



(19) **UA** (11) **82 485** (13) **C2**
(51) МПК

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
УКРАИНЫ

ГОСУДАРСТВЕННЫЙ ДЕПАРТАМЕНТ
ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ УКРАИНЫ

(21), (22) Заявка: 97084103, 06.11.1996
(24) Дата начала действия патента: 25.04.2008
(30) Приоритет: 06.11.1995 US 60/007,255
28.02.1996 US 08/608,423
28.08.1996 US 08/705,484
(46) Дата публикации: 25.04.2008 A01N 63/00
20060101CFI20070115RMUA C07K 14/195
20060101CLI20070115RMUA C12N 1/20
20060101CLI20070115RMUA
(86) Заявка PCT:
PCT/US96/18003, 19961106

(72) Изобретатель:
Инсайн Джеральд К., US,
Бауен Девид Дж., US,
Пителл Джеймс, US,
Фатиг Реймонд, US,
Шуновер Сю, US,
Ффренч-Констант Ричард Г., US,
Рочело Томас А., US,
Блексберн Майкл Б., US,
Хей Тимоти Д., US,
Мерло Дональд Дж., US,
Орр Грегори Л., US,
Робертс Джин Л., US,
Стрикленд Джеймс А., US,
Гуо Лининг, US,
Киче Тодд А., US
(73) Патентовладелец:
ВИСКОНСИН АЛЮМНИ РИСЕЧ ФАУНДЕЙШЕ,
US

(54) ИНСЕКТИЦИДНЫЕ БЕЛКОВЫЕ ТОКСИНЫ ИЗ PHOTORHABDUS

(57) Реферат:
Изобретение относится к белковому токсину, выделенному из бактерий рода Photorhabdus, что имеет аминокислотную последовательность SEQ ID NO: 47, полинуклеотида, кодирующего этот белок, где нуклеотидная последовательность представляет собой SEQ ID NO: 46, и к их применению в качестве инсектицидов.

Официальный бюллетень "Промышленная собственность". Книга 1 "Изобретения, полезные модели, топографии интегральных микросхем", 2008, N 8, 25.04.2008. Государственный департамент интеллектуальной собственности Министерства образования и науки Украины.



(19) **UA** (11) **82 485** (13) **C2**
(51) Int. Cl.

MINISTRY OF EDUCATION AND SCIENCE OF
UKRAINE

STATE DEPARTMENT OF INTELLECTUAL
PROPERTY

(12) **DESCRIPTION OF PATENT OF UKRAINE FOR INVENTION**

(21), (22) Application: 97084103, 06.11.1996

(24) Effective date for property rights: 25.04.2008

(30) Priority: 06.11.1995 US 60/007,255
28.02.1996 US 08/608,423
28.08.1996 US 08/705,484

(46) Publication date: 25.04.2008 A01N 63/00
20060101CFI20070115RMUA C07K 14/195
20060101CLI20070115RMUA C12N 1/20
20060101CLI20070115RMUA

(86) PCT application:
PCT/US96/18003, 19961106

(72) Inventor:

ENSIGN JERALD C, US,
BOWEN DAVID J, US,
PETELL JAMES, US,
FATIG RAYMOND, US,
SCHOONOVER SUE, US,
FFRENCH CONSTANT RICHARD H, US,
ROCHELEAU THOMAS A, US,
BLACKBURN MICHAEL B, US,
HEY TIMOTHY D, US,
MERLO DONALD J, US,
ORR GREGORY L, US,
ROBERTS JEAN L, US,
STRICKLAND JAMES A (US), US,
GUO LINING, US,
CICHE TODD A, US

(73) Proprietor:

WISCONSIN ALUMNI RESEARCH
FOUNDATION, US

(54) **INSECTICIDE PROTEIN TOXINS FROM PHOTORHABDUS**

(57) Abstract:

Proteins from the genus *Photorhabdus* are toxic to insects upon exposure. *Photorhabdus luminescens* (formerly *Xenorhabdus luminescens*) have been found in mammalian clinical samples and as a bacterial symbiont of entomopathogenic nematodes of genus *Heterorhabditis*. These protein toxins can be applied to, or genetically

engineered into, insect larvae food and plants for insect control.

Official bulletin "Industrial property". Book 1 "Inventions, utility models, topographies of integrated circuits", 2008, N 8, 25.04.2008. State Department of Intellectual Property of the Ministry of Education and Science of Ukraine.



(19) **UA** (11) **82 485** (13) **C2**
(51)МПК

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ ВЛАСНОСТІ

(12) ОПИС ВІНАХОДУ ДО ПАТЕНТУ УКРАЇНИ

(21), (22) Дані стосовно заявки:
97084103, 06.11.1996

(24) Дата набуття чинності: 25.04.2008

(30) Дані стосовно пріоритету відповідно до Паризької
конвенції : 06.11.1995 US 60/007,255
28.02.1996 US 08/608,423
28.08.1996 US 08/705,484

(46) Публікація відомостей про видачу патенту
(деклараційного патенту): 25.04.2008 A01N 63/00
20060101CFI20070115RMUA C07K 14/195
20060101CLI20070115RMUA C12N 1/20
20060101CLI20070115RMUA

(86) Номер та дата подання міжнародної заявки
відповідно до договору PCT:
PCT/US96/18003, 19961106

(72) Винахідник(и):

Інсайн Джеральд К., US,
Бауєн Девід Дж., US,
Пітелл Джеймс, US,
Фатіг Реймонд, US,
Шуновер Сью, US,
Ффренч-Констант Річард Г., US,
Рочело Томас А., US,
Блекберн Майкл Б., US,
Хей Тімоті Д., US,
Мерло Дональд Дж., US,
Орр Грегорі Л., US,
Робертс Джін Л., US,
Стрікленд Джеймс А., US,
Гуо Лінінг, US,
Кіче Тодд А., US

(73) Власник(и):

ВІСКОНСІН АЛЮМНІ РІСЬОЧ ФАУНДЕЙШН, US

(54) ІНСЕКТИЦИДНІ БІЛКОВІ ТОКСИНИ З PHOTORHABDUS

(57) Реферат:

Винахід стосується білкового токсину,
виділеного з бактерій роду Photorhabdus, що має
амінокислотну послідовність SEQ ID NO: 47,

полінуклеотиду, що кодує цей білок, де
нуклеотидна послідовність являє собою SEQ ID
NO: 46, та до їх застосування як інсектицидів.

Опис винаходу

Цей винахід стосується токсинів, виділених з бактерій, та до застосування цих токсинів як інсектицидів.

Численні комахи широко розглядаються як шкідники володарями будинків, учасниками пікніків, садівниками й фермерами та іншими людьми, вкладення яких у сільськогосподарські продукти часто пропадають або зменшуються внаслідок пошкодження комахами польових культур. Зокрема, в зонах, де вегетаційний сезон є коротким, значне пошкодження комахами може означати втрату всіх прибутків для фермерів і сильне зниження врожайності. Мізерне постачання конкретних сільськогосподарських продуктів неминуче призводить до більш високих цін для людей, що займаються переробкою харчових продуктів, і потім - для кінцевих споживачів харчових рослин та продуктів, які одержують з цих рослин.

Відвертання пошкодження комахами сільськогосподарських культур та квітів і виключення неприємностей, пов'язаних з комахами-шкідниками, як правило засновано на використанні сильних пестицидів та інсектицидів з широкими спектрами токсичності. Ці синтетичні продукти стикаються з різким неприйняттям усього населення як занадто жорстокі по відношенню до довкілля та об'єктів, що їх піддають дії таких агентів. Таким чином, поза сільським господарством, володарі будинків задовольнилися б тим, щоб комахи уникали їх будинків та місць пікніків, без знищення цих комах.

Інтенсивне застосування хімічних інсектицидів викликало занепокоєність щодо довкілля та здоров'я у фермерів, компаній, які продукують ці інсектициди, державних органів, груп захисту суспільних інтересів та громадськості в цілому. Розробка менш жорстких стратегій захисту від шкідників стимулювалася разом з стурбованістю суспільства про довкілля та з розвитком біологічних засобів, що використовують біологічні механізми боротьби з комахами. Засоби біологічного захисту є перспективною альтернативою хімічним інсектицидам.

Організми на кожному рівні еволюційного розвитку розвивали засоби для підсилення їх власного поступу й виживання. Використання біологічних молекул як інструментів захисту й агресії відомо в царстві тварин і в рослинному світі. Крім того, відносно нові інструменти генної інженерії дозволяють модифікувати біологічні інсектициди для втілення конкретних рішень конкретних проблем.

Один з таких агентів, *Bacillus thuringiensis* (Bt), є ефективним інсектицидним агентом і, як такий, широко використовується комерційно. Фактично, інсектицидний агент бактерії Bt являє собою білок, який має таку обмежену токсичність, що його можна використовувати на харчових культурах людини в день збирання врожаю. Для організмів, які не є мішенню, токсин Bt є нетоксичним білком, що легко засвоюється.

Іншим відомим класом біологічних агентів захисту від комах є деякі роди нематод, котрі, як відомо, є векторами переносу бактеріальних симбіонтів, що вбивають комах. Нематоди, що містять інсектицидні бактерії, заселяють личинки комах. Потім ці бактерії вбивають личинки. Нематоди розмножуються у вбитих личинках. Потомство нематод поїдає вбитих личинок зсередини. Одержане у такий спосіб потомство нематод, що містить бактерії, може потім заселяти інші личинки.

У минулому інсектицидних нематод у родах *Steinernema* та *Heterorhabditis* використовували як агентів захисту від комах. Певно, кожен рід нематод є жителем для конкретного виду бактерій. У нематодах роду *Heterorhabditis* симбіотичною бактерією є *Photorhabdus luminescens*.

Хоча ці нематоди є ефективними агентами захисту від комах, зараз важкою, дорогою й неефективною справою є одержання, підтримання та розповсюдження нематод для боротьби з комахами.

У цій галузі було відомо, що з *Photorhabdus luminescens* можна виділити інсектицидний токсин, який є активним лише при введенні у личинки лускокрилих та твердокрилих комах. Це робило неможливим ефективне застосування інсектицидних властивостей цих нематод чи їх бактеріального симбіонту. Корисним був би більш практичний, менш трудомісткий, такий, що охоплює великі простори, спосіб доставляння інсектицидного токсину, який зберігав би його біологічні властивості після доставки. Вельми бажаним було б відкриття токсинів з пероральною активністю, що їх продукує рід *Photorhabdus*. Виділення й застосування цих токсинів є бажаним завдяки їх ефективності. До відкриттів, що їх було здійснено заявниками, такі токсини не було виділено або охарактеризовано.

Нативні токсини являють собою білкові комплекси, що їх продукують та секретують бактеріальні клітини роду *Photorhabdus*, які ростуть, та увагу привертають білки, що їх продукує вид *Photorhabdus luminescens*. Білкові комплекси з розміром молекул приблизно 1000кДа можуть бути поділені аналізом за допомогою електрофорезу у ПААГ-ДСН на численні складові білки. Ці токсини не містять гемолізіну, ліпази, фосфоліпази С або нуклеазних активностей. Вони виявляють значну токсичність при введенні ряду комах.

Цей винахід забезпечує інсектицидний білок, що його може бути легко введено, а також експресію токсину у гетерологічній системі.

Цей винахід забезпечує також спосіб доставляння інсектицидних токсинів, які є функціонально активними та ефективними проти багатьох рядів комах.

