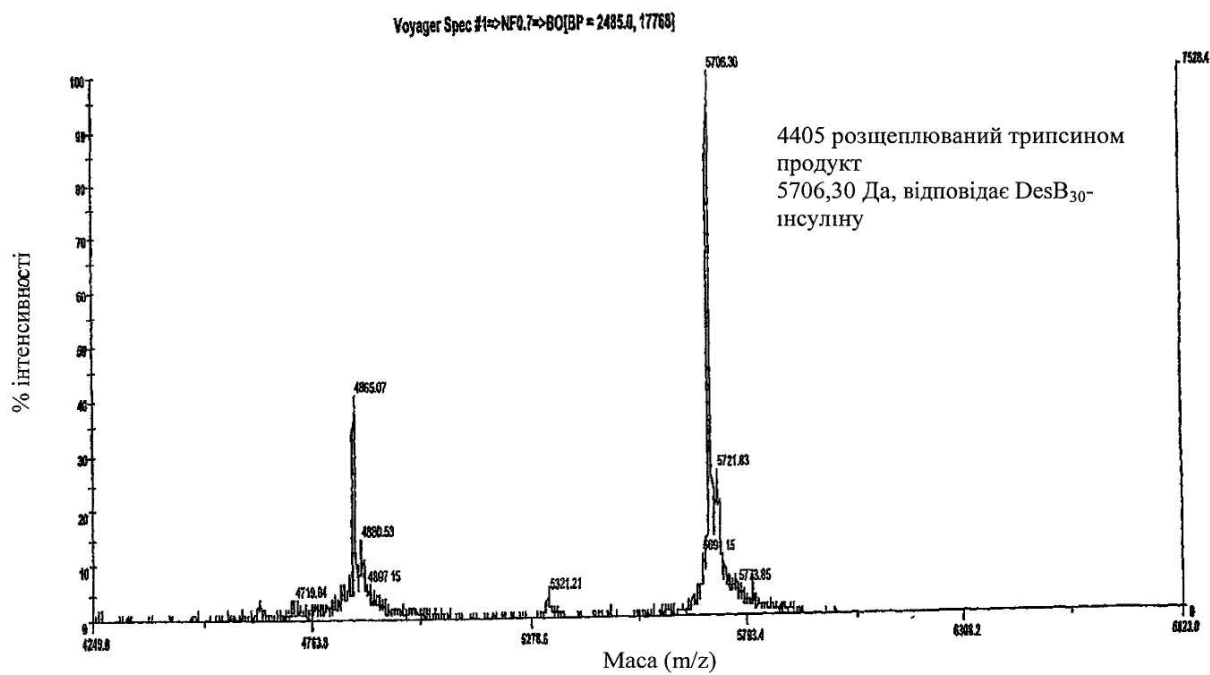
Фіг. 8



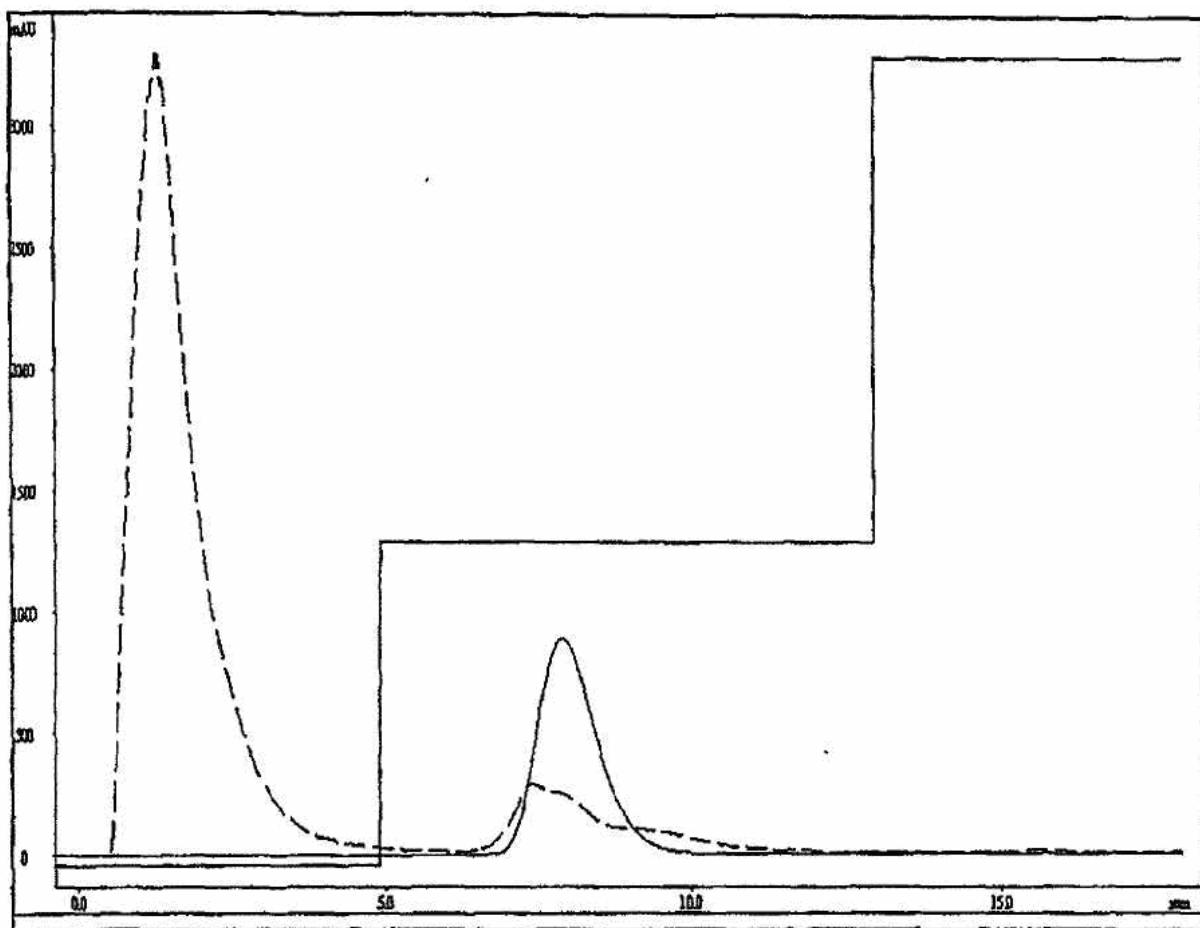
Фіг. 9



Фіг. 10

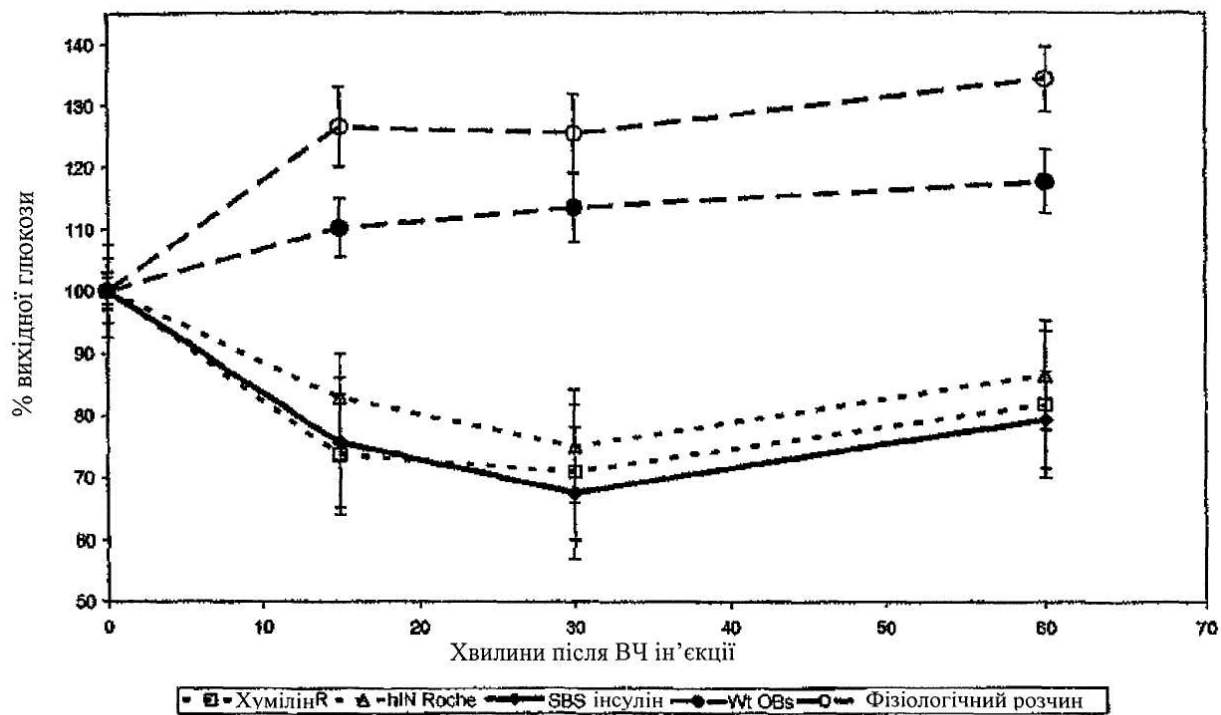


Фіг. 11

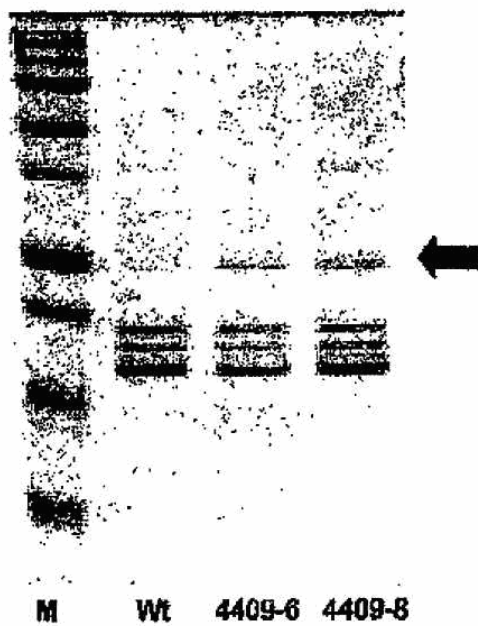


Фіг. 12

Тести на сприйнятливість до інсуліну на самцях мишей C57Bl/6



Фіг. 13



Фіг. 14

СПИСОК ПОСЛІДОВНОСТЕЙ

<110> SemBioSys Genetics Inc.