Цілі, переваги й ознаки цього винаходу стануть очевидними з подальшого опису.

Стислий опис малюнків

Фіг.1 - ілюстрація збігу клонованих ізолятів ДНК, що використовуються як частина послідовності генів для токсину цього винаходу.

Фіг.2 - карта трьох плазмид, що використовуються у процесі секвенування.

Фіг.3 - карта, що ілюструє взаємовідношення кількох часткових фрагментів ДНК.

Фіг.4 - ілюстрація аналізу гомології між білковими послідовностями білків TcbA_{ij} та TcaB_{ij}.

Фіг.5 - фенотипна карта штамів *Photorhabdus*. Спорідненість штамів *Photorhabdus* визначали за допомогою гер-PCR. Верхня частина Фіг.5 вимірює процентну схожість штамів на основі бальної оцінки продуктів гер-PCR (тобто, 0,0 [немає схожості] - 1,0 [100% схожості]). На правій осі цифри й літери означають різні штам, які було випробувано; 14=W-14, Hm=Hm, H9=H9, 7=WX-7, 1=WX-1, 2=WX-2, 88=HP88, NC-1=NC-1, 4=WX-4, 9=WX-9, 8=WX-8, 10=WX-10, WIR=WIR, 3=WX-3, 11=WX-11, 5=WX-5, 6=WX-6, 12=WX-12, 14=WX-14, 15=WX-15, Hb=Hb, B2=B2, 48-52=ATCC43948-ATCC43952. Вертикальні лінії, що поділяють горизонтальні лінії, вказують ступінь спорідненості (що його зчитують з перетинання, що екстраполюється, вертикальної лінії з верхньою віссю) між штамми або групами штамів при основі горизонтальних ліній (наприклад, штам W-14 приблизно на 60% є подібним до штаму H9 і Hm).

Фіг.6 - ілюстрація геномних карт штаму W-14.

Цей винахід спрямовано на відкриття унікального класу інсектицидних білкових токсинів з виду *Photorhabdus*, що мають пероральну токсичність в відношенні до комах. Унікальною ознакою *Photorhabdus* є його біолюмінесценція. *Photorhabdus* може бути виділено з різноманітних джерел. Одним з таких джерел є нематоди, більш конкретно - нематоди виду *Heterorhabditis*. Інше подібне джерело виділено з клінічних проб з ран людини, [див. Farmer et al., 1989, J. Clin. Microbiol., 27pp., 1594-1600]. Ці сапрофітні штам депоновано в [American Type Culture Collection (Rockville, MD) ATCC №№43948, 43949, 43950, 43951 і 43952] та включено сюди як посилання. Можливо, що інші джерела могли б захоплювати бактерії *Photorhabdus*, які продукують інсектицидні токсини. Такі джерела в довідці могли б бути або такими, що живуть у землі чи на землі, або водними.

Рід *Photorhabdus* таксономічно визначається як член родини Enterobacteriaceae, хоча він має деякі ознаки, атипичні для цієї родини. Наприклад, штам цього виду не відновлюють нітрати, продукують жовтий та червоний пігмент та мають таку ознаку, як біолюмінесценція. Ця остання ознака не є відомою для інших членів родини Enterobacteriaceae. *Photorhabdus* тільки недавно було описано як вид, що відрізняється від *Xenorhabdus* [Boemare et al., 1993, Int. J. Syst. Bacteriol., 43, 249-255]. Ця диференціація заснована на дослідженнях гібридизації ДНК-ДНК, фенотипічних відмінностях (наприклад, наявності (*Photorhabdus*) або відсутності (*Xenorhabdus*) каталази та біолюмінесценції) та родині живителя-нематоди (*Xenorhabdus*; *Steinernematode*, *Photorhabdus*; *Heterorhabditidae*). Порівняльні аналізи жирних кислот у клітинах [Janse et al., 1990, Lett. Appl. Microbiol., 10, 131-135; Suzuki et al., 1990, J. Gen. Appl. Microbiol., 36, 393-401] підтверджують диференціацію *Photorhabdus* від *Xenorhabdus*.

Щоб встановити, що колекція штамів, яку описано тут, складається з штамів *Photorhabdus*, ці штам характеризували на основі ознак, що їх можна впізнати, які визначають *Photorhabdus* та відрізняють його від інших Enterobacteriaceae і від видів *Xenorhabdus* [Farmer, 1984, Bergey's Manual of Systemic Bacteriology, vol.1, pp.510-511; Akhurst and Boemare, 1988, J. Gen. Microbiol., 134, pp.1835-1845; Boemare et al., 1993, Int. J. Syst. Bacteriol., 43, pp.249-255], що їх включено сюди як посилання). Було досліджено такі ознаки: грамнегативні палички, розмір організму, пігментація колоній, тіла включення, наявність каталази, спроможність відновлювати нітрати, біолюмінесценція, зростання на елективних середовищах, температура росту, виживання за анаеробних умов та рухливості. Аналіз жирних кислот використовували, щоб підтвердити те, що всі ці штам належать до одного роду *Photorhabdus*.

Зараз бактеріальний рід *Photorhabdus* складається з єдиного визначеного виду, *Photorhabdus luminescens* [ATCC Type strain №29999, Poinar et al., 1977, Nematologica, 23, 97-102]. Численні споріднені штам було описано в літературі [наприклад, Akhurst et al., 1988, J. gen. Microbiol., 134, 1835-1845; Boemare et al., 1993, Int. J. Syst. Bacteriol., 43, pp.249-255; Putz et al., 1990, Appl. Environ. Microbiol., 56, 181-186]. Численні штам *Photorhabdus* було описано тут. Ці штам наведено в списку таблиці 18 у Прикладах. Оскільки зараз є тільки один вид (*luminescens*), який визначено у роді *Photorhabdus*, ознаки виду *luminescens* використали для характеристики штамів, що їх описано тут. Не можна не передбачити, що в майбутньому можуть бути виявлені інші види *Photorhabdus*, які матимуть деякі характерні ознаки виду *luminescens*, а також деякі властивості, що відрізняються, які зараз не визначено як ознаку *Photorhabdus luminescens*. Однак, обсяг цього винаходу охоплює будь-які види або штам *Photorhabdus*, що продукують білки, які мають функціональну активність як агенти боротьби з комахами, незалежно від інших ознак та відмінних властивостей.

Крім того, як показано тут, бактерії роду *Photorhabdus* продукують білки, які мають функціональну активність, що її визначено тут. Особливий інтерес представляють білки, які продукує вид *Photorhabdus luminescens*. Цей винахід не повинен обмежуватися штамми, які описано в ньому. Ці штам ілюструють вперше, що білки, які продукують різноманітні ізоляти *Photorhabdus*, є токсичними при обробленні ними комах. Таким чином, до описаного тут винаходу належать штам, які заявлено тут, і будь-які їх мутанти, а також будь-які штам або види роду *Photorhabdus*, що мають описану тут функціональну активність.

Тут використано кілька термінів, які мають визначене значення:

Під "функціональною активністю" тут розуміють те, що білкові токсини діють як агенти боротьби з комахами, у тому значенні, що вони є активними перорально або виявляють токсичну дію чи здатні порушувати або погіршувати живлення, що може спричинювати або не спричинювати загибель комах. При вступі комах в контакт з ефективною кількістю токсину, що поставляється через експресію трансгенної рослини, через приготовані білкові композиції, білкові композиції, які розприскують, матрикс принади або іншу систему доставки, результатами, як правило, є загибель комах, або комах не поїдають джерело, яке робить ці токсини доступними для комах.

Термін "токсин *Photorhabdus*" означає будь-який білок, що його продукує штам мікроорганізму *Photorhabdus*, який має функціональну активність у відношенні до комах, причому токсин *Photorhabdus* може бути приготовано у вигляді композиції, що її розприскують, експресовано трансгенною рослиною, приготовано у вигляді матриксу

принади, доставлятися за допомогою бакуловірусу або за допомогою будь-якого іншого прийнятного живителя або будь-якої іншої системи доставки.

Білкові токсини, які обговорюються тут, як правило, називають "інсектицидами". Під інсектицидами тут розуміють те, що ці білкові токсини мають "функціональну активність", визначену тут раніше, та використовуються як агенти боротьби з комахами.

Під терміном "олігонуклеотиди" розуміють макромолекулу, що складається з короткого ланцюга нуклеотидів або РНК, чи ДНК. Такою довжиною може бути, принаймні, один нуклеотид, але, як правило, довжина знаходиться у діапазоні від приблизно 10 до приблизно 12 нуклеотидів. Визначення довжини олігонуклеотиду є добре відомим для фахівців у цій галузі та не є обмеженням у цьому винаході. Таким чином, олігонуклеотиди можуть мати менше 10 або більше 12 нуклеотидів.

Термін "токсичний" або "токсичність", як його застосовано тут, означає, що токсини, які продукує *Photobhabdus*, мають визначену тут "функціональну активність".

Термін "генетичний матеріал", що використовується тут, стосується до генів, нуклеїнової кислоти, ДНК або РНК.

Ферментаційні бульйони від вибраних штамів, які описано у Таблиці 18, використовували для визначення: широти продукування інсектицидних токсинів видом *Photobhabdus*, інсектицидного спектру цих токсинів, і для забезпечення вихідного матеріалу для очищення комплексів токсинів. Було показано, що штами, які охарактеризовано тут, мають пероральну токсичність для численних рядів комах. Такі ряди комах включають (але не обмежуються ними) *Coleoptera*, *Homoptera*, *Lepidoptera*, *Diptera*, *Acarina*, *Hymenoptera* та *Dictyoptera*.

Як і у випадку інших бактеріальних токсинів, швидкість мутації бактерій у популяції спричинює утворення багатьох споріднених токсинів, які трохи відрізняються за їх послідовністю. Токсини, що описано тут, являють собою такі токсини, що утворюють білкові комплекси, токсичні для різних комах, які експонуються із ними, як це описано тут. Більш прийнятно, ці токсини є активними проти *Lepidoptera*, *Coleoptera*, *Homoptera*, *Diptera*, *Hymenoptera*, *Dictyoptera* та *Acarina*. Цей винахід стосується отримання білкових токсинів, гомологічних білковим токсинам, що їх продукують описані тут штами та будь-які утворені з них штами, а також будь-яких білкових токсинів, продукованих *Photobhabdus*. Ці гомологічні білки можуть відрізнятися за їх послідовністю, але вони не відрізняються за функцією від описаних тут токсинів. Розуміється, що гомологічні токсини включають білкові комплекси з розміром між 300кДа та 2000кДа і складаються, принаймні, з двох (2) субодиниць, причому субодиниця являє собою пептид, що може бути таким самим, як друга субодиниця, або відрізнятися від неї. Різноманітні білкові субодиниці були ідентифіковано та наведено в Прикладах цього винаходу. Як правило, субодиниці білка мають розмір між приблизно 18кДа та приблизно 230кДа; між приблизно 160кДа та приблизно 230кДа; 100кДа - 160кДа; приблизно 80кДа - приблизно 100кДа; і приблизно 50кДа - приблизно 80кДа.

Як обговорювалось вище, деякі штам *Photobhabdus* може бути виділено з нематод. Деякі нематод, подовжені циліндричні паразитичні черв'яки типу *Nematoda*, розвинули здатність використовувати личинки комах як більш прийнятне середовище для росту. Личинки комах забезпечують джерело їжі для нематод, що ростуть, і середовище, де вони розмножуються. Драматичним ефектом після інвазії личинок деякими нематодами є загибель личинок. Смерть личинок зумовлена наявністю (у деяких нематодах) бактерій, що продукують інсектицидний токсин, який зупиняє зростання личинок та інгібує активність живлення.