Moloney, Maurice M.

Boothe, Joseph

Keon, Richard

Nykiforuk, Cory

Van Rooijen, Gijs

<120> Спосіб одержання інсуліну в рослинах

<130> 9369-296

<150> 60/478,818

<151> 2003-06-17

<150> 60/549,539

<151> 2004-03-04

<160> 196

<170> PatentIn версія 3.1

<210> 1

<211> 1143

<212> днк

<213> Штучна послідовність

<220>

<223> Послідовність нуклеїнової кислоти злитого білка інсуліну

<400> 1
atgaacttcc ttaagtcttt ccctttctac gctttccttt gtttcgggtca atacttcggt 60
gctgttaacg atgctgacat tgtgatgaca cagtctccat cctccctggc tatgtcagtg 120
ggacagcggg tcaactatgcg ctgcaagtcc agtcagagcc ttttaaaaag taccaatcaa 180
aagaactatt tggcctggta ccagcagaaa ccaggacagt ctctaaaact tctggtatac 240
tttgcatcca ctagggaatc tggggtcctt gatcgcttca taggcagtggt atctgggaca 300
gatttcactc ttaccatcag cagtgtgcag gctgaagacc tggcagatta cttctgtcag 360
caacattata acactcctcc caggttcggt gctgggacca agctggagct taagcgggtc 420
ccgaacgggtg cttctcatag cggttctgca ccaggcacta gctctgcac tggatctcag 480
gtgcacctgc agcagctctg agctgagctg atgaagcctg gggcctcaat gaagatatcc 540
tgcaaggcta ctggctacac attcagtagc tactggatag agtgggtaaa gcagaggcct 600
ggacatggcc ttgagtggat tggagagatt ttacctggca gtggtagtac tacctacaat 660
gagaagttca agggcaaggc cacattcact gcagatacat cctccaacac agcctacatg 720
caactcagca gcctgacatc tgaggactct gccgtctatt actgtgcaag attggatgtt 780
gactcctggg gccaaaggcac cactctcaca gtctcgagtc aaccaattga tgacactgaa 840
tcccagacca cgtcagtgaa cctcatggcc gatgatactg agagcgcgtt tgctacacaa 900
acaaattcgg gaggtcttga cgttgctcga ttgatctcca tggctaagag agaagaagga 960
gagcctaagt ttgttaatca acatctttgt ggatctcatc ttgttgaggc tctctacctt 1020
gtgtgtggag aaagaggatt tttctacact cctaaggctg ctaagggaat tgttgaacaa 1080
tgttgcactt ctatttgctc actttaacaa ttggagaact attgcaacaa ggatgaactt 1140
tga 1143

<210> 2

<211> 380

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<400> 2

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
 1 5 10 15
 Gln Tyr Phe Val Ala Val Thr His Ala Asp Ile Val Met Thr Gln Ser
 20 25 30
 Pro Ser Ser Leu Ala Met Ser Val Gly Gln Arg Val Thr Met Arg Cys
 35 40 45
 Lys Ser Ser Gln Ser Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu
 50 55 60
 Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Val Tyr
 65 70 75 80
 Phe Ala Ser Thr Arg Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser
 85 90 95
 Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Glu
 100 105 110
 Asp Leu Ala Asp Tyr Phe Cys Gln Gln His Tyr Asn Thr Pro Pro Thr
 115 120 125
 Phe Gly Ala Gly Thr Lys Leu Glu Leu Lys Arg Ser Pro Asn Gly Ala
 130 135 140
 Ser His Ser Gly Ser Ala Pro Gly Thr Ser Ser Ala Ser Gly Ser Gln
 145 150 155 160
 Val His Leu Gln Gln Ser Gly Ala Glu Leu Met Lys Pro Gly Ala Ser
 165 170 175
 Met Lys Ile Ser Cys Lys Ala Thr Gly Tyr Thr Phe Ser Ser Tyr Trp
 180 185 190
 Ile Glu Trp Val Lys Gln Arg Pro Gly His Gly Leu Glu Trp Ile Gly
 195 200 205
 Glu Ile Leu Pro Gly Ser Gly Ser Thr Thr Tyr Asn Glu Lys Phe Lys
 210 215 220

Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr Met
225 230 235 240

Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys Ala
245 250 255

Arg Leu Asp Val Asp Ser Trp Gly Gln Gly Thr Thr Leu Thr Val Ser
260 265 270

Ser Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu
275 280 285

Met Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly
290 295 300

Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys Arg Glu Glu Gly
305 310 315 320

Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu
325 330 335

Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys
340 345 350

Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu
355 360 365

Tyr Gln Leu Glu Asn Tyr Cys Asn Lys Asp Glu Leu
370 375 380

<210> 3

<211> 945

<212> ДНК

<213> Штучна послідовність

<220>

<223> Послідовність нуклеїнової кислоти злитого білка інсуліну

<400> 3

atggcggata cagctagagg aaccatcac gatatcatcg gcagagacca gtaccgatg 60

atgggcccag accgagacca gtaccagatg tccggacgag gatctgacta ctccaagtct 120

aggcagattg ctaaagctgc aactgctgtc acagctggtg gttccctcct tgttctctcc 180
 agccttaccc ttgttggaac tgtcatagct ttgactgttg caacacctct gctcgttato 240
 ttcagcccaa tccttgtccc ggctctcatc acagttgcac tcctcatcac cggttttctt 300
 tcctctggag ggtttgcat tgcgctata accgttttct ottggattta cgcaacggga 360
 gagcaccac agggatcaga caagttggac agtgcaagga tgaagttggg aagcaaagct 420
 caggatctga aagacagagc tcagtactac ggacagcaac atactggtgg ggaacatgac 480
 cgtgaccgta ctcggtgttg ccagcacact accatggctg agatcacccg cattcctctc 540
 taaaaggtta agtctctccg taaggcgtg aaggaacatg gacttctaga agacttcttg 600
 cagaaacaac agtatggcat ctcgagcaag ttccaacca ttgatgacac tgaatcccag 660
 accacgtcag tgaacctcat ggcgatgat actgagagcg cgtttgctac acaacaaat 720
 tcgggaggtc ttgacgttgt cggattgatc tccatggcta agagagaaga aggagagcct 780
 aagtttgta atcaacatct ttgtggatct catcttggtg aggtctctta ccttggtgtg 840
 ggagaaagag gatttttcta cactcctaag gctgctaagg gaattgttga acaatgttgc 900
 acttctattt gctcacttta ccaattggag aactattgca actga 945

<210> 4

<211> 314

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<400> 4

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
 50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
 65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
 85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
 100 105 110

Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
 115 120 125

Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys
 130 135 140

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
 145 150 155 160

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr Met Ala Glu Ile Thr
 165 170 175

Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu
 180 185 190

His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser
 195 200 205

Ser Lys Phe Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val
 210 215 220

Asn Leu Met Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn
 225 230 235 240

Ser Gly Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys Arg Glu
 245 250 255

Glu Gly Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
 260 265 270

Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
 275 280 285

Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
 290 295 300

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
 305 310

<210> 5

<211> 1011

<212> ДНК

<213> Штучна послідовність

<220>

<223> Послідовність нуклеїнової кислоти злитого білка інсуліну

<400> 5

atgaacttcc ttaagtcttt ccccttctac gctttccttt gtttcggtca atacttcggt	60
gctgttacgc atgcctttgt taatcaacat ctttgtggat ctcatcttgt tgaggctctc	120
taccttgtgt gtggagaaaag aggatttttc tacactccta agactagaag aaagagagga	180
attgttgaac aatgttgacac ttctatttgc tcactttacc aattggagaa ctattgcaac	240
agaagaaaaga gagacattgt gatgacacag tctccatcct cctgggtat gtcagtggga	300
cagcgggtca ctatgcgctg caagtccagt cagagccttt taaaaagtac caatcaaaag	360
aactatttgg cctggtacca gcagaaacca ggacagtctc ctaaaacttc ggtatacttt	420
gcatecacta gggaatctgg ggtccctgat cgcttcatag gcagtggatc tgggacagat	480
ttcactctta ccacagcag tgtgcaggct gaagacctgg cagattactt ctgtcagcaa	540
cattataaca ctccctccac gttcgggtgt gggaccaagt tggagcttaa gcgggtctccg	600
aacggtgctt ctcatagcgg ttctgcacca ggcactagct ctgcatctgg atctcaggtg	660
cacctgcagc agtctggagc tgagctgatg aagcctgggg cctcaatgaa gatatactgc	720
aaggctactg gctacacatt cagtagctac tggatagagt gggtaaagca gaggcctgga	780
catggccttg agtggatttg agagatttta cctggcagtg gtagtactac ctacaatgag	840
aagttcaagg gcaaggccac attcactgca gatacatcct ccaacacagc ctacatgcaa	900
ctcagcagcc tgacatctga ggactctgcc gtctattact gtgcaagatt ggatgttgac	960
tcttggggcc aaggcaccac tctcacagt agctcaaagg atgagctttg a	1011