Цікаво, що, по-видимому, кожен рід паразитичних нематод з комахами є жителем певного виду бактерії, який є унікально пристосованим для симбіотичного росту з цією нематою. Протягом часу, що минув з початку цього дослідження, назву бактеріального роду *Xenorhabdus* було змінено на *Xenorhabdus* та *Photobhabdus*. Бактерії роду *Photobhabdus* охарактеризовано як симбіонт нематод *Heterorhabditis*, тоді як види *Xenorhabdus* є симбіонтами видів *Steinernema*. Цю модифікацію в номенклатурі відображено у цій заявці, але модифікація в номенклатурі ніяк не змінює обсягу описаного тут винаходу.

Пептиди та гени, які описано тут, називаються відповідно до основних вказівок, що їх опубліковано недавно в [Journal of Bacteriology "Instructions to Authors", p.i-xii (Jan. 1996)], які включено тут як посилання. Названі нижче пептиди й гени було виділено з штаму W-14 *Photobhabdus*.

Номенклатура пептидів /генів комплексу токсинів (Tc)

Назва пептиду	Назва гена	Патентований SEQ ID №
Геномний район tca		
TcaA	tcaA	12
TcaA _{iii}	tcaA	4
TcaB _i	tcaB	3 (19, 20)
TcaB _{ii}	tcaB	5
TcaC	tcaC	2
Геномний район tcb		
TcbA	tcbA	16
TcbA _i	tcbA	(пропептид)
TcbA _{ii}	tcbA	1 (21, 22, 23, 24)
TcbA _{iii}	tcbA	40
Геномний район tcc		
TccA	tccA	8

TccB	tccB	7
Геномний район tcd		
TcdA _i	tcdA	(пропептид)
TcdA _{ii}	tcdA	13, (38, 39, 17, 18)
TcdA _{iii}	tcdA	41, (42, 43)
TcdB	tcdB	14

5

(послідовність у дужках вказує внутрішню амінокислотну послідовність, яку одержано на основі продуктів розщеплення трипсином)

Наведені вище послідовності згруповано на основі району геному. Ген tcbA експресували у *E. coli* у вигляді двох білкових фрагментів TcbA та TcbA_{iii}, як показано у Прикладах. Може бути вигідним мати протеолітичне розрізування деяких послідовностей для одержання більш високої активності токсинів для комерційних трансгенних прикладень.

Токсини, що їх описано тут, є цілковито унікальними у тому розумінні, що ці токсини мають функціональну активність, яка є ключовою для розроблення стратегії боротьби з комахами. У розробці стратегії боротьби з комахами можна уповільнити чи обійти процес деградації білка шляхом введення білка безпосередньо у організм, обминаючи травний тракт. У такому разі білок, який введено в організм, буде зберігати його функцію до тих пір, поки він не денатурується, неспецифічно зруйнується чи його буде еліміновано імунною системою у вищих організмах. Ін'єкція у комах інсектицидного токсину має потенційне прикладення лише у лабораторії і, крім того, тільки у разі великих комах, які легко ін'єкувати. Спостереження, що інсектицидні білкові токсини, що їх описано тут, виявляють їх токсичну активність після проковтування через рот або контакту з токсинами, дозволяє розробити стратегію боротьби з комахами, яка ґрунтується лише на можливості включення цих білкових токсинів до раціону комах. Така стратегія веде до створення принад для комах.

Токсини *Photorhabdus* можна вводити комахам у очищеному вигляді. Токсини можна також поставляти кількістю від приблизно 1 до приблизно 100мг/л бульйону. Кількість може змінюватись залежно від умов приготування, умов джерела інокуляту, способів виділення токсину тощо. Токсини можна вводити у вигляді білка секреції ексудату чи клітинного білка, який спочатку було експресовано у гетерологічному прокаріотичному чи еукаріотичному живителі. Живителями, у який експресуються білки, є, як правило, бактерії. Еукаріотичними живителями могли би бути (але не обмежуються ними) рослини, комахи й дріжджі. Альтернативно, ці токсини можуть продукуватись у бактеріях чи трансгенних рослинах у полі або комах з використанням бакуловірусного вектора. Як правило, ці токсини вводять комасі шляхом включення одного чи кількох токсинів у їжу комах.

Повна летальність щодо комах, які живляться, є застосовуваною, але не є потрібною для досягнення корисної токсичності. Якщо комахи уникають токсину або припиняють житись, це уникання їжі буде вигідним у деяких прикладеннях, навіть якщо ці ефекти є сублетальними. Наприклад, якщо бажаними є стійкі до комах трансгенні сільськогосподарські рослини, небажання комах житись на цих рослинах так само цінно, як і летальна токсичність для цих комах, оскільки кінцевою метою є захист рослин, а не вбивання комах.

Існує багато інших шляхів того, як токсини може бути введено у їжу комах. Наприклад, можна підробити джерело їжі личинок за допомогою токсичного білка шляхом обприскування їжі розчином білка, як це описано тут. Альтернативно, очищений білок можна ввести за допомогою генної інженерії у бактерію, нешкідливу у інших відношеннях, яка могла би потім рости у культурі і або наноситись на джерело їжі, або приміщуватись у ґрунт у тій зоні, де бажаним є винищення комах. Також можна було б ввести одержаний за допомогою генної інженерії білок безпосередньо у джерело їжі комах. Наприклад, основним джерелом живлення багатьох личинок комах є рослинний матеріал.

Внаслідок включення генетичного матеріалу, який кодує інсектицидні властивості токсинів *Photorhabdus*, до геному рослини, що його поїдає конкретна комаха-шкідник, доросла комаха або личинки гинуть після живлення цією рослиною. Численні члени однодольних та дводольних родів було трансформовано. Трансгенні агрономічні культури, а також фрукти й овочі, являють собою комерційний інтерес. Такі сільськогосподарські культури включають (але не обмежуються ними) кукурудзу, рис, сою, сапола, соняшник, люцерну, сорго, пшеницю, бавовник, земляний горіх, томати, картоплю тощо. Існує декілька способів для введення чужорідного генетичного матеріалу у клітини рослин та для одержання рослин, що стабільно зберігають та експресують ген, який введено. Такі способи включають прискорення генетичного матеріалу, який нанесено на мікрочастки, для введення безпосередньо у клітини [U.S. Patents 4945050 to Cornell та 5141131 to Dow Elanco]. Рослини може бути трансформовано відповідно до технології з використанням *Agrobacterium*, [див. U.S. Patent 5177010 to University of Toledo, 5104310 to Texas A and M, European Patent Application 0131624B1, European Patent Applications 120516, 159418B1 та 176112 to Schilperoot, U.S. Patents 5149645, 5469976, 5464763 та 4940838 та 4693976 to Schilperoot, European Patent Applications 116718, 290799, 320500 to MaxPlanck, European Patent Applications 604662 and 627752 to Japan Tobacco, European Patent Applications 0267159 and 0292435, and U.S. Patent 5231019 to Ciba Geigy, U.S. Patents 5463174 та 4762785 to Calgene та U.S. Patents 5004863 і 5159135 to Agracetus]. Інші технології трансформацію включають "whiskers"-технологію, [див. U.S. Patents 5302523 та 5464765 to Zeneca]. Технологію електропорації також використовували для трансформації рослин, [див. WO 87/06614 to Boyce Thompson Institute, 5472869 та 5384253 to Dekalb, WO 9209696 і WO 9321335 to PGS]. Усі ці патенти й публікації щодо трансформації включено як посилання. Окрім численних технологій для трансформації рослин, можна такою варіювати тип тканини, яку приводять до контакту з чужорідними генами. Така тканина може включати (але не обмежуватись ними) ембріогенну тканину, калусну тканину типів I та II, гіпокотиль, меристему тощо. Майже усі тканини рослини може бути трансформовано під час дедиференціації з

використанням відповідних способів, що відомі фахівцям у цій галузі.

Іншою змінною є вибір маркера, що його селектують. Більш прийнятний конкретний маркер визначають за розсудом фахівця, але можна використовувати будь який з наведених далі маркерів, які селектують, разом з будь-яким іншим геном, який не названо тут, що він міг би функціонувати як маркер, який селектують. Такі маркери, які селектують, включають (але не обмежуються ними) ген аміноглікозилфосфотрансферази транспозону Tn5 (Aph II), який кодує стійкість до антибіотиків канаміцину, неомицину та G418, а також гени, що кодують стійкість чи толерантність до гліфозату, гігроміцину, метотрексату, фосфінотрицину (біалофосу), імідазолінонів, сульфонілсечовин та триазопіримідинових гербіцидів, таких як хлорсульфон, бромоксініл, далапон тощо.

Крім маркера, який селектують, може бути бажаним використання репортерного гена. У деяких випадках репортерний ген може бути використано без маркера, який селектують. Репортерні гени являють собою гени, які, як правило, не є наявними чи не експресуються у організмі- або тканині-реципієнті. Репортерний ген часто кодує білок, що забезпечує якусь фенотипічну зміну чи ферментативну властивість. Приклади таких генів наведено у [K. Weising et al., Ann. Rev. Genetics, 22, 421 (1988)], що включено тут як посилання. Більш прийнятним репортерним геном є ген глюкуронідази (GUS).

Незалежно від способу трансформації ген більш прийнятно включають до вектора переносу гена, який є пристосованим для експресії токсинів *Photobacterium* у клітині рослини шляхом включення до цього вектору рослинного промотору. Крім промоторів рослин можна використовувати ефективно промотори з різноманітних джерел у клітинах рослин для експресування чужорідних генів. Наприклад, можна використовувати промотори бактеріального походження, такі як промотор октопінсинтази, промотор нопалінсинтази, промотор маннопінсинтази; промотори вірусного походження, такі як промотор вірусу тютюнової мозаїки (35S та 19S) тощо. Промотори рослин включають (але не обмежуються ними) промотор малої субодиниці (MC) рибулозо-1,6-бісфосфат(РБФ)карбоксилази, промотор бета-конгліциніну, промотор фазеоліну, промотор ADH, промотори теплового шоку та тканиноспецифічні промотори. Промотори можуть також містити деякі елементи енхансерної послідовності, які можуть покращувати ефективність транскрипції. Типові енхансери включають (але не обмежуються ними) Adh-інтрон I та Adh-інтрон 6. Може бути використано й конститутивні промотори. Конститутивні промотори регулюють неперервну експресію генів в усіх типах клітин та у будь-який час (наприклад, актину, убіквітину, CaMV 35S). Тканиноспецифічні промотори відповідають за експресію генів у специфічних типах клітин чи тканин, таких як листя або насіння (наприклад, зеїну, олеозіну, напіну, ACP), і ці промотори можна також використовувати. Можна також застосовувати промотори, що є активними протягом певної стадії розвитку рослини, а також активними у тканинах рослин та їх органах. Прикладами таких промоторів є (але не обмежуються ними) специфічні для пилку, специфічні для ембріону, специфічні для кукурудзяного "шовку", специфічні для волокон бавовнику, специфічні для коріння, специфічні для ендосперму насіння промотори тощо.