<210> 6

<211> 336

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<400> 6

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala Phe Val Asn Gln His Leu Cys
20 25 30

Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly
35 40 45

Phe Phe Tyr Thr Pro Lys Thr Arg Arg Lys Arg Gly Ile Val Glu Gln
50 55 60

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
65 70 75 80

Arg Arg Lys Arg Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ala
85 90 95

Met Ser Val Gly Gln Arg Val Thr Met Arg Cys Lys Ser Ser Gln Ser
100 105 110

Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln
115 120 125

Lys Pro Gly Gln Ser Pro Lys Leu Leu Val Tyr Phe Ala Ser Thr Arg
130 135 140

Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser Gly Ser Gly Thr Asp
145 150 155 160

Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Glu Asp Leu Ala Asp Tyr
165 170 175

Phe Cys Gln Gln His Tyr Asn Thr Pro Pro Thr Phe Gly Ala Gly Thr
180 185 190

Lys Leu Glu Leu Lys Arg Ser Pro Asn Gly Ala Ser His Ser Gly Ser
195 200 205

Ala Pro Gly Thr Ser Ser Ala Ser Gly Ser Gln Val His Leu Gln Gln
210 215 220

Ser Gly Ala Glu Leu Met Lys Pro Gly Ala Ser Met Lys Ile Ser Cys
225 230 235 240

Lys Ala Thr Gly Tyr Thr Phe Ser Ser Tyr Trp Ile Glu Trp Val Lys
245 250 255

Gln Arg Pro Gly His Gly Leu Glu Trp Ile Gly Glu Ile Leu Pro Gly
260 265 270

Ser Gly Ser Thr Thr Tyr Asn Glu Lys Phe Lys Gly Lys Ala Thr Phe
275 280 285

Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr Met Gln Leu Ser Ser Leu
290 295 300

Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys Ala Arg Leu Asp Val Asp
305 310 315 320

Ser Trp Gly Gln Gly Thr Thr Leu Thr Val Ser Ser Lys Asp Glu Leu
325 330 335

<210> 7

<211> 110

<212> PRT

<213> Homo sapiens

<400> 7

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 8

<211> 31

<212> PRT

<213> Equus przewalskii

<400> 8

Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro
1 5 10 15

Gly Leu Gly Gly Leu Gln Pro Leu Ala Leu Ala Gly Pro Gln Gln
20 25 30

<210> 9

<211> 86

<212> PRT

<213> Equus caballus

<220>

<221> MISC_FEATURE

<222> (31) .. (32)

<223> X = будь-яка амінокислота

<220>

<221> MISC_FEATURE

<222> (64) .. (65)

<223> X = будь-яка амінокислота

<400> 9

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

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

Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro
35 40 45

Gly Leu Gly Gly Leu Gln Pro Leu Ala Leu Ala Gly Pro Gln Gln Xaa
50 55 60

Xaa Gly Ile Val Glu Gln Cys Cys Thr Gly Ile Cys Ser Leu Tyr Gln
65 70 75 80

Leu Glu Asn Tyr Cys Asn
85

<210> 10

<211> 110

<212> PRT

<213> *Oryctolagus cuniculus*

<400> 10

Met Ala Ser Leu Ala Ala Leu Leu Pro Leu Leu Ala Leu Leu Val Leu
1 5 10 15

Cys Arg Leu Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Glu Leu Gln Val Gly
50 55 60

Gln Ala Glu Leu Gly Gly Gly Pro Gly Ala Gly Gly Leu Gln Pro Ser
65 70 75 80

Ala Leu Glu Leu Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 11

<211> 21

<212> PRT

<213> Balaenoptera physalus

<400> 11

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 12

<211> 21

<212> PRT

<213> Balaenoptera borealis

<400> 12

Gly Ile Val Glu Gln Cys Cys Ala Ser Thr Cys Ser Leu Tyr Gln Leu
 1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 13

<211> 108

<212> PRT

<213> Sus scrofa

<400> 13

Met Ala Leu Trp Thr Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
 1 5 10 15

Trp Ala Pro Ala Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
 20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
 35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Ala Glu Asn Pro Gln Ala Gly
 50 55 60

Ala Val Glu Leu Gly Gly Gly Leu Gly Gly Leu Gln Ala Leu Ala Leu
 65 70 75 80

Glu Gly Pro Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser
 85 90 95

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
 100 105

<210> 14

<211> 21

<212> PRT

<213> Elephas maximus

<400> 14

Gly Ile Val Glu Gln Cys Cys Thr Gly Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 15

<211> 105

<212> PRT

<213> Bos taurus

<400> 15

Met Ala Leu Trp Thr Arg Leu Arg Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Pro Pro Pro Pro Ala Arg Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gln Val Gly
50 55 60

Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Gly Leu Glu Gly Pro
65 70 75 80

Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser
85 90 95

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 16

<211> 105

<212> PRT

<213> Ovis aries

<400> 16

Met Ala Leu Trp Thr Arg Leu Val Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Ala Pro Ala Pro Ala His Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gln Val Gly
50 55 60

Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Gly Leu Glu Gly Pro
65 70 75 80

Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Ala Gly Val Cys Ser
85 90 95

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 17

<211> 21

<212> PRT

<213> Camelus dromedaries

<400> 17

Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 18

<211> 110

<212> PRT

<213> Canis sp.

<400> 18

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Ala Pro Ala Pro Thr Arg Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Asp Leu Gln Val Arg
50 55 60

Asp Val Glu Leu Ala Gly Ala Pro Gly Glu Gly Gly Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 19

<211> 21

<212> PRT

<213> Hystrix cristata

<400> 19

Gly Ile Val Asp Gln Cys Cys Thr Gly Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 20

<211> 108

<212> PRT

<213> Aotus trivirgatus

<400> 20

Met Ala Leu Trp Met His Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Glu Pro Ala Pro Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Ala Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Ser Ile Thr Gly Ser Leu Pro Pro Leu
65 70 75 80

Glu Gly Pro Met Gln Lys Arg Gly Val Val Asp Gln Cys Cys Thr Ser
85 90 95

Ile Cys Ser Leu Tyr Gln Leu Gln Asn Tyr Cys Asn
100 105

<210> 21

<211> 110

<212> PRT

<213> Macaca fascicularis

<400> 21

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Pro Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 22

<211> 110

<212> PRT

<213> Cercopithecus aethiops

<400> 22

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Val Pro Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 23

<211> 110

<212> PRT

<213> Pan troglodytes

<400> 23

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Val Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ser Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 24

<211> 21

<212> PRT

<213> Ornithorhynchus anatinus

<400> 24

Gly Ile Val Glu Glu Cys Cys Lys Gly Val Cys Ser Met Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 25

<211> 65

<212> PRT

<213> Pongo pygmaeus

<400> 25

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
1 5 10 15

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln
20 25 30

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln
35 40 45

Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln
50 55 60

Cys
65

<210> 26

<211> 110

<212> PRT

<213> Gorilla gorilla

<400> 26

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
 50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
 65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
 85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
 100 105 110

<210> 27

<211> 51

<212> PRT

<213> Saimiri sciureus

<400> 27

Phe Val Asn Gln His Leu Cys Gly Pro His Leu Val Glu Ala Leu Tyr
 1 5 10 15

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

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

Tyr Cys Asn
 50

<210> 28

<211> 110

<212> PRT

<213> Cricetulus longicaudatus

<400> 28

Met Thr Leu Trp Met Arg Leu Leu Pro Leu Leu Thr Leu Leu Val Leu
 1 5 10 15

Trp Glu Pro Asn Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Gly Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Asp Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Gln Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 29

<211> 110

<212> PRT

<213> Rattus norvegicus

<400> 29

Met Ala Leu Trp Met Arg Phe Leu Pro Leu Leu Ala Leu Leu Val Leu
1 5 10 15

Trp Glu Pro Lys Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Asp Pro Gln Val Pro
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Glu Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 30

<211> 110

<212> PRT

<213> Rattus norvegicus

<400> 30

Met Ala Leu Trp Ile Arg Phe Leu Pro Leu Leu Ala Leu Leu Ile Leu
1 5 10 15

Trp Glu Pro Arg Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 31

<211> 21

<212> PRT

<213> Acomys cahirinus

<400> 31

Gly Ile Val Asp Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 32

<211> 108

<212> PRT

<213> Mus musculus

<400> 32

Met Ala Leu Leu Val His Phe Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Glu Pro Lys Pro Thr Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Asp Pro Gln Val Glu
50 55 60

Gln Leu Glu Leu Gly Gly Ser Pro Gly Asp Leu Gln Thr Leu Ala Leu
65 70 75 80

Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys Thr Ser
85 90 95

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 33

<211> 110

<212> PRT

<213> Mus musculus

<400> 33

Met Ala Leu Trp Met Arg Phe Leu Pro Leu Leu Ala Leu Leu Phe Leu
1 5 10 15

Trp Glu Ser His Pro Thr Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Gln Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 34

<211> 21

<212> PRT

<213> Chinchilla brevicaudata

<400> 34

Gly Ile Val Asp Gln Cys Cys Thr Ser Ile Cys Thr Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 35

<211> 110

<212> PRT

<213> *Cavia porcellus*

<400> 35

Met Ala Leu Trp Met His Leu Leu Thr Val Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asn Thr Gly Gln Ala Phe Val Ser Arg His Leu Cys Gly
20 25 30

Ser Asn Leu Val Glu Thr Leu Tyr Ser Val Cys Gln Asp Asp Gly Phe
35 40 45

Phe Tyr Ile Pro Lys Asp Arg Arg Glu Leu Glu Asp Pro Gln Val Glu
50 55 60

Gln Thr Glu Leu Gly Met Gly Leu Gly Ala Gly Gly Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Gly Thr Cys Thr Arg His Gln Leu Gln Ser Tyr Cys Asn
100 105 110

<210> 36

<211> 109

<212> PRT

<213> *Octodon degus*

<400> 36

Met Ala Pro Trp Met His Leu Leu Thr Val Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asn Ser Val Gln Ala Tyr Ser Ser Gln His Leu Cys Gly
20 25 30

Ser Asn Leu Val Glu Ala Leu Tyr Met Thr Cys Gly Arg Ser Gly Phe
35 40 45

Tyr Arg Pro His Asp Arg Arg Glu Leu Glu Asp Leu Gln Val Glu Gln
 50 55 60

Ala Glu Leu Gly Leu Glu Ala Gly Gly Leu Gln Pro Ser Ala Leu Glu
 65 70 75 80

Met Ile Leu Gln Lys Arg Gly Ile Val Asp Gln Cys Cys Asn Asn Ile
 85 90 95

Cys Thr Phe Asn Gln Leu Gln Asn Tyr Cys Asn Val Pro
 100 105

<210> 37

<211> 21

<212> PRT

<213> Didelphis virginiana

<400> 37

Gly Ile Val Glu Gln Cys Cys Asn Ser Ile Cys Ser Leu Tyr Gln Leu
 1 5 10 15

Glu Thr Tyr Cys Asn
 20

<210> 38

<211> 108

<212> PRT

<213> Rodentia sp.

<400> 38

Met Ala Leu Trp Ile Leu Leu Pro Leu Leu Ala Leu Leu Ile Leu Trp
 1 5 10 15

Gly Pro Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly Ser
 20 25 30

His Leu Val Glu Ala Leu Tyr Ile Leu Val Cys Gly Glu Arg Gly Phe
 35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Ala Gly Pro Gly Ala Gly Ser Glu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Ala Arg Ile Val Gln Gln Cys Thr Ser
85 90 95

Gly Ile Cys Ser Leu Tyr Gln Glu Asn Tyr Cys Asn
100 105

<210> 39

<211> 110

<212> PRT

<213> Psammomys obesus

<400> 39

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Phe Leu Ile Leu
1 5 10 15

Trp Glu Pro Ser Pro Ala His Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Phe Arg Arg Gly Val Asp Asp Pro Gln Met Pro
50 55 60

Gln Leu Glu Leu Gly Gly Ser Pro Gly Ala Gly Asp Leu Arg Ala Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Gly Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 40

<211> 110

<212> PRT

<213> *Spermophilus tridecemlineatus*

<400> 40

Met Ala Leu Trp Thr Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Leu Gly Pro Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Glu Gln Gln Gly Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Leu Pro Gln Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 41

<211> 29

<212> PRT

<213> *Cavia porcellus*

<400> 41

Glu Leu Glu Asp Pro Gln Val Glu Gln Thr Glu Leu Gly Met Gly Leu
1 5 10 15

Gly Ala Gly Gly Leu Gln Pro Leu Gln Gly Ala Leu Gln
20 25

<210> 42

<211> 107

<212> PRT

<213> Ballus gallus

<400> 42

Met Ala Leu Trp Ile Arg Ser Leu Pro Leu Leu Ala Leu Leu Val Phe
1 5 10 15

Ser Gly Pro Gly Thr Ser Tyr Ala Ala Ala Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Ser Pro Lys Ala Arg Arg Asp Val Glu Gln Pro Leu Val Ser
50 55 60

Ser Pro Leu Arg Gly Glu Ala Gly Val Leu Pro Phe Gln Gln Glu Glu
65 70 75 80

Tyr Glu Lys Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr
85 90 95

Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 43

<211> 81

<212> PRT

<213> Anas platyrhynchos

<220>

<221> MISC_FEATURE

<222> (31)..(32)

<223> X = будь-яка амінокислота

<220>

<221> MISC_FEATURE

<222> (59) .. (60)

<223> X = будь-яка амінокислота

<400> 43

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

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

Asp Val Glu Gln Pro Leu Val Asn Gly Pro Leu His Gly Glu Val Gly
35 40 45

Glu Leu Pro Phe Gln His Glu Glu Tyr Gln Xaa Xaa Gly Ile Val Glu
50 55 60

Gln Cys Cys Glu Asn Pro Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys
65 70 75 80

Asn

<210> 44

<211> 21

<212> PRT

<213> Anser anser

<400> 44

Gly Ile Val Glu Gln Cys Cys Glu Asn Pro Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 45

<211> 103

<212> PRT

<213> *Selasphorus rufus*

<400> 45

Ile Gln Ser Leu Pro Leu Leu Ala Leu Leu Ala Leu Ser Gly Pro Gly
1 5 10 15

Thr Ser His Ala Ala Val Asn Gln His Leu Cys Gly Ser His Leu Val
20 25 30

Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Ser Pro
35 40 45

Lys Ala Arg Arg Asp Ala Glu His Pro Leu Val Asn Gly Pro Leu His
50 55 60

Gly Glu Val Gly Asp Leu Pro Phe Gln Gln Glu Glu Phe Glu Lys Val
65 70 75 80

Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr
85 90 95

Gln Leu Glu Asn Tyr Cys Asn
100

<210> 46

<211> 108

<212> PRT

<213> *Danio rerio*

<400> 46

Met Ala Val Trp Leu Gln Ala Gly Ala Leu Leu Val Leu Leu Val Val
1 5 10 15

Ser Ser Val Ser Thr Asn Pro Gly Thr Pro Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe
35 40 45

Tyr Asn Pro Lys Arg Asp Val Glu Pro Leu Leu Gly Phe Leu Pro Pro
50 55 60

Lys Ser Ala Gln Glu Thr Glu Val Ala Asp Phe Ala Phe Lys Asp His
65 70 75 80

Ala Glu Leu Ile Arg Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys
85 90 95

Pro Cys Ser Ile Phe Glu Leu Gln Asn Tyr Cys Asn
100 105

<210> 47

<211> 108

<212> PRT

<213> Cyprinus carpio

<400> 47

Met Ala Val Trp Ile Gln Ala Gly Ala Leu Leu Phe Leu Leu Ala Val
1 5 10 15

Ser Ser Val Asn Ala Asn Ala Gly Ala Pro Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe
35 40 45

Tyr Asn Pro Lys Arg Asp Val Asp Pro Pro Leu Gly Phe Leu Pro Pro
50 55 60

Lys Ser Ala Gln Glu Thr Glu Val Ala Asp Phe Ala Phe Lys Asp His
65 70 75 80

Ala Glu Val Ile Arg Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys
85 90 95

Pro Cys Ser Ile Phe Glu Leu Gln Asn Tyr Cys Asn
100 105

<210> 48

<211> 21

<212> PRT

<213> Batrachoididae gen. sp.

<400> 48

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asp Ile Phe Asp Leu
1 5 10 15

Gln Ser Tyr Cys Asn
20

<210> 49

<211> 21

<212> PRT

<213> Thunnus thynnus

<400> 49

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 50

<211> 21

<212> PRT

<213> Katsuwonus pelamis

<400> 50

Gly Ile His Glx Glx Cys Cys His Lys Pro Cys Asx Ile Phe Glx Leu
 1 5 10 15

Glx Asx Tyr Cys Asn
 20

<210> 51

<211> 116

<212> PRT

<213> *Lophius piscatorius*

<400> 51

Met Ala Ala Leu Trp Leu Gln Ser Phe Ser Leu Leu Val Leu Leu Val
 1 5 10 15

Val Ser Trp Pro Gly Ser Gln Ala Val Ala Pro Ala Gln His Leu Cys
 20 25 30

Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
 35 40 45

Phe Phe Tyr Asn Pro Lys Arg Asp Val Asp Gln Leu Leu Gly Phe Leu
 50 55 60

Pro Pro Lys Ser Gly Gly Ala Ala Ala Ala Gly Ala Asp Asn Glu Val
 65 70 75 80

Ala Glu Phe Ala Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly
 85 90 95

Ile Val Glu Gln Cys Cys His Arg Pro Cys Asn Ile Phe Asp Leu Gln
 100 105 110

Asn Tyr Cys Asn
 115

<210> 52

<211> 115

<212> PRT

<213> Myxine glutinosa

<400> 52

Met Ala Leu Ser Pro Phe Leu Ala Ala Val Ile Pro Leu Val Leu Leu
1 5 10 15

Leu Ser Arg Ala Pro Pro Ser Ala Asp Thr Arg Thr Thr Gly His Leu
20 25 30

Cys Gly Lys Asp Leu Val Asn Ala Leu Tyr Ile Ala Cys Gly Val Arg
35 40 45

Gly Phe Phe Tyr Asp Pro Thr Lys Met Lys Arg Asp Thr Gly Ala Leu
50 55 60

Ala Ala Phe Leu Pro Leu Ala Tyr Ala Glu Asp Asn Glu Ser Gln Asp
65 70 75 80

Asp Glu Ser Ile Gly Ile Asn Glu Val Leu Lys Ser Lys Arg Gly Ile
85 90 95

Val Glu Gln Cys Cys His Lys Arg Cys Ser Ile Tyr Asp Leu Glu Asn
100 105 110

Tyr Cys Asn
115

<210> 53

<211> 105

<212> PRT

<213> Oncorhynchus keta

<400> 53

Met Ala Phe Trp Leu Gln Ala Ala Ser Leu Leu Val Leu Leu Ala Leu
1 5 10 15

Ser Pro Gly Val Asp Ala Ala Ala Ala Gln His Leu Cys Gly Ser His
20 25 30

Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Lys Gly Phe Phe Tyr
35 40 45