За певних обставин може бути бажаним застосування індукованого промотору. Індукований промотор відповідає за експресію генів у відповідь на специфічний сигнал, такий як: фізичний фактор (гени теплового шоку); світло (РБФ-карбоксилаза); гормон (Em); метаболіти й стрес. Можуть використовуватись інші бажані елементи транскрипції та трансляції. У цій галузі відомі численні специфічні для рослин вектори переносу генів.

Крім того, відомо, що для одержання високої експресії бактеріальних генів у рослинах краще перебудувати бактеріальні гени таким чином, щоб вони більш ефективно експресувались у цитоплазмі рослин. Кукурудза є однією з таких рослин, для яких більш прийнятним є реконструювання бактеріального гена (генів) перед трансформацією для підвищення рівня експресії токсину у цій рослині. Однією з причин такої перебудови генної інженерії є низький вміст G+C нативного бактеріального гена (генів) (та подальше відхилення у бік високого вмісту A+T). Це спричинює утворення послідовностей, що імітують або дублюють регуляторні послідовності генів рослин, які, як відомо, вельми багаті на A+T. Наявність деякої кількості багатих на A+T послідовностей у ДНК гена (генів), які вводять у рослини (наприклад, районів TATA-боксу, що у нормі є наявним у промоторах генів), може спричинювати аберантну транскрипцію цього гена (генів). З іншого боку, присутність інших регуляторних послідовностей, наявних у мРНК, що транскрибується (наприклад, послідовностей сигналу поліаденілування (AAUAAA), або послідовностей, які є комплементарними до малих ядерних РНК, що беруть участь у сплайсінгу пре-мРНК), може спричинювати нестабільність РНК. Таким чином, однією з цілей у конструюванні бактеріального гена (генів), що його (їх) перебудовують, та який, більш прийнятно, називають оптимізованим для рослин геном (генами), є утворення послідовності ДНК, що має високий вміст G+C, та, більш прийнятно, послідовності, близької до послідовності генів рослин, які кодують метаболічні ферменти. Іншою метою побудови оптимізованого для рослини гена (генів) є утворення послідовності ДНК, яка не тільки має більш високий вміст G+C, але її має бути виготовлено, шляхом модифікації змін послідовності, таким чином, щоб не заважати трансляції.

Прикладом рослини, що має високий вміст G+C, є кукурудза. Наведена нижче таблиця ілюструє, наскільки високим є вміст G+C у кукурудзі. Вважають, що, як і у кукурудзі, вміст G+C у інших рослинах є теж високим.

Таблиця 1		
Зведення вмісту G+C кодуючих білок районів генів кукурудзи		
Клас білка ^a	Діапазон G+C, %	Середнє G+C ^b , %
Метаболічні ферменти (40)	44.4-75.3	59.0 (8.0)
Запасні білки		

Група I (23)	46.0-51.9	48.1 (1.3)
Група II (13)	60.4-74.3	67.5 (3.2)
Група I+II (36)	46.0-74.3	55.1 (9.6) ^c
Структурні білки (18)	48.6-70.5	63.6 (6.7)
Регуляторні білки (5)	57.2-68.9	62.0 (4.9)
Білки, які не охарактеризовано (9)	41.5-70.3	64.3 (7.2)
Усі білки (108)	44.4-75.3	60.8 (5.2)
^a У дужках наведено кількість генів у класі.		
^b У дужках наведено стандартне відхилення.		
^c Середнє об'єднаних груп, що не враховується у розрахунках загального середнього.		

Для даних у Таблиці 1 кодуючі райони генів вибирали з GenBank (Release 71), та склад основ розраховували, використовуючи програму MacVector™ (IBI, New Haven, CT). Послідовності нітронів не брали до уваги у цих розрахунках. Послідовності генів запасних білків груп I та II розрізняли за помітною відмінністю у складі основ.

Завдяки гнучкості, що її надавав надмір генетичного коду (тобто тим, що деякі амінокислоти кодуються більш ніж одним кодоном), еволюція геномів різних організмів або класів організмів спричинила диференційне використання надмірних кодонів. Це таксономічне "зміщення кодонів" відображено у середньому складі основ районів, що кодують білки. Наприклад, організми з відносно низьким вмістом G+C використовують кодони, що мають А чи Т у третій позиції надмірних кодонів, тоді як кодони, що мають високий вміст G+C, використовують кодони, які мають G чи С у третій позиції.

Припускають, що наявність "мінорних" кодонів у мРНК гена може зменшувати абсолютну швидкість трансляції цієї мРНК, особливо, коли відносний достаток заряджених тРНК, які відповідають мінорному кодону, є низьким. Продовженням цього є те, що зменшення швидкості трансляції окремими мінорними кодонами було би, принаймні, адитивно для численних мінорних кодонів. Таким чином, мРНК, що мають високий відносний вміст мінорних кодонів, матимуть, відповідно, більш низькі швидкості трансляції. Ця швидкість відбивається у синтезі низьких рівнів білка, що кодується.

Для перебудови бактеріального гена (генів) визначають зміщення кодонів рослини. Зміщення кодонів являє собою статистичний розподіл кодонів, які рослина використовує для кодування її білків. Визначивши це зміщення, знаходять частоту стрічання цих кодонів у гені (генах) що являє інтерес. Треба визначити первинні, більш прийнятні для рослини кодони, а також другий та третій вибір більш прийнятних кодонів. Амінокислотна послідовність білка, що представляє інтерес, зворотно трансклюється таким чином, що одержана послідовність нуклеїнової кислоти кодує той самий білок, що й нативний бактеріальний ген, але ця отримана послідовність нуклеїнової кислоти відповідає першим більш прийнятним кодонам цільової рослини. Нову послідовність аналізують на сайти рестриктаз, які може бути створено шляхом цієї модифікації. Ідентифіковані сайти модифікують далі заміною кодонів більш прийнятними кодонами другого й третього вибору. Іншими сайтами у цій послідовності, що можуть впливати на транскрипцію чи трансляцію цільового гена, є місця з'єднання 5' або 3' екзон: інтрон, сигнали приєднання полі А чи сигнали термінації РНК-полімерази. Цю послідовність аналізують далі та модифікують, щоб зменшити частоту стрічання дуплетів ТА або GC. Крім цих дуплетів, блоки послідовностей з G або C, які мають більше приблизно 4 залишків, можуть впливати на транскрипцію цієї послідовності. Тому ці блоки також модифікують, замінюючи кодони першого чи другого вибору тощо наступним більш прийнятним кодоном вибору. Більш прийнятним є те, щоб оптимізований для рослин ген (гени) містив приблизно 63% кодонів першого вибору, приблизно 22% - приблизно 37% кодонів другого вибору та 15% - 0% кодонів третього вибору, причому загальний процент дорівнює 100%. Найбільш прийнятний оптимізований для рослин ген (гени) містить близько 63% кодонів першого вибору, принаймні, близько 22% кодонів другого вибору, близько 7,5% кодонів третього вибору та близько 7,5% кодонів четвертого вибору, причому загальний процент дорівнює 100%. Спосіб, який описано вище, дає змогу фахівцеві у цій галузі модифікувати ген (гени), чужорідні для конкретної рослини таким чином, що ці гени оптимально експресуються у рослинах. Спосіб ілюстровано додатково у попередній заявці, що очікує розглядання, [U.S. 60/005405, яку подано 13 жовтня 1995 року], що її включено сюди як посилання.

Такий чином, для побудови оптимізованого для рослини гена (генів) амінокислотну послідовність токсинів зворотно трансклюють (переводять) у послідовність ДНК, що використовує ненадмірний генетичний код, який визначено з таблиці зміщення кодонів, яку компільовано для ДНК-послідовності гена для конкретної трансформованої рослини. Одержану послідовність ДНК, яка є цілком гомогенною щодо використання кодонів, модифікують далі для утворення послідовності ДНК, яка, крім того, що має високий ступінь різноманіття кодонів, містить також стратегічно розташовані сайти впізнавання рестриктаз, бажаний склад основ та не містить послідовностей, які можуть перешкоджати транскрипції цього гена чи трансляції продукту мРНК.

Теоретично припускається, що бактеріальні гени можуть легше експресуватись у рослинах, якщо ці бактеріальні гени експресуються у пластидах. Таким чином, можна експресувати бактеріальні гени у рослинах, не оптимізуючи ці гени для експресії рослинами, та одержувати високий рівень експресії білка [див. U.S. Patents №№4762785, 5451513 та 5545817, що їх включено тут як посилання].

Одним з результатів з точки зору комерційного використання трансгенних рослин є управління стійкістю. Це особливо стосується токсинів *Bacillus thuringiensis*. Існують численні компанії, що комерційно використовують *Bacillus thuringiensis*, та було чимало занепокоєння щодо токсинів Bt, які стають стійкими. Однією з стратегій для управління стійкістю до комах може бути комбінування токсинів, які продукує *Photographus*, з токсинами, такими як Bt, вегетативними білками комах (Ciba Geigy) чи іншими токсинами. Такі комбінації можна було б

готувати для використання шляхом розприскування або вони могли би бути молекулярними комбінаціями. Рослини можна трансформувати генами *Photorhabdus*, що продукують токсини комах, та генами токсинів інших комах, таких як Bt.

[В European Patent Application 0400246A1] описано трансформацію 2 генів Bt у рослину, причому ці 2 гени могли би бути будь-якими іншими двома генами. Інший спосіб отримання трансгенної рослини, що містить більше, ніж один ген стійкості до комах, міг би складатись з тримання двох рослин, кожна з яких має ген стійкості до комах. Ці рослини можна було б піддати зворотному схрещуванню з використанням традиційного способу схрещування рослин для одержання рослини, що містить більше одного гена стійкості до комах.

Крім одержання трансформованої рослини, яка має оптимізований для рослини ген (гени), існують інші системи доставляння для випадків, коли бажаною є перебудова бактеріального гена (генів). За тим самим типом генетично сконструйований білковий токсин, що легко виділяється, злитий разом з молекулою, яка атрагує комах, як джерело живлення та інсектицидної активності токсину, може бути генетично сконструйовано та експресовано у бактеріях або в еукаріотичних клітинах за допомогою стандартних, добре відомих способів. Після очищення у лабораторії такий токсичний агент з вбудованою "принадою" може бути упаковано у стандартних пастках для комах.

Інша схема доставки являє собою включення генетичного матеріалу токсинів до бакуловірусного вектора. Бакуловіруси інфікують конкретних комах-живителів, у тому числі тих, яких поражають токсини *Photorhabdus*. Інфекційний бакуловірус, що несе експресійну конструкцію для токсинів *Photorhabdus*, міг би вводитись у зони зараження комахами для інтоксикації чи отруєння інфікованих комах.

Перенос інсектицидних властивостей вимагає, щоб послідовності нуклеїнових кислот, що кодують амінокислотні послідовності токсинів *Photorhabdus*, було інтегровано до вектора експресії білка, придатного для застосування у живителі, у якому цей вектор буде знаходитись. Одним з способів одержання послідовності нуклеїнової кислоти, що кодує білок з інсектицидними властивостями, є виділення нативного генетичного матеріалу, який продукує токсини *Photorhabdus*, з використанням інформації, що її розшифровано з амінокислотної послідовності токсину, більші частки якої наведено нижче. Як описано нижче, наведено також способи очищення білків, що відповідають за активність токсину.

Використовуючи дані про N-кінцеву амінокислотну послідовність, що їх наведено нижче, можна сконструювати олігонуклеотиди, комплементарні усім чи частині основ ДНК, які кодують перші амінокислоти токсину. Ці олігонуклеотиди можуть бути радіоактивно мічені та можуть використовуватись як молекулярні зонди для виділення генетичного матеріалу з геномної генетичної бібліотеки, яку побудовано з генетичного матеріалу, що його виділено з штамів *Photorhabdus*. Генетичну бібліотеку може бути клоновано у плазмидному, космидному, фаговому чи фагмидному векторах. Бібліотеку можна трансформувати у *Escherichia coli* та скринювати на утворення токсину трансформованими клітинами з використанням антитіл, які вироблено проти токсину, або прямими аналізами на токсичність по відношенню до комах.

Цей підхід вимагає отримання набору олігонуклеотидів, оскільки вироджений генетичний код дає змогу кодувати одну амінокислоту у ДНК будь-якої з кількох трьохнуклеотидних комбінацій. Наприклад, амінокислота аргінін може кодуватись триплетами CGA, CGC, CGG, CGT, AGA та AGG. Оскільки неможливо передбачити, який триплет використовується у цих позиціях у гені токсину, слід приготувати олігонуклеотиди з кожним потенційно характерним триплетом. Може виникнути потреба у більш, ніж одній молекулі ДНК, що відповідає субодиниці балка, для конструювання достатньої кількості олігонуклеотидних зондів для визначення усіх субодиниць білка, що необхідні для пероральної токсичності.

На основі амінокислотної послідовності очищеного білка генетичні матеріали, які відповідають за продукування токсинів, може бути легко виділено та клоновано, повністю або частково, у експресуючий вектор з використання будь-якого з кількох способів, що добре відомі фахівцям у галузі молекулярної біології. Типовим експресуючим вектором є ДНК-плазида, хоча передбачаються також і інші переносники, у тому числі (але не тільки), космиди, фагмиди й фаги. Крім ознак, потрібних чи бажаних для реплікації плазмид, таких як запал реплікації та стійкість до антибіотиків чи інша форма маркера, який селектують, наприклад, ген *bar* *Streptomyces hydropiscus* або *viridochromogenes*, вектори експресії білка часто потребують додатково касети експресії, яка включає цис-діючі послідовності, необхідні для транскрипції і трансляції цільового гена. Цис-діючі послідовності, необхідні для експресії в прокаріотах, відрізняються від подібних елементів, що потрібні в еукаріотах та рослинах.

Еукаріотична касета експресії вимагає наявності транскрипційного промотору проти ходу транскрипції (5') відносно цільового гена, району термінації транскрипції, такого як сайт приєднання полі-А, та місця зв'язування рибосом проти ходу транскрипції від першого кодону цільового гена. У бактеріальних клітинах застосовувані транскрипційні промотори, який можна було б включити до вектора, є промотор, що зв'язує РНК-полімеразу T7. Як описано вище, промотори ефективно підсилюють транскрипцію мРНК. Також проти ходу транскрипції (зліва) від цільового гена вектор може мати нуклеотидну послідовність, що кодує сигнальну послідовність, яка, як відомо, направляє ковалентно зв'язаний з нею білок до конкретного компартменту клітин-живителів, такий як клітинна поверхня.

Відомо, що віруси комах, або бакуловіруси, інфікують деяких комах та несприятливо діють на них. Дія цих вірусів на комах є повільною, та віруси не припиняють живлення комах. Тому віруси не вважаються такими, що можуть застосовуватись як агенти боротьби з комахами. Введення генів токсинів *Photorhabdus* у бакуловірусний вектор могло б стати ефективним засобом передавання токсинів з одночасним підвищенням летальності такого вірусу. Крім того, оскільки різні бакуловіруси є специфічними до різних комах, можна використовувати конкретний токсин для селективного націлювання на конкретних шкідників-комах. Особливо цінним вектором для

генів токсинів є вірус ядерного поліедрозу. Вектори-переносники, що використовують цей вірус, було описано у літературі та зараз їх може бути вибрано як вектори для перенесення чужорідних генів у комах. Рекомбінант вірус-ген токсину може бути сконструйовано у формі, що передається перорально. Бакуловіруси часто інфікують комах через слизову оболонку середньої кишки. Ген токсину, що його вбудовано за сильним промотором білка оболонки вірусу, буде експресуватись та швидко вбивати інфіковану комаху.

Крім вірусу комах або бакуловірусу чи системи доставки трансгенної рослини для білкових токсинів цього винаходу, ці білки може бути інкапсульовано, використовуючи технологію *Bacillus Thuringiensis*, таку як (але не тільки) [U.S. Patents №№4695455, 4695462, 4861595], які включено тут як посилання. Іншою системою доставки для білкових токсинів цього винаходу є введення цього білка у матрикс принади, яку можна було б потім використовувати у надземних та підземних ділянках пробування. Приклади такої технології наведено у (але не тільки) [PST Patent Application WO 93/23998], яку включено тут як посилання.

Як описано вище, може бути необхідною модифікація послідовності, що кодує цільовий білок, при експресії її у ненативному живителі, оскільки прийнятність кодонів інших живителів може відрізнятись від прийнятності кодонів *Photobacterium*. У такому разі трансляція може бути цілком неефективною у новому живителі, якщо не ввести компенсуючих модифікацій у кодуючу послідовність. Крім того, можуть бути бажаними модифікації амінокислотної послідовності, щоб уникнути інгібіторної перехресної реактивності з білками нового живителя, чи для вдосконалення інсектицидних властивостей білка у новому живителі. Генетично модифікований ген токсину може кодувати токсин, що виявляє, наприклад, підсилену чи послаблену токсичність, змінений розвиток стійкості до комах, змінену стабільність чи модифіковану специфічність щодо виду-мішені.

Крім генів *Photobacterium*, що кодують його токсини, обсяг цього винаходу включає споріднені послідовності нуклеїнових кислот, які кодують амінокислотні біополімери, гомологічні білкам токсинів, та які зберігають токсичну дію білків *Photobacterium* у видах комах після перорального проковтування.

Наприклад, токсини, що використовуються у цьому винаході, певно, спочатку інгібують живлення личинок перед настанням смерті. Шляхом маніпулювання послідовністю нуклеїнової кислоти токсинів *Photobacterium* або її регуляторними послідовностями генні інженери, приміщуючи ген токсину у рослини, могли би модулювати його силу або спосіб дії, наприклад, для зберігання інгібуючої живлення активності з одночасним виключенням абсолютної токсичності для личинок. Ця зміна могла би дати змогу трансформованій рослині виживати до збирання врожаю, не справляючи необов'язкового драматичного впливу на екосистему внаслідок знищення усіх ком ах-мішеней. Передбачається, що усі подібні модифікації гена, що кодує токсин, або білка, що його кодує цей ген, входять до обсягу цього винаходу.

Інші передбачені модифікації нуклеїнової кислоти включають приєднання направляючих послідовностей для направлення токсину у конкретні частини личинок комах для підсилення ефективності.

Штами ATCC 55397, 43948, 43950, 43951, 43952 було депоновано у [American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852 USA]. Дані про амінокислотну послідовність та нуклеотидну послідовність для нативного токсину W-14 (ATCC 55397) наведено нижче. Виділення геномної ДНК для токсинів з бактеріальних живителів також наведено тут як приклад.

У цій заявці використано й описано стандартні способи та способи молекулярної біології. Додаткову інформацію можна знайти у [Sambrook J., Fritsch, E.F., and Maniatis, T. (1989), *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press], яку включено тут як посилання.

У Прикладах використано такі аббревіатури:

Трис= трис(гідроксиметил)амінометан;

ДСН= додецилсульфат натрію;

ЕДТК= етилендіамінтетраоцтова кислота;

ІПТГ= ізопропілтіо-бета-галактозид;

Х-гал= 5-бром-4-хлор-3-індоліл-бета-D-галактозид;

ЦТАБ= цетилтриметиламонійбромід;

т.п.О.= тисячі пар основ;

dATP, dCTP, dGTP, dTTP, I= 2'-дезоксинуклеозид-5'-трифосфати аденіну, цитозину, гуаніну, тиміну та інозин, відповідно;

ATP= аденозин-5'-трифосфат.

Приклад 1

Очищення токсину з *P. luminescens* та демонстрація токсичності після пероральної доставки очищеного токсину

Інсектицидний токсин цього винаходу очищували з штаму W-14 *P. luminescens*, ATCC Accession №55397. Вихідні культури *P. luminescens* зберігали у чашках Петрі, що містили 2% Протеоза Пептон №3 (тобто PP3, Difco Laboratories, Detroit MI) у 1,5% агарі, інкубували при 25°C та пересіювали щотижня. Колонії первинної форми цих бактерій інокулювали у 200мл PP3-бульйону, доповненого 0,5% моностеаратом поліоксетиленсорбітану (Твін 60, Sigma Chemical Company, St. Louis, MO), в однілітровій колбі. Бульйонні культури вирощували протягом 72 годин при 30°C на ротаційному шейкері. Білки токсину здобували з культур, які було вирощено у присутності чи відсутності Твіну; однак, відсутність Твіну може впливати на форму бактерій, що їх вирощують, та профіль білків, що їх продукують бактерії. При відсутності Твіну може мати місце варіантне зміщення, оскільки молекулярна маса принаймні одного ідентифікованого токсину зміщується від приблизно 200кДа до приблизно 185кДа.

72-годинні культури центрифугували при 10000×g протягом 30 хвилин для видалення клітин та залишків клітин (дебрису). Фракцію супернатанту, що містить інсектицидну активність, зливали та доводили до 50mM

K_2HPO_4 шляхом додавання підходящого об'єму $1,0M K_2HPO_4$. pH доводили до 8,6, додавши гідроксид калію. Потім цю фракцію супернатанту змішували з DEAE-Sephacel (Pharmacia LKB Biotechnology), який було врівноважено $50mM K_2HPO_4$. Токсична активність адсорбувалась на DEAE. Потім цю суміш виливали у колонку $2,6 \times 40cm$ та промивали $50mM K_2HPO_4$ при кімнатній температурі при швидкості потоку $340ml/год.$ до тих пір, поки ефлюент не досягав постійної базової лінії поглинання УФ при $280nm$. Потім колонку промивали $150mM RC1$ до тих пір, поки ефлюент знову не досягав постійної базової лінії при $280nm$. Нарешті, колонку промивали $300mM KCl$ та збирали фракції.

Фракції, що містять токсин, об'єднували та стерилізували, пропускаючи крізь стерилізуючий фільтр, яким був мембранний фільтр з порами $0,2$ мікрон. Потім токсин концентрували й зрівноважували до $100mM KPO_4$, pH6,9, застосовуючи ультрафільтраційну мембрану з відсіканням молекулярної маси $100kDa$ при $4^\circ C$ (Centryprep 100, Amicon Division-W.R., Grace and Company). Пробу ($3ml$) концентрату токсину наносили на верхню частину гель-фільтраційної колонки (Pharmacia LKB Biotechnology). Елюючим буфером був $100mM KPO_4$, pH6,9, швидкість потоку $17ml/год.$, при $4^\circ C$. Ефлюент піддавали моніторингу при $280nm$.

Фракції збирали та тестували на токсичну активність. Токсичність хроматографічних фракцій випробували у біологічному тесті з використанням личинок *Manduca sexta*. Фракції або наносили на корм комах (корм з зародків пшениці для шовкопряда непарного, ICN Biochemical Division - ICN Biomedicals, Inc.), або вводили внутрішньогомощельною ін'єкцією проби 5 мкл через першу несправжню ніжку за допомогою голки калібру 30 . Вагу кожної личинки у групі, що одержувала токсин, реєстрували з інтервалами 24 години. Токсичність припускалась, якщо комаха припиняла їсти та вмирала у межах кількох днів поїдання обробленого корму для комах, або якщо смерть наставала у межах 24 годин після ін'єкції фракції.