Thr Pro Lys Arg Asp Val Asp Pro Leu Ile Gly Phe Leu Ser Pro Lys
50 55 60

Ser Ala Lys Glu Asn Glu Glu Tyr Pro Phe Lys Asp Gln Thr Glu Met
65 70 75 80

Met Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn
85 90 95

Ile Phe Asp Leu Gln Asn Tyr Cys Asn
100 105

<210> 54

<211> 21

<212> PRT

<213> *Myoxocephalus scorpius*

<400> 54

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asn Ile Arg Val Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 55

<211> 21

<212> PRT

<213> *Lepisosteus spatula*

<400> 55

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Thr Ile Tyr Glu Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 56

<211> 21

<212> PRT

<213> *Platichthys flesus*

<400> 56

Gly	Ile	Val	Glu	Gln	Cys	Cys	His	Lys	Pro	Cys	Asn	Ile	Phe	Asp	Leu
1				5				10						15	

Gln	Asn	Tyr	Cys	Asn
			20	

<210> 57

<211> 21

<212> PRT

<213> *Hydrolagus colliei*

<400> 57

Gly	Ile	Val	Glu	Gln	Cys	Cys	His	Asn	Thr	Cys	Ser	Leu	Ala	Asn	Leu
1				5				10						15	

Glu	Gly	Tyr	Cys	Asn
			20	

<210> 58

<211> 22

<212> PRT

<213> *Squalus acanthias*

<400> 58

Gly	Ile	Val	Glu	His	Cys	Cys	His	Asn	Thr	Cys	Ser	Leu	Tyr	Asp	Leu
1				5				10						15	

Glu Gly Tyr Cys Asn Gln
20

<210> 59

<211> 21

<212> PRT

<213> *Torpedo marmorata*

<400> 59

Gly Ile Val Glu His Cys Cys His Asn Thr Cys Ser Leu Phe Asp Leu
1 5 10 15

Glu Gly Tyr Cys Asn
20

<210> 60

<211> 89

<212> PRT

<213> *Callorhinchus milii*

<400> 60

Val Pro Thr Gln Arg Leu Cys Gly Ser His Leu Val Asp Ala Leu Tyr
1 5 10 15

Phe Val Cys Gly Glu Arg Gly Phe Phe Tyr Ser Pro Lys Gln Ile Arg
20 25 30

Asp Val Gly Pro Leu Ser Ala Phe Arg Asp Leu Glu Pro Pro Leu Asp
35 40 45

Thr Glu Met Glu Asp Arg Phe Pro Tyr Arg Gln Gln Leu Ala Gly Ser
50 55 60

Lys Met Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser
65 70 75 80

Leu Val Asn Leu Glu Gly Tyr Cys Asn
85

<210> 61

<211> 21

<212> PRT

<213> *Petromyzon marinus*

<400> 61

Gly Ile Val Glu Gln Cys Cys His Arg Lys Cys Ser Ile Tyr Asp Met
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 62

<211> 21

<212> PRT

<213> *Oncorhynchus gorboscha*

<400> 62

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 63

<211> 21

<212> PRT

<213> *Amia calva*

<400> 63

Gly Ile Val Glu Gln Cys Cys Leu Lys Pro Cys Thr Ile Tyr Glu Met
1 5 10 15

Glu Lys Tyr Cys Asn
20

<210> 64

<211> 21

<212> PRT

<213> *Anguilla rostrata*

<400> 64

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 65

<211> 113

<212> PRT

<213> *Oreochromis niloticus*

<400> 65

Met Ala Ala Leu Trp Leu Gln Ala Phe Ser Leu Leu Val Leu Met Met
1 5 10 15

Val Ser Trp Pro Gly Ser Gln Ala Val Gly Gly Pro Gln His Leu Cys
20 25 30

Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
35 40 45

Phe Phe Tyr Asn Pro Arg Arg Asp Val Asp Pro Leu Leu Gly Phe Leu
50 55 60

Pro Pro Lys Ala Gly Gly Ala Val Val Gln Gly Gly Glu Asn Glu Val
65 70 75 80

Thr Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly Ile Val Glu
85 90 95

Glu Cys Cys His Lys Pro Cys Thr Ile Phe Asp Leu Gln Asn Tyr Cys
 100 105 110

Asn

<210> 66

<211> 21

<212> PRT

<213> *Acipenser gueldenstaedti*

<400> 66

Gly Ile Val Glu Gln Cys Cys His Ser Pro Cys Ser Leu Tyr Asp Leu
 1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 67

<211> 21

<212> PRT

<213> *Piaractus mesopotamicus*

<400> 67

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
 1 5 10 15

Gln Asn Tyr Cys Asn
 20

<210> 68

<211> 115

<212> PRT

<213> *Verasper moseri*

<400> 68

Met Ala Ala Leu Trp Leu Gln Ser Val Ser Leu Leu Val Leu Met Leu
1 5 10 15

Val Ser Trp Ser Gly Ser Gln Ala Val Leu Pro Pro Gln His Leu Cys
20 25 30

Gly Ala His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly
35 40 45

Phe Phe Tyr Thr Pro Lys Arg Asp Val Asp Pro Leu Leu Gly Phe Leu
50 55 60

Pro Ala Lys Ser Gly Gly Ala Ala Ala Gly Gly Glu Asn Glu Val Ala
65 70 75 80

Glu Phe Ala Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly Ile
85 90 95

Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu Gln Asn
100 105 110

Tyr Cys Asn
115

<210> 69

<211> 30

<212> PRT

<213> *Anquilla anguilla*

<400> 69

Asp Val Glu Pro Leu Leu Gly Phe Leu Ser Pro Lys Ser Gly Gln Glu
1 5 10 15

Asn Glu Val Asp Asp Phe Pro Tyr Lys Gly Gln Gly Glu Leu
20 25 30

<210> 70

<211> 106

<212> PRT

<213> *Xenopus laevis*

<400> 70

Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Phe Phe
1 5 10 15

Ser Thr Pro Asn Thr Glu Ala Leu Val Asn Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35 40 45

Tyr Tyr Pro Lys Val Lys Arg Asp Met Glu Gln Ala Leu Val Ser Gly
50 55 60

Pro Gln Asp Asn Glu Leu Asp Gly Met Gln Leu Gln Pro Gln Glu Tyr
65 70 75 80

Gln Lys Met Lys Arg Gly Ile Val Glu Gln Cys Cys His Ser Thr Cys
85 90 95

Ser Leu Phe Gln Leu Glu Ser Tyr Cys Asn
100 105

<210> 71

<211> 106

<212> PRT

<213> *Xenopus laevis*

<400> 71

Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Leu Phe
1 5 10 15

Ser Thr Pro Asn Thr Glu Ala Leu Ala Asn Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35 40 45

Tyr Tyr Pro Lys Ile Lys Arg Asp Ile Glu Gln Ala Gln Val Asn Gly
50 55 60

Pro Gln Asp Asn Glu Leu Asp Gly Met Gln Phe Gln Pro Gln Glu Tyr
65 70 75 80

Gln Lys Met Lys Arg Gly Ile Val Glu Gln Cys Cys His Ser Thr Cys
85 90 95

Ser Leu Phe Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 72

<211> 21

<212> PRT

<213> *Trachemys scripta*

<400> 72

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 73

<211> 21

<212> PRT

<213> *Alligator mississippiensis*

<400> 73

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 74

<211> 21

<212> PRT

<213> Zaocys dhumnades

<400> 74

Gly	Ile	Val	Glu	Gln	Cys	Cys	Glu	Asn	Thr	Cys	Ser	Leu	Tyr	Glu	Leu
1				5				10						15	

Glu	Asn	Tyr	Cys	Asn
			20	

<210> 75

<211> 21

<212> PRT

<213> Crotalus atrox

<400> 75

Gly	Ile	Val	Glu	Gln	Cys	Cys	Glu	Asn	Thr	Cys	Ser	Leu	Tyr	Gln	Leu
1				5				10						15	

Glu	Asn	Tyr	Cys	Asn
			20	

<210> 76

<211> 114

<212> PRT

<213> Штучна послідовність

<220>

<223> Препроінсулін

<400> 76

Met Gly Leu Trp Ile Arg Leu Leu Pro Leu Ile Ala Leu Leu Ile Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Glu Phe Arg Met Phe Val Asn Gln
20 25 30

His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly
35 40 45

Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp
50 55 60

Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser
65 70 75 80

Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val
85 90 95

Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr
100 105 110

Cys Asn

<210> 77

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 77

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 78

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 78

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

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

<210> 79

<211> 53

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 79

Phe	Val	Asp	Gln	His	Leu	Cys	Gly	Ser	His	Leu	Val	Glu	Ala	Leu	Tyr
1				5					10					15	