Токсичні фракції об'єднували та концентрували за допомогою Centryprep-100 і потім аналізували за допомогою ВЕРХ з використанням гель-фільтраційної колонки $7,5mM \times 60cm$ TSK-GEL G-4000 SW з $100mM$ фосфатом калію, pH6,9, швидкість потоку елюючого буфера $0,4ml/хвил.$ Цей аналіз виявив, що білок токсину міститься у єдиному гострому піку, що його елюювали з колонки з часом утримування приблизно $33,6$ хвилини. Цей час утримування відповідає молекулярній масі $1000kDa$, яку було оцінено. Фракції піку збирали для подальшого очищення, а фракції, що не містили цього білка, викидали. Пік, що його елюювали з колонки ВЕРХ, поглинає УФ світло при 218 та $280nm$, але не поглинає світло при $405nm$. Було показано, що поглинання при $405nm$ є ознакою ксенорабдинових антибіотичних сполук.

Електрофорез об'єднаних фракцій піку у неденатуруючому агарозному гелі (Metaphor Agarose, FMC BioProducts) показав, що у піку наявні два білкових комплекси. Матеріал піку, який було забуферено у $50mM$ Трис- HCl , pH7,0, розділяли на $1,5\%$ агарозному концентруючому гелі, що його було забуферено $100mM$ Трис- HCl при pH7,0, та $1,9\%$ агарозному розділяючому гелі, який було забуферено $200mM$ Трис-борату при pH8,3, за стандартних буферних умов (анодний буфер $1M$ Трис- HCl , pH8,3; катодний буфер $0,025M$ Трис, $0,192M$ гліцин). Електрофорез проводили при $13mA$ постійного струму при $15^\circ C$ до тих пір, поки барвник-"свідок", феноловий червоний, не доходив кінця гелю. Дві білкові смуги візуалізували в агарозних гелях забарвленням Кумасі (Coomassie Brilliant blue).

Смугу, що переміщувалась повільніше, називали "білковою полоскою 1", а смугу, що переміщувалась швидше, називали "білковою полоскою 2". Дві білкові смуги були наявні приблизно у рівних кількостях. Забарвлені Кумасі агарозні гелі використовували як шаблон для точного вирізання двох білкових смуг з незабарвлених частин гелів. Вирізані ділянки, що містили ці білкові полоси, мацерували та додавали невелику кількість стерильної води. Для контролю також вирізали частину гелю, яка не містила білка, та обробляли її так само, як і ділянки гелю, що містили білок. Білок здобували з шматочків гелю за допомогою електрофорезу у $100mM$ Трис-боратному буфері, pH8,3, при 100 вольтях (постійна напруга) протягом двох годин. Альтернативно, білок пасивно елюювали з шматочків гелю додаванням рівного об'єму $50mM$ Трис- HCl , pH7,0 до шматочків гелю з подальшим інкубуванням при $30^\circ C$ протягом 16 годин. Це дозволяло білку дифундувати з гелю до буфера, який потім збирали.

Результати тестів токсичності для комах з використанням очищеного за допомогою ВЕРХ токсину (пік $33,6$ хвил.) та очищеного за допомогою агарозного гелю токсину продемонстрували токсичність цих екстрактів. Ін'єкція $1,5mkg$ ВЕРХ-очищеного білка вбиває у межах 24 годин. Обидві білкові смуги 1 та 2, здобуті з агарозних гелів пасивною елюцією або електроелюцією, були летальними при ін'єкції. Концентрація білка, визначена для цих проб була меншою за $50ng$ на личинку. Порівняння збільшення маси та смертності між групами личинок, які було ін'єктовано білковими смугами 1 та 2, показує, що білкова смуга 1 була більш токсичною при доставлянні шляхом ін'єкції.

При нанесенні ВЕРХ-очищеного токсину на корм для личинок при концентрації $7,5mkg$ на личинку відбувалось припинення збільшення маси личинки (досліджували 24 личинки). Личинки починали житись кормом, але після поїдання лише вельми малої частини корму, який було оброблено токсином, вони починали виявляти патологічні симптоми, що їх викликав токсин, та припиняли поїдати корм. Екскременти комах ставали безбарвними, та більшість личинок виявляла ознаки діареї. Значна смертність комах спостерігалась, коли кілька доз по $5mkg$ наносили на корм протягом періоду $7 - 10$ днів.

Одержана шляхом розділу на агарозі смуга 1 значно інгібувала збільшення маси личинок при дозі $200ng$ на личинку. Личинки, що отримували подібні концентрації білкової смуги 2 з кормом, не інгібувались та давали збільшення маси з тією самою швидкістю, що й контрольні личинки. Дванадцять личинок годували елююваним білком та 45 личинок годували шматочками агарози, що містили білок. Ці два набори даних показують, що білкова смуга 1 була перорально токсичною для *Manduca sexta*. У цьому експерименті, напевно, білкова смуга 2 не була токсичною для *Manduca sexta*.

Подальший аналіз білкових смуг 1 та 2 за допомогою електрофорезу у ППАГ-ДСН за денатуруючих умов показав, що кожна смуга складалась з кількох менших білкових субодиниць. Білки візуалізували забарвленням Кумасі з подальшим забарвленням сріблом для досягнення максимальної чутливості.

Білкові субодиниці у цих двох смугах були дуже схожими. Білкова смуга 1 містить 8 білкових субодиниць 25,1, 56,2, 60,8, 65,6, 166, 171, 184 та 208кДа. Білкова смуга 2 мала ідентичний профіль за винятком того, що не були наявними білки 25,1, 60,8 та 65,6кДа. Білки 56,2, 60,8, 65,6 та 184кДа були наявні у комплексі білкової смуги 1 у приблизно однакових концентраціях та складали 80% чи більше від загального вмісту білка цього комплексу.

Нативний ВЕРХ-очищений токсин характеризували далі таким чином. Токсин був термолабільним, оскільки після нагрівання до 60 °C протягом 15 хвилин він втрачав здатність вбивати чи інгібувати збільшення маси при ін'єкції або надання з кормом личинкам *M. sexta*. Було створено тести для виявлення ліпазної активності, активності фосфоліпази С, нуклеазної активності чи активності гемолізу еритроцитів, та їх проводили з очищеним токсином. Не було виявлено жодної з цих активностей. Проводили також тести антибіотичного зонального інгібування, та очищений токсин не інгібував ріст грамнегативних та грампозитивних бактерій, дріжджів або ниткоподібних грибів, а це свідчило про те, що він не є ксенорабдиновим антибіотиком.

Нативний очищений ВЕРХ токсин тестували на здатність вбивати комах, інших ніж *Manduca sexta*. У Таблиці 2 наведено список комах, що їх вбивав очищений ВЕРХ токсин *P. luminescens* у цьому дослідженні.

Таблиця 2			
Комахи, яких вбиває токсин <i>P. Luminescens</i>			
Звичайна назва	Ряд	Рід та вид	Спосіб доставки
Бражник	Lepidoptera	<i>Manduca sexta</i>	пероральний та ін'єкція
"Борошняний черв'як"	Coleoptera	<i>Tenebrio molitor</i>	пероральний
Фараонів мурашка	Hymenoptera	<i>Monomorium pharaonis</i>	пероральний
Прусак	Dictyoptera	<i>Blattella germanica</i>	пероральний та ін'єкція
Комар жовтогарячковий	Diptera	<i>Aedes aegypti</i>	пероральний

Приклад 2

Застосовність як інсектициду

Далі було охарактеризовано застосовність та токсичність *Photographus luminescens*. Культуральний бульйон штамму W-14 *Photographus luminescens* одержали у такий спосіб. Продукційним середовищем був 2% Bacto Proteose Peptone* Number 3 (PP3, Difco Laboratories, Detroit, Michigan) у деіонізованій Milli-Q* воді. Колби для культури для засівання, що містили 175мл середовища, приміщували у перегороджену на три частини по 500мл колбу з горлом Делонга, закривали Кап'ю та автоклавували протягом 20 хвилин при температурі 250°F (121,1°C). Продукційні колби містили 500мл у колбі на 2,8л, яку було перегороджено на три частини по 500мл, з горлом Делонга, закрито пробкою з кремнійорганічного пенопласту Shin-etsu. Їх автоклавували протягом 45 хвилин при 250°F (121,1°C). Культуру для засіву інкубували при 28 °C при 150об./хвил. у термостаті, що обертався та струшував, з ексцентриситетом 2 дюйми. Після 16 годин росту 1% культури для засівання приміщували у продукційну колбу та давали їм рости протягом 24 годин до збирання. Продуктування токсину відбувається, напевно, упродовж логарифмічної фази росту. Бульйон, що містив мікроби, переносили до центрифужової пляшки на 1л, та клітинну біомасу осаджували (30 хвилин при 2500об./хвил. при 4 °C [R.C.F.=~1600] HG-4L Rotor RC3Sorval centrifuge, Dupont, Wilmington, Delaware). Первинний бульйон охолоджували при 4°C протягом 8 - 16 годин та повторно центрифугували принаймні протягом 2 годин (за умов, що описано вище) для подальшого освітлення бульйону шляхом вилучення можливого мукополісахариду, який осаджується при стоянні. (Альтернативний спосіб обробки об'єднував обидві ці стадії та включав використання 16-годинного центрифугування для освітлення за описаних вище умов). Потім бульйон зберігали при 4 °C до біотесту або фільтрування.

Культуральний бульйон *Photographus* та білковий токсин (токсини), які було очищено з цього бульйону, виявили активність (смертність та/або інгібування росту, зменшене вилуплення дорослих особин) щодо ряду комах. Більш конкретно, було виявлено активність проти блішки довговусої (личинки та дорослих особин), колорадського жука та газонних червоподібних личинок хрущика японського, що є членами ряду *Coleoptera*. Інші члени ряду *Coleoptera* включають дротяників, квітогриза рапсового, земляних блошек, жуків, що поїдають насіння, та жуків-довгоносику. Також спостерігали активність проти цикадки айстр, що є членом ряду *Homoptera*. Інші члени *Homoptera* включають цикад, листоблішку грушеву, мідяницю яблуневу, червеців, білокрилок та spittle-жуків, а також численні специфічні у відношенні до живителя види тлі. Бульйон та очищені фракції є активними також проти совки малої, совки капустяної, совки іпсилон, совки тютюнової, точильника зернового, совки бавовняної та плодожерки яблуневий, що є членами ряду *Lepidoptera*. Іншими типовими членами цього ряду є справжня міль, вогнівка амбарна південна, метелики-листовійки, совка капустяна, совка бавовняна американська, мішечниця поденкоподібна, коконопряд кільчастий американський, лучний метелик та совка трав'яна. Активність також спостерігалася проти плодової мушки звичайної та личинок комарів, що є членами ряду *Diptera*. Іншими членами ряду *Diptera* є галіця горохова, муха морквяна, муха капустяна весняна, муха ріпи, муха цибульна, довгоніг, муха кімнатна та різні види комарів. Активність спостерігали проти мурашки-деревоотця, мурашки аргентинського, які є членами ряду, що включає також мурашку Ріхтера, мурашку кімнатного та мурашок малих.