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

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

Glu	Asp	Tyr	Cys	Asn
	50			

<210> 80

<211> 53

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 80

Phe Val Glu Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

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

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

Glu Glu Tyr Cys Asn
50

<210> 81

<211> 53

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 81

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

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

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

Glu Asn Tyr Cys Gly
50

<210> 82

<211> 53

<212> PRT

<213> Штучна послідовність

<220>

<223> Білковий продукт без назви з гомологією інсуліну

<400> 82

Phe Val Thr Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

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

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

Glu His Tyr Cys Ser
50

<210> 83

<211> 57

<212> PRT

<213> Штучна послідовність

<220>

<223> Проінсулін

<400> 83

Asn Ser Asn Gly Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
1 5 10 15

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

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

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

<210> 84

<211> 58

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 84

Asn Ser Asn Gly Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
1 5 10 15

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

Pro Lys Thr Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
35 40 45

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

<210> 85

<211> 50

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 85

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

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

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

Cys Asn
50

<210> 86

<211> 54

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 86

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 Tyr Cys Asn
50

<210> 87

<211> 61

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 87

Lys Glu Thr Leu Thr Ile Thr Cys Ala Val Pro Thr Trp Leu Lys Leu
1 5 10 15

Trp Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Asn Leu Arg Leu
20 25 30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Pro Cys Thr
35 40 45

Asn Trp Lys Thr Thr Ala Thr Arg Arg Ser Pro Gln Ala
50 55 60

<210> 88

<211> 61

<212> PRT

<213> Штучна послідовність

<220>

<223> Препроінсулін

<400> 88

Lys Asp Ser Leu Thr Asn Thr Cys Ala Val Ser Thr Trp Leu Lys Leu
1 5 10 15

Cys Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Leu Leu Arg Leu
20 25 30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Asn Tyr Thr
35 40 45

Asn Trp Lys Thr Thr Ala Thr Arg Arg Ser Pro Gln Ala
50 55 60

<210> 89

<211> 87

<212> PRT

<213> Штучна послідовність

<220>

<223> Проінсулін

<400> 89

Met 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 Thr Arg
20 25 30

Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly
35 40 45

Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln
50 55 60

Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr
65 70 75 80

Gln Leu Glu Asn Tyr Cys Asn
85

<210> 90

<211> 35

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 90

Arg Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly
1 5 10 15

Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu
20 25 30

Gln Lys Arg
35

<210> 91

<211> 23

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 91

Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln Phe
1 5 10 15

Val Cys Gly Asp Arg Gly Phe
20

<210> 92

<211> 124

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 92

Met Lys Leu Lys Thr Val Arg Ser Ala Val Leu Ser Ser Leu Phe Ala
1 5 10 15

Ser Gln Val Leu Gly Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr
20 25 30

Ser Val Asn Leu Met Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln
35 40 45

Thr Asn Ser Gly Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys
50 55 60

Arg Glu Glu Gly Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser
65 70 75 80

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe
85 90 95

Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser
100 105 110

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
115 120

<210> 93

<211> 171

<212> ДНК

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 93

tttgtcaatc agcacctttg tggttctcac ctggtggagg ctctgtacct ggtgtgtggg 60
gaacgtgggtt tctttctacac acccaagacc cgtcgttaagc ttaagcgtgg cattgtggag 120
cagtgcctgca ccagcatctg ctccctctac caactggaga actactgcaa c 171

<210> 94

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Аналог інсуліну

<400> 94

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 95

<211> 26

<212> PRT

<213> Штучна послідовність

<220>

<223> Аналог інсуліну

<220>

<221> MISC_FEATURE

<222> (26)..(26)

<223> X = будь-яка амінокислота

<400> 95

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

Leu Val Cys Gly Glu Arg Gly Phe Phe Xaa
20 25

<210> 96

<211> 21

<212> PRT

<213> Homo sapiens

<400> 96

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 97

<211> 29

<212> PRT

<213> Homo sapiens

<400> 97

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

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

<210> 98

<211> 21

<212> PRT

<213> Homo sapiens

<400> 98

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 99

<211> 30

<212> PRT

<213> Homo sapiens

<400> 99

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

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

<210> 100

<211> 21

<212> PRT

<213> Homo sapiens

<400> 100

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 101

<211> 30

<212> PRT

<213> Homo sapiens

<400> 101

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

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

<210> 102

<211> 21

<212> PRT

<213> Homo sapiens

<400> 102

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 103

<211> 30

<212> PRT

<213> Homo sapiens

<400> 103

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

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

<210> 104

<211> 21

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> X = будь-яка амінокислота

<400> 104

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 105

<211> 30

<212> PRT

<213> Homo sapiens

<400> 105

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

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

<210> 106

<211> 21

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> X = будь-яка амінокислота

<400> 106

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 107

<211> 30

<212> PRT

<213> Homo sapiens

<400> 107

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

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

<210> 108

<211> 21

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X = будь-яка амінокислота

<400> 108

Gly Xaa Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 109

<211> 30

<212> PRT

<213> Homo sapiens

<400> 109

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

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

<210> 110
<211> 21
<212> PRT
<213> Homo sapiens

<400> 110
Gly Ala Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 111
<211> 30
<212> PRT
<213> Homo sapiens

<400> 111
Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

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

<210> 112
<211> 21
<212> PRT
<213> Штучна послідовність

<220>
<223> Аналог інсуліну
<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X = будь-яка амінокислота

<400> 112

Gly	Xaa	Val	Glu	Gln	Cys	Cys	Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu
1				5					10					15	

Glu	Asn	Tyr	Cys	Asn
			20	

<210> 113

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Аналог інсуліну

<400> 113

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

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

<210> 114

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Аналог інсуліну

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X = будь-яка амінокислота

<400> 114

Gly Xaa Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 115

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Аналог інсуліну

<400> 115

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

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

<210> 116

<211> 21

<212> PRT

<213> Homo sapiens

<400> 116

Gly Gly Gly Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 117

<211> 30

<212> PRT

<213> Homo sapiens

<400> 117

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

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

<210> 118

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант інсуліну Homo sapiens

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> X = будь-яка амінокислота

<400> 118

Xaa Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 119

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант інсуліну Homo sapiens

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> X = будь-яка амінокислота

<400> 119

Xaa	Val	Asn	Gln	His	Leu	Cys	Gly	Asp	His	Leu	Val	Glu	Ala	Leu	Tyr
1			5					10						15	

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

<210> 120

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант інсуліну Homo sapiens

<400> 120

Gly	Ile	Val	Glu	Gln	Cys	Cys	Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu
1				5				10						15	

Glu	Asn	Tyr	Cys	Asn
			20	

<210> 121

<211> 29

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант інсуліну Homo sapiens

<400> 121

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

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

<210> 122

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<400> 122

Gly Ile Val Glu Gln Ser Cys Thr Ser Ile Ser Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 123

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<400> 123

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

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

<210> 124

<211> 79

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 124

Met Asp Pro Gly Asp Pro Glu Cys Leu Glu Gln Leu Leu Arg Arg Leu
1 5 10 15

Gly Gly Ser Val Glu Val Glu Val Thr Gly Gly Thr Val His Val Glu
20 25 30

Val Ser Pro Glu Asp Pro Gly Asp Pro Glu Cys Leu Glu Gln Leu Leu
35 40 45

Arg Arg Leu Gly Gly Ser Val Glu Val Glu Val Thr Gly Gly Thr Val
50 55 60

His Val Glu Val Ser Pro Gly Glu Arg Gly Phe Phe Tyr Cys Asn
65 70 75

<210> 125

<211> 87

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін

<400> 125

Met Leu Lys Glu Lys Lys Tyr Ser Pro Asp Pro Gly Asp Pro Glu Cys
1 5 10 15

Leu Glu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu Val Glu Val
20 25 30

Thr Gly Gly Thr Val His Val Glu Val Ser Pro Glu Asp Pro Gly Asp
35 40 45

Pro Glu Cys Leu Glu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu
50 55 60

Val Glu Val Thr Gly Gly Thr Val His Val Glu Val Ser Pro Gly Glu
65 70 75 80

Arg Gly Phe Phe Tyr Cys Asn
85

<210> 126

<211> 23

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<220>

<221> MISC_FEATURE

<222> (5) .. (5)

<223> X = будь-яка амінокислота

<220>

<221> MISC_FEATURE
<222> (18)..(18)
<223> X = будь-яка амінокислота

<400> 126

Met Ala Thr Ser Xaa Ser Thr Lys Lys Thr Gln Leu Gln Leu Glu His
1 5 10 15

Leu Xaa Leu Asp Leu Gln Met
20

<210> 127

<211> 23

<212> PRT

<213> Escherichia coli

<220>

<221> MISC_FEATURE

<222> (14)..(14)

<223> X = будь-яка амінокислота

<220>

<221> MISC_FEATURE

<222> (17)..(17)

<223> X = будь-яка амінокислота

<400> 127

Thr Met Ile Thr Asp Ser Leu Ala Val Val Leu Gln Arg Xaa Asp Trp
1 5 10 15

Xaa Pro Gly Val Thr Gln Leu
20

<210> 128

<211> 96

<212> PRT

<213> Brevibacillus brevis

<400> 128

Asn Ser Val Leu Ala Ser Ala Leu Ala Leu Thr Val Ala Pro Met Ala
1 5 10 15

Phe Ala Asn Ser Asp Ser Glu Ser Pro Leu Ser His Asp Gly Tyr Ser
20 25 30

Leu His Asp Gly Val Ser Met Tyr Ile Glu Ala Leu Asp Lys Phe Val
35 40 45

Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val
50 55 60

Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Gly Ile Val Glu Gln
65 70 75 80