Описані бульйон та фракція можуть застосовуватись для зменшення популяції комах та їх використовували у

способі інгібування популяція комах. Спосіб може включати внесення до осередку комах ефективної кількості, яка інгібує комах, описаного активного начала. Результати наведено в таблиці 3.

Активність проти личинок білшки довговусої тестували таким чином. Культуральний бульйон *Photobhabdus* (який було стерилізовано за допомогою стерилізуючого фільтру, безклітинний) або очищені за допомогою ВЕРХ фракції наносили безпосередньо на поверхню (~1,5см²) 0,25мл штучного корму у вигляді аліквот по 30мкл після розведення в контрольному середовищі або 10ММ фосфатному буфері, рН7.0, відповідно. Планшетам з кормом давали висохнути на повітрі у стерильному ламінарному боксі, та лунки заражали однією новонародженою *Diabrotica undecimpunctata howardi* (білшкою кукурудзяною довговусою (Південною), SCR), що вилупилася з поверхнево стерилізованого яйця), а личинки другої вікової стадії SCR росли на штучному кормі, або личинки другої вікової стадії *Diabrotica virgifera virgifera* (Western corn rootworm (білшки довговусої західної), WCR) вирощували на проростках кукурудзи, що росли в Metromix*. Личинки другої стадії зважували перед приміщенням на корм. Планшети герметизували, приміщували у зволожену камеру для вирощування та тримали при 27°C протягом підходячого періоду часу (4 дні для новонароджених та дорослих особин SCR, 2 - 5 днів для личинок WCR). Визначення смертності та маси оцінювали у балах, як зазначено. Як правило, в усіх дослідженнях використовували 16 комах на обробку. Смертність у контролях була такою: новонароджені личинки <5%, дорослі жуки - 5%.

Активність проти колорадського жука тестували так. Культуральний бульйон *Photobhabdus* або контрольне середовище наносили на поверхню (~2,0см²) 1,5мл стандартного штучного корму, що знаходився у лунках 24-лункового планшета для культури тканини.

Кожна лунка отримувала 50мкл бульйону, що його тестували, (обробка) та лункам давали висохнути на повітрі. Після цього окремі личинки другої вікової стадії колорадського жука (*Leptinotarsa decemlineata*, CPB) саджали на корм, та смертність оцінювали у балах після 4-го дня. В усіх дослідженнях використовували 10 личинок на обробку. Смертність в контролі сягала 3,3%.

Активність проти личинок та жуків (дорослих особин) хрущика японського тестували так. Газонні червоподібні личинки хрущика японського (*Popillia japonica*, 2-ої, 3-ої вікової стадії) збирали з зараженого лужку та тримали у лабораторії в суміші ґрунту з торфом та з шматочками моркви, що її додавали як додатковий корм. Газонних жуків ловили у пастках з феромоном, які було розташовано в деяких місцях, та тримали у лабораторії в пластикових контейнерах з листям клену, що було кормом. Після нанесення нерозчиненого культурального бульйону *Photobhabdus* або контрольного середовища на штучний корм для білшки довговусої (30мкл/1,54см², для жуків) або на шматочки моркви (для личинок) комах обох стадій саджали по одній у лунки з кормом та спостерігали за смертністю та живленням. В обох випадках спостерігали явне зменшення поїдання корму (та утворення екскрементів).

Активність проти личинок комарів тестували так. Тест проводили у 96-лунковому мікротитраційному планшеті. Кожна лунка містила 200мкл водного розчину (культурального бульйону *Photobhabdus*, контрольного середовища або води) та приблизно 20 одноденних личинок (*Aedes aegypti*) Для кожного варіанту використовували 6 лунок. Результати реєстрували за 2 години після зараження, та вони не змінювалися протягом триденного періоду спостереження. У контролі не було виявлено смертності.

Активність проти плодової мушки звичайної тестували так. Середовище для *Drosophila melanogaster* готували з 50% купленого сухого порошку та 50% рідини (води, контрольного середовища або культурального бульйону *Photobhabdus*). Це виконували, приміщуючи 8,0мл сухого середовища у кожен з трьох флаконів для вирощування для одного варіанту та додаючи 8,0мл відповідної рідини. Після цього у кожен флакон саджали десять личинок *Drosophila melanogaster*. Флакони тримали на лабораторному столі при кімнатній температурі під світлом флуоресцентних ламп, що знаходились на стелі. Підрахунок ляльок або дорослих особин проводили після 3, 7 та 10 днів експозиції. Включення культурального бульйону до кормового середовища для личинок плодової мушки звичайної спричинило невелике (17%), але значуще зменшення вилуплення на 10-ий день дорослих особин порівняно з водою та контрольним середовищем (3% зменшення).

Активність проти цикадки айстр тестували так. Тест з внутрішнім прийманням для цикадки айстр (*Macrosteles severini*) побудовано таким чином, щоб допустити тільки проковтування активного начала, без іншого зовнішнього контакту. Резервуар для розчину активного почала/"корму" готували, зробивши 2 дірки у центрі донної частини чашок Петрі розміром 35х10мм. Квадрат з Parafilm M*2 дюйми приміщували на верхню частину чашки та закріплювали "0"-кільцем. Після цього пластикову чашку на 1 унцію заражали приблизно 7 цикадками та зазначений резервуар приміщували на цю чашку парафіном донизу. Потім додавали тест-розчин до резервуару крізь дірки. У тестах з використанням нерозчиненого культурального бульйону* *Photobhabdus* бульйон та контрольне середовище діалізували проти води, щоб зменшити контрольну смертність. Смертність реєстрували на 2-й день, коли спостерігали 26,5% смертність у контролі. У тестах із застосуванням очищених фракцій (200мг білка на мл) використовували сахарозу у остаточній концентрації 5% в усіх варіантах для покращання виживання цикадок. Тест проводили в термостаті при 28°C, 70% відносної вологості з фотоперіодом 16 годин світла/8 годин темряви. Смертність визначали при 72 годинах. Смертність у контролі сягала 5,5%.

Активність проти мурашки аргентинського тестували так. Аліквоту 1,5мл 100% культурального бульйону *Photobhabdus*, контрольного середовища або води піпетували у прозорі скляні флакони на 2,0 мл. Флакони закупорювали шматочком ватного зубного тампону, який зволожували розчином відповідного варіанту. Кожний флакон приміщували до окремої чашки Петрі розміром 60х16мм з 8 - 12 дорослими особинами мурашок аргентинських (*Linepithema humile*). В кожному варіанті було три повторності. Чашки біотесту тримали на лабораторному столі при кімнатній температурі під світлом розташованих на стелі флуоресцентних ламп.

Смертність реєстрували після 5 днів експозиції. Смертність у контролі сягала 24%.

Активність проти мурашки-деревоточця тестували так. Робочих особин чорного мурашки-деревоточця (*Camponotus pennsylvanicus*) збирали з дерев маєтку DowElanco в Індіанополісі, IN. Тести з культуральним бульйоном *Photorhabdus* проводили так. Кожен пластиковий контейнер для біотесту (7 1/8 дюйма × 3 дюйми) містив 15 робочих особин, паперову підкладку та 10мл бульйону або контрольного середовища у пластиковій дозуючій склянці. Ватний тампон доставляв відповідний розчин до мурашок через отвір у кришці дозуючої склянки. Усі розчини містили 5% сахарозу. Біотести тримали у темряві при кімнатній температурі та оцінювали при 19 днях. Смертність у контролі дорівнювала 9%. У тестах, в яких доставляли очищені фракції, використовували штучний корм для мурашок, що його було змішано з відповідним розчином (очищеною фракцією або контрольним розчином) при співвідношенні 0,2мл розчину на 2,0г корму у пластиковій тест-пробірці. Кінцева концентрація білка очищеної фракції була меншою за 10мкг на грам корму. 10 мурашок на варіант, джерело води, підкладку та оброблений корм приміщували у запаковані (які було герметизовано) пластикові контейнери та тримали в темряві при 27°C у зволоженому термостаті. Смертність оцінювали у балах на 10-ий день. У контролі не спостерігали смертності.

Активність проти личинок різноманітних лускокрилих комах (*Lepidoptera*) тестували так. Культуральний бульйон *Photorhabdus* або очищені фракції наносили безпосередньо на поверхню (~1,5см²) 0,25мл стандартного штучного корму у вигляді аліквот по 30мкл після розведення у контрольному середовищі або 10мМ буфері з фосфату натрію, pH7,0, відповідно. Планшетам з кормом давали висохнути на повітрі у стерильному ламінарному боксі та лунки заражали окремими новонародженими личинками. Яйця точильника зернового (*Ostrinia nubilalis*) та совки кукурудзяної (*Helicoverpa zea*) одержували з комерційних джерел та вилуплення проводили в лабораторії, тоді як личинки совки малої (*Spodoptera exigua*), совки капустяної (*Trichoplusia ni*), совки тютюнової (*Heliothis virescens*), плодожерки яблуневої (*Laspeyresia pomonella*) та совки іпсилон (*Agrotis ipsilon*) збирали самі. Після зараження личинками планшети з кормом герметизували, приміщували до зволжених камер для вирощування та тримали у темряві при 27 °C протягом підходящого періоду. Визначення смертності та маси у балах проводили на 5-7-ий дні для культурального бульйону *Photorhabdus* та на 4 - 7-ий дні для очищеної фракції. Як правило, в усіх дослідженнях використовували 16 комах для кожного варіанту. Контрольна смертність була у межах 4 - 12,5% у випадку контрольного середовища та була меншою за 10% для фосфатного буферу.

Таблиця 3				
Вплив культурального бульйону <i>Photorhabdus luminescens</i> (штаму W-14) та фракції очищеного токсину на смертність та інгібування росту різних рядів/видів комах				
Ряд/вид комах	Бульйон		Очищена фракція	
	% смертн.	% I.P.	% смертн.	% I.P.
COLEOPTERA				
Блішка довговуса				
Південна/новонароджена личинка	100	нз	100	нз
Південна/2-а вікова стадія	нз	38,5	нв	нв
Південна/доросла особина	45	нв	нв	нв
Західна/2-а вікова стадія	нз	35	нв	нв
Колорадський жук				
2-а вікова стадія	93	нв	нв	нв
Газонні черв'яки (хрущик японський)				
3-я вікова стадія	нз	аф	нв	нв
дорослі особини	нз	аф	нв	нв
DIPTERA				
Плодова мушка (дорослі особини, що вилупилися)	17	нв	нв	нв
Личинки комарів	100	нз	нв	нв
НОМОPTERA				
Цикадка айстр	96,5	нз	100	нз
HYMENOPTERA				
Мурашка аргентинський	75	нз	нв	нз
Мурашка-деревоточець	71	нз	100	нз
LEPIDOPTERA				
Совка мала	12,5	36	18,75	41,4
Совка іпсилон	нв	нв	0	71,2
Совка капустяна	нв	нв	21,9	66,8
Плодожерка яблунева	нв	нв	6,25	45,9
Совка кукурудзяна	56,3	94,2	97,9	нз
Точильник зерновий	96,7	98,4	100	нз
Совка тютюнова	13,5	52,5	19,4	85,6
Смертн. = смертність, I.P. = інгібування росту, нз= не є застосовним, нв= не визначали, аф= антифідінг.				