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
85 90 95

<210> 129

<211> 25

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<400> 129

Asp Thr Thr Met Pro Ala Gly Gly Gly Gly Gly Gly Gln His Leu Cys
1 5 10 15

Gly Pro His Leu Val Glu Ala Leu Tyr
20 25

<210> 130
<211> 6
<212> PRT
<213> Штучна послідовність

<220>
<223> Злитий білок інсуліну
<400> 130

Leu Glu Asn Tyr Cys Asn
1 5

<210> 131
<211> 22
<212> PRT
<213> Штучна послідовність

<220>
<223> Злитий білок інсуліну
<400> 131

Met Thr Met Ile Thr Asp Ser Leu Glu Phe Gln Ala Trp Gly Gly Gly
1 5 10 15

Gly Gly Trp Met Arg Phe
20

<210> 132
<211> 22
<212> PRT
<213> Штучна послідовність

<220>

<223> Злитий білок інсуліну

<400> 132

Met Val Leu Arg Phe Leu Pro Leu Leu Ala Leu Leu Val Leu Trp Glu
1 5 10 15

Pro Lys Pro Ala Gln Ala
20

<210> 133

<211> 60

<212> PRT

<213> Штучна послідовність

<220>

<223> Міні-проінсулін

<400> 133

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

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

Tyr Pro Gly Asp Val Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser
35 40 45

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55 60

<210> 134

<211> 21

<212> PRT

<213> Homo sapiens

<400> 134

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 135

<211> 30

<212> PRT

<213> Homo sapiens

<400> 135

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

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

<210> 136

<211> 21

<212> PRT

<213> Homo sapiens

<400> 136

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 137

<211> 30

<212> PRT

<213> Homo sapiens

<400> 137

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

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

<210> 138

<211> 14

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<400> 138

Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Gly
1 5 10

<210> 139

<211> 21

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> X = будь-яка послідовність

<400> 139

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 140

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<400> 140

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

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

<210> 141

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> X = будь-яка амінокислота

<400> 141

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 142

<211> 30

<212> PRT

<213> Штучна послідовність

<220>

<223> Інсулін Homo sapiens

<400> 142

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

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

<210> 143

<211> 21

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант міні-проінсуліну

<400> 143

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 144

<211> 29

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант міні-проінсуліну

<400> 144

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

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys
20 25

<210> 145

<211> 50

<212> PRT

<213> Штучна послідовність

<220>

<223> Мутант міні-проінсуліну

<400> 145

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

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys Gly Ile Val
20 25 30

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

Cys Asn
50

<210> 146

<211> 3

<212> PRT

<213> Штучна послідовність

<220>

<223> С-пептид інсуліну

<400> 146

Ala Ala Lys

1

<210> 147

<211> 3

<212> PRT

<213> Штучна послідовність

<220>

<223> С-пептид інсуліну

<400> 147

Asn Lys Arg

1

<210> 148

<211> 6

<212> PRT

<213> Штучна послідовність

<220>

<223> С-пептид інсуліну

<400> 148

Arg Arg Lys Gln Lys Arg

1

5

<210> 149

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Сайт розщеплення

<400> 149

Arg Arg Lys Arg
1

<210> 150

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Послідовність, яка утримує ER

<400> 150

Lys Asp Glu Leu
1

<210> 151

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Послідовність, яка утримує ER

<400> 151

His Asp Glu Leu
1

<210> 152

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Послідовність, яка утримує ER

<400> 152

Asp Asp Glu Leu
1

<210> 153

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Послідовність, яка утримує ER

<400> 153

Ala Asp Glu Leu
1

<210> 154

<211> 4

<212> PRT

<213> Штучна послідовність

<220>

<223> Послідовність, яка утримує ER

<400> 154

Ser Asp Glu Leu
1

<210> 155

<211> 4

<212> PRT

<213> *Lycopersicon esculentum* Mill.

<400> 155

His Asp Glu Phe
1

<210> 156

<211> 118

<212> PRT

<213> *Arabidopsis thaliana*

<400> 156

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys
115

<210> 157

<211> 187

<212> PRT

<213> Brassica napus

<400> 157

Met Ala Asp Thr Ala Arg Thr His His Asp Val Thr Ser Arg Asp Gln
1 5 10 15

Tyr Pro Arg Asp Arg Asp Gln Tyr Ser Met Ile Gly Arg Asp Arg Asp
20 25 30

Gln Tyr Ser Met Met Gly Arg Asp Arg Asp Gln Tyr Asn Met Tyr Gly
35 40 45

Arg Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Val Thr Ala Val
50 55 60

Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu Val Gly
65 70 75 80

Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile Phe Ser
85 90 95

Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile Thr Gly
100 105 110

Phe Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala Ile Thr Val Phe Ser
115 120 125

Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
130 135 140

Leu Asp Ser Ala Arg Met Lys Leu Gly Thr Lys Ala Gln Asp Ile Lys
145 150 155 160

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
165 170 175

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
180 185

<210> 158

<211> 748

<212> ДНК

<213> *Arabidopsis thaliana*

<400> 158

```
taccatgggg tcaaagacgg agatgatgga gagagacgca atggctacgg tggctcccta 60
tgcgccggtc acttaccatc gccgtgctcg tgttgacttg gatgatagac ttccctaaacc 120
ttatatgcc aagacattgc aagcaccaga cagagaacac ccgtacggaa ctccaggcca 180
taagaattac ggacttagtg ttcttcaaca gcatgtctcc ttcttcgata tcgatgataa 240
tggcatcatt tacccttggg agacctactc tggactgcga atgcttgggt tcaatatcat 300
tgggtcgttt ataatagccg ctgttatcaa cctgacctt agctatgcc ctcttcggg 360
gtggttacct tcacctttct tccctatata catacacaac atacacaagt caaagcatgg 420
aagtgattca aaaacatatg acaatgaagg aagggttatg ccggtgaatc ttgagttgat 480
atttagcaaa tatgcgaaaa ccttgccaga caagttgagt cttggagaac tatgggagat 540
gacagaagga aaccgtgacg cttgggacat ttttgatgg atgcaggca aaatagagtg 600
gggactgttg tacttgctag caagggatga agaaggggtt ttgtcaaaag aagctattag 660
gcgggtgttt gatggaagct tgttcgagta ctgtgccaaa atctacgctg gtatcagtga 720
agacaagaca gcatactacg ccatggat 748
```

<210> 159

<211> 738

<212> ДНК

<213> *Arabidopsis thaliana*

<400> 159

```
atgggggtcaa agacggagat gatggagaga gacgcaatgg ctacgggtggc tccctatgog 60
ccgggtcactt accaccggcg tgctcgtgtt gacttgatg atagacttcc taaaccttat 120
atgccaaagag cattgcaagc accagacaga gaacacccgt acggaactcc aggccataag 180
aattacggac ttagtgttct tcaacagcat gtctccttct tcgatatcga tgataatggc 240
atcatttacc cttgggagac ctactctgga ctgcgaatgc ttggtttcaa tatcattggg 300
tcgcttataa tagccgctgt tatcaacctg acccttagct atgccactct tccgggggtg 360
```

ttaccttcac ctttcttccc tatatacata cacaacatac acaagtcaaa gcatggaagt	420
gattcaaaaa catatgacaa tgaaggaagg tttatgccgg tgaatcttga gttgatattt	480
agcaaatatg cgaaaacctt gccagacaag ttgagtcttg gagaactatg ggagatgaca	540
gaaggaaacc gtgacgcttg ggacattttt ggatggatcg caggcaaat agagtgggga	600
ctgttgact tgctagcaag ggatgaagaa gggtttttgt caaaagaagc tattaggcgg	660
tgtttogatg gaagcttggt cgagtactgt gccaaaatct acgctggtat cagtgaagac	720
aagacagcat actactaa	738

<210> 160

<211> 1047

<212> ДНК

<213> Sesamum indicum

<400> 160

atggatctaa tccacacttt cctcaactta atagctcccc ctttcacott cttcttcott	60
ctctttttct tgccaccott ccagattttc aagttcttcc tttcaatott gggcaccott	120
ttcagcgagg atgtcgctgg aaaagtcgtc gtcatcacog gcgcctctc cggcatcggc	180
gaaagtcttg cttacgagta tgctaagaga ggggcgtgct tggtgcttgc tgcaagaagg	240
gaaaggagtc ttcaagaagt ggcgaaaagg gcgcgcgatt tggggtcgcc ggacgtcgtg	300
gtggtcgggg ccgatgttcc gaaggcggag gactgcagga aggttggtga tcagactatg	360
aatcgctttg gaagattgga tcacctggtc aataacgctg gaattatgtc agtttcaatg	420
ctggaagaag ttgaagatat tactggttac agagaaacta tggatatcaa cttctggggc	480
tatgtgtata tgacccgatt tgccgcccca taccttagga atagcagagg ccgaattgtt	540
gtactttctt catccagttc ttggatgcct actccgagga tgagttttta caatgcaagc	600
aaagcggcga tttcacaatt ttttgagaca ctgcgggtgg aattcggccc cgatataggc	660
ataacccttg tgactccagg attcatagaa tetgaactta cccaaggcaa attctacaat	720
gctggcgaac gtgtaattga tcaggacatg agagatgtac aagtgagcac gactccaatc	780
ctgaggggtg aaagtgcggc aaggtaatc gtgaggagcg cgatccgtgg agaaagatac	840
gtgacagagc cggcctgggt tagggttact tattgggtga agctattctg ccctgaggtg	900
atggagtggg tatttagact gatgtacttg gccagcccg gtgagccgga gaaggaaacg	960

tttggcaaga aggttttgga ttacacagga gtgaagtctt tgctttaccc ggaaaccgtg 1020
 caagttccgg agcccaagaa tgattaa 1047

<210> 161

<211> 25

<212> PRT

<213> Тютюн, сигнальна послідовність білка, спорідненого з
 патогенезом (PR-S)

<400> 161

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
 1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala
 20 25

<210> 162

<211> 64

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора .

<400> 162

Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Glu Ala Val
 1 5 10 15

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

Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asx Thr Thr Ile
 35 40 45

Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Met Ala Lys Arg
 50 55 60

<210> 163

<211> 63

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 163

Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Glu Ala Val
1 5 10 15

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

Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asx Thr Thr Ile
35 40 45

Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Met Ala Lys Arg
50 55 60

<210> 164

<211> 15

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 164

Gln Pro Ile Asp Glu Asp Asn Asp Thr Ser Ser Met Ala Lys Arg
1 5 10 15

<210> 165

<211> 43

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 165

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40

<210> 166

<211> 43

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 166

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40

<210> 167

<211> 43

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 167

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
35 40

<210> 168

<211> 43

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 168

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
35 40

<210> 169

<211> 45

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 169

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Gly
20 25 30

Gly Leu Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40 45

<210> 170

<211> 44

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 170

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly Gly
20 25 30

Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys Arg
35 40

<210> 171

<211> 44

<212> PRT

<213> Штучна послідовність

<220>

<223> Лідерна послідовність альфа-фактора

<400> 171

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly Gly
20 25 30

Leu Asp Val Val Gly Leu Ile Ser Met Ala Ala Ala
35 40

<210> 172

<211> 10

<212> PRT

<213> Пептид роздільника

<400> 172

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys
1 5 10

<210> 173

<211> 6

<212> PRT

<213> Штучна послідовність

<220>

<223> Пептид роздільника

<400> 173

Glu Glu Gly Glu Pro Lys
1 5

<210> 174

<211> 40

<212> PRT

<213> Штучна послідовність

<220>

<223> Сайт розщеплення

<400> 174

Met Ala Glu Ile Thr Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg
1 5 10 15

Lys Ala Leu Lys Glu His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln
20 25 30

Gln Tyr Gly Ile Ser Ser Lys Phe
35 40

<210> 175

<211> 27

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 175
gcatgctgac attgtgatga cacagtc

27

<210> 176

<211> 43

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 176
aagcttgcat ttaaatactc gagactgtga gagtgggtgcc ttg

43

<210> 177

<211> 70

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 177

gaagaaggag agcctaagtt tgttaatcaa catctttgtg gatctcatct tgttgaggct 60

ctctaccttg 70

<210> 178

<211> 56

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 178

ccttaggagt gtagaaaaat cctctttctc cacacacaag gtagagagcc tcaaca 56

<210> 179

<211> 21

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 179

ctaaggctgc taagggaatt g 21

<210> 180

<211> 83

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 180

aagcttcagt tgcaatagtt ctccaattgg taaagtgagc aaatagaagt gcaacattgt 60

tcaacaattc ccttagcagc ctt 83

<210> 181

<211> 30

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 181

ctcgagtcaa ccaattgatg acactgaatc 30

<210> 182

<211> 41

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 182

aagcttcaaa gttcatcctt gttgcaatag ttctccaatt g 41

<210> 183

<211> 21

<212> ДНК
 <213> Штучна послідовність

 <220>
 <223> Праймер
 <400> 183
 aagcttcagt tgcaatagtt c 21

 <210> 184
 <211> 26
 <212> ДНК
 <213> Штучна послідовність

 <220>
 <223> Праймер
 <400> 184
 gcatgssaa ccaattgatg acactg 26

 <210> 185
 <211> 34
 <212> ДНК
 <213> Штучна послідовність

 <220>
 <223> Праймер
 <400> 185
 gcatgcatgc ctttgtaat caacatcttt gtgg 34

 <210> 186
 <211> 54
 <212> ДНК
 <213> Штучна послідовність

<220>

<223> Праймер

<400> 186

acattgttca acaattcctc tctttcttct agtcttagga gtgtagaaaa atcc

54

<210> 187

<211> 29

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 187

gcataagctt caaagctcat cctttgagc

29

<210> 188

<211> 387

<212> ДНК

<213> Штучна послідовність

<220>

<223> Послідовність нуклеинової кислоти злитого білка інсуліну

<400> 188

atgaacttcc ttaagtcttt ccctttctac gctttccttt gtttcgggtca atacttcggtt 60

gctgttacgc atgccaacc aattgatgac actgaatccc agaccacgtc agtgaacctc 120

atggccgatg atactgagag cgcgtttgct acacaaacaa attcgggagg tcttgacggt 180

gtcggattga tctccatggc taagagagaa gaaggagagc ctaagtttgt taatcaacat 240

ctttgtggat ctcatcttgt tgaggctctc taccttggtg gtggagaaag aggatttttc 300

tacactccta aggctgctaa gggaattggt gaacaatggt gcacttcctat ttgctcactt 360

taccaatttq aqaactattq caactqa 387

<210> 189

<211> 128

<212> PRT

<213> Штучна послідовність

<220>

<223> Білок фактора інсуліну

<400> 189

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala Gln Pro Ile Asp Asp Thr Glu
20 25 30

Ser Gln Thr Thr Ser Val Asn Leu Met Ala Asp Asp Thr Glu Ser Ala
35 40 45

Phe Ala Thr Gln Thr Asn Ser Gly Gly Leu Asp Val Val Gly Leu Ile
50 55 60

Ser Met Ala Lys Arg Glu Glu Gly Glu Pro Lys Phe Val Asn Gln His
65 70 75 80

Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu
85 90 95

Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln
100 105 110

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
115 120 125

<210> 190

<211> 18

<212> ДНК

<213> Штучна послідовність

<220>		
<223>	Праймер	
<400>	190	
	ttcgtgaacc aacacttg	18
<210>	191	
<211>	20	
<212>	ДНК	
<213>	Штучна послідовність	
<220>		
<223>	Праймер	
<400>	191	
	aagctttcag ttacagtagt	20
<210>	192	
<211>	17	
<212>	ДНК	
<213>	Штучна послідовність	
<220>		
<223>	Праймер	
<400>	192	
	gcatgcatgt gttgagc	17
<210>	193	
<211>	16	
<212>	ДНК	
<213>	Штучна послідовність	
<220>		
<223>	Праймер	

<400> 193
ggtagtgtgc tggcca

16

<210> 194

<211> 38

<212> ДНК

<213> Штучна послідовність

<220>

<223> Праймер

<400> 194
ggtggccagc acactacett cgtgaacca cacttgtg

38

<210> 195

<211> 1020

<212> ДНК

<213> Штучна послідовність

<220>

<223> Послідовність нуклеїнової кислоти злитого білка інсуліну

<400> 195
atggcgata sagctagagg aaccatcac gatatcatcg gcagagacca gtaccgatg 60
atgggcccag accgagacca gtaccagatg tccggacgag gatctgacta ctccaagtct 120
aggcagattg ctaaagctgc aactgctgtc acagctgggtg gttccctcct tgttctctcc 180
agccttacct ttgttggaac tgtcatagct ttgactgttg caacacctct gctcgttatc 240
ttcagcccaa tccttgtccc ggctctcatc acagttgcac tcctcatcac cggttttctt 300
tcctctggag ggtttgcat tgcgctata accgttttct cttggattta caagtaagca 360
cacatttatc atcttacttc ataattttgt gcaatatgtg catgcatgtg ttgagccagt 420
agctttggat caattttttt ggtagaataa caaatgtaac aataagaaat tgcaaattct 480
agggaacatt tggttaacta aatacgaaat ttgacctagc tagcttgaat gtgtctgtgt 540
atatcatcta tataggtaaa atgcttggtg tgatacctat tgattgtgaa taggtacgca 600

acgggagagc acccacaggg atcagacaag ttggacagtg caaggatgaa gttgggaagc 660
aaagctcagg atctgaaaga cagagctcag tactacggac agcaacatac tgggtggggaa 720
catgaccgtg accgtactcg tgggtggccag cacactacct tcgtgaacca acacttgtgt 780
ggatctcatc tcgttgaagc totctacttg gtttgtggtg agagaggatt cttctacact 840
cctaagacca gaagggaagc tgaggacttg caggtgggac aagttgagtt ggggtggaggt 900
cctggagcag gatctttgca acctctcgct ttggaagggt ctttgcagaa gagaggaatc 960
gttgaacaat gttgcacttc aatctgttct ttgtatcagt tggagaacta ctgtaactga 1020

<210> 196

<211> 257

<212> PRT

<213> Штучна послідовність

<220>

<223> Злитий блок інсуліну

<400> 196

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
115 120 125

Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys
130 135 140

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
145 150 155 160

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr Phe Val Asn Gln His
165 170 175

Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu
180 185 190

Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu
195 200 205

Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu
210 215 220

Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu
225 230 235 240

Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys
245 250 255

Asn