Приклад 3

Застосовність як інсектицидів при внесенні до ґрунту

Було показано, що культуральний бульйон штаму W-14 *Photorhabdus luminescens* є активним проти блішки довговусої при внесенні безпосередньо до ґрунту або до ґрунтової суміші (Metromix*). Активність проти новонароджених SCR та WCR у Metromix* тестували так (Таблиця 4). Тест проводили з використанням проростків кукурудзи (United Agriseeds brand CL614), що пророщували на світлі на вологому фільтрувальному папері протягом 6 днів. Після того як коріння досягало довжини 3 - 6см, окреме зерно/проросток висаджували у прозору пластикову чашку на 591мл з 50г сухого Metromix*. Після цього двадцять новонароджених SCR або WCR приміщували безпосередньо на коріння проростку та закривали Metromix*. При зараженні проростки потім змочували розрідженим розчином бульйону, загальний об'єм якого сягав 50мл. Після змочування чашки герметизували та залишали при кімнатній температурі на світлі впродовж 7 днів. Після цього проростки промивали для видалення всього Metromix* та коріння відрізали та зважували. Активність виражали у вигляді відсотків коріння кукурудзи, що залишилося, у відношенні до контрольних рослин та у вигляді пошкодження листя, що його спричинили личинки, які його поїдали. Пошкодження листя оцінювали у балах візуально та виражали як -, +, ++ або +++, причому - означає відсутність пошкодження, а +++ означає значне пошкодження.

Активність проти новонароджених SCR в ґрунті тестували так (Таблиця 5). Тест виконували, використовуючи проростки кукурудзи (United Agriseeds Brand CL614), які пророщували на світлі на вологому фільтрувальному папері протягом 6 днів. Після того як коріння сягало довжини 3 - 6см, окреме зерно/проросток висаджували у прозору пластикову чашку на 591мл з 150г ґрунту, який було взято з поля у Lebanon, IN, де минулого року вирощували кукурудзу. Цей ґрунт раніше не обробляли інсектицидами. Після цього двадцять новонароджених SCR приміщували безпосередньо на коріння проростка та покривали ґрунтом. Після зараження проростки змочували розрідженим розчином бульйону, загальний об'єм якого сягав 50мл. Після змочування незакриті чашки інкубували у камері з високою відносною вологістю (80%) при 78°C. Після цього проростки промивали, щоб видалити весь ґрунт, та коріння відрізали та зважували. Активність виражали у вигляді відсотків коріння кукурудзи, що залишилося, у відношенні до контрольних рослин та у вигляді пошкодження листя, що його спричинювали личинки, що поїдали його. Пошкодження листя оцінювали візуально у балах та виражали у вигляді -, +, ++ або +++, причому - означає відсутність пошкодження, а +++ означає серйозне пошкодження.

Таблиця 4				
Вплив культурального бульйону штаму W-14 <i>Photorhabdus luminescens</i> на личинки блішки довговусої після змочування бульйоном після зараження (Metromix*)				
Варіант	Личинки	Пошкодження листя	Маса коріння (г)	%
Блішка довговуса Південна				
Вода	-	-	0,4916±0,023	100
Середовище (2,0% об /об)	-	-	0,4416±0,029	100
Бульйон (6,25% об /об)	-	-	0,4641±0,081	100
Вода	+	+++	0,1410±0,006	28,7
Середовище (2,0% об/об)	+	+++	0,1345±0,028	30,4
Бульйон (1,56% об/об)	+	-	0,4830±0,031	104
Блішка довговуса Західна				
Вода	-	-	0,4446±0,019	100
Бульйон (2,0% об/об)	-	-	0,4069±0,026	100
Вода	+	-	0,2202±0,015	49
Бульйон (2,0% об/об)	+		0,3879±0,013	95

Таблиця 5				
Вплив культурального бульйону штаму W-14 <i>Photorhabdus luminescens</i> на личинки блішки довговусої Південної після змочування бульйоном після зараження (ґрунт)				
Варіант	Личинки	Пошкодження листя	Маса коріння (г)	%
Вода	-	-	0,2148±0,014	100
Бульйон (50% об/об)	-	-	0,2260±0,016	103
Вода	+	+++	0,0916±0,009	43
Бульйон (50% об/об)	+		0,2428±0,032	113

Активність культурального бульйону штаму W-14 *Photorhabdus luminescens* проти газонних червоподібних личинок другої вікової стадії в Metromix* спостерігали у тестах, які проводили так (Таблиця 6). Приблизно 50г сухого Metromix* додавали до прозорої пластикової чашки на 591мл. Потім Metromix* просочували загальним об'ємом 50мл 50% (об./об.) розрідженого розчину бульйону *Photorhabdus*, вихідний бульйон розбавляли водою, причому 50% бульйон готували, додаючи 25мл вихідного бульйону до 25мл води з отриманням загального обсягу 50мл. 1% (маса/об'єм) розчин Proteose Pepton №3 (PP3), що є 50% розбавленням концентрації нормального середовища, використовували як контрольний бульйон. Після просочування 5 газонних черв'яків другої вікової стадії приміщували на верхню поверхню зволоженого Metromix*. Здорові личинки швидко ховалися в Metromix*. Личинки, що не ховалися в межах 1 години, вилучали та замінювали свіжими личинками. Чашки герметизували та приміщували в термостат на 28°C в темряві. Після 7 днів личинок виймали з Metromix* та

оцінювали їх смертність у балах. Активність виражали у вигляді відсотку смертності по відношенню до контролю.

Таблиця 6		
Вплив культурального бульйону штаму W-14 Photorhabdus luminescens на газонні личинки хрущика японського після просочування бульйоном перед зараженням (Metromix*)		
Варіант	Смертність*	% смертності
Вода	7/15	47
Контрольне середовище (1,0% маса /об'єм)	12/19	63
Бульйон (50% об/об)	17/20	85
* яку виражено у вигляді відношення мертві/живі личинки		

Приклад 4
Застосовність як інсектицидів при нанесенні на лист

Активність бульйону Photorhabdus проти точильника зернового спостерігали, коли бульйон наносили безпосередньо на поверхню листя кукурудзи (Таблиця 7). В цих тестах бульйон Photorhabdus розбавляли в 100 разів культуральним середовищем та наносили вручну на поверхню відрізаного листя кукурудзи при ~6,0мкл/см поверхні листя. Листя сушили на повітрі та розрізали на смуги рівного розміру приблизно 2×2 дюйми. Листя скручували, закріплювали скріпками для паперу та приміщували в пластикові дозовані склянки на 1 унцію з 0,25 дюйму 2% агару на дні для забезпечення вологості. Дванадцять новонароджених точильників приміщували на згорнутий лист, та чашки герметично закривали. Після інкубування протягом 5 днів при 27 °C в темряві проби оцінювали в балах на пошкодження, що спричинюється поїданням листя, та витягали личинки.

Таблиця 7			
Вплив культурального бульйону штаму W-14 Photorhabdus luminescens на личинки точильника зернового після нанесення бульйону на зрізане листя кукурудзи перед зараженням			
Варіант	Пошкодження листя	Личинки, які було витягнуто	Маса (мг)
Вода	Значне	55/120	0,42
Контрольне середовище	Значне	40/120	0,50
Бульйон	Сліди	3/120	0,15

Активність культурального бульйону проти новонароджених личинок совки тютюнової (Heliothis virescens) було показано з використанням методології занурення листя. Свіже листя бавовнику відрізали від рослини та вирізали листові диски (висічки) пробочним свердлом 18,5мм. Диски окремо занурювали до контрольного середовища (PP3) або культуральний бульйон штаму W-14 Photorhabdus luminescens, що його було сконцентровано приблизно у 10 разів за допомогою Amicon (Beverly, MA) з тангенціальною фільтраційною системою Proflux M12 з фільтром 10кДа. Надлишок рідини видаляли та випрямлену скріпку для паперу закріплювали через центр диску. Після цього скріпку закріплювали (втикали) у пластиковій дозованій склянці на 1 унцію, що містила приблизно 2,0мл 1% агару. Це робили для того, щоб підвісити диск вище рівня агару. Після висушування листового диску одну новонароджену личинку совки приміщували на диск та чашку закривали. Потім чашки герметизували у пластиковому мішку та приміщували до затемненого термостату при 27°C на 5 днів. Після цього періоду часу личинок, що залишалися, та листовий матеріал зважували, щоб визначити міру пошкодження листя (Таблиця 8).

Таблиця 8		
Вплив культурального бульйону штаму W-14 Pbotorhabdus luminescens на новонароджених личинок совки в тесті занурення листя		
Варіант	Листовий диск	Кінцеві маси (мг) личинок
Контрольне листя	55,7±1,3	нз*
Контрольне середовище	34,0±2,9	4,3±0,91
Бульйон Photorhabdus	54,3±1,4	0,0**
* - не є застосовним, ** - не виявлено живих личинок		

Приклад 5. Частина А
Характеристика пептидних компонентів токсину

У подальшому аналізі білкові субодиниці токсину смуг, які було виділено так, як описано у Прикладі 1, розділяли на 7% пол і акриламідному гелі для електрофорезу з співвідношенням 30:0,8 (акриламід:біс-акриламід), який містить ДСН. Цей гелевий матрикс дає найкраще розділення великих білків. Система гелю, яку застосовували для визначення молекулярних мас субодиниць Смуги 1 та Смуги 2 у Прикладі 1, була 18% гелем з співвідношенням 38:0,18 (акриламід :біс-акриламід), що дозволяло розділити більш широкий діапазон розмірів, але давало гірше розділення високомолекулярних компонентів.

В цьому аналізі одержували скоріше 10, ніж 8, смуг. У Таблиці 9 наведено розраховані молекулярні маси 10 отриманих смуг та безпосередньо порівняно молекулярні маси, які було визначено за цих умов, з молекулярними масами попереднього прикладу. Не є дивним те, що було виявлено додаткові смуги за різних умов розділення в цьому прикладі. Варіації між попередніми та новими визначеннями молекулярної маси слід було очікувати через відмінність умов аналізу. При аналізі цього прикладу вважали, що більші величини молекулярних мас є більш точними, ніж величини у Прикладі 1, внаслідок кращого розрішення. Однак, ці оцінки ґрунтуються на аналізі за допомогою електрофорезу в ПААГ-ДСН, що, як правило, не є аналітично точним та дає

оцінки пептидів, які може бути додатково змінено в результаті пост- та котрансляційних модифікацій.

Визначали амінокислотні послідовності для N-кінцевих частин п'яти з 10 отриманих пептидів. У Таблиці 9 наведено порівняння молекулярної маси білків та послідовностей, які було ідентифіковано. У випадку SEQ ID №2 деякі аналізи припускають, що пролін у позиції 5 може бути аспарагіном (asn). В випадку SEQ ID №3 деякі аналізи припускають, що амінокислотні залишки в позиціях 13 та 14 обидва є аргініном (arg). В випадку SEQ ID №4 деякі аналізи припускають, що амінокислотний залишок у позиції 6 може бути аланіном (ala) або серином (ser). В випадку SEQ ID №5 деякі аналізи припускають, що амінокислотний залишок у позиції 3 може бути аспарагіновою кислотою (asp).

Таблиця 9		
Приклад 1 Оцінка	Нова оцінка*	Послідовності
208	200,2кДа	SEQ ID №1
184	175,0кДа	SEQ ID №2
65,6	68,1кДа	SEQ ID №3
60,8	65,1кДа	SEQ ID №4
56,2	58,3кДа	SEQ ID №5
25,1	23,2кДа	SEQ ID №15

* Нові оцінки було отримано на основі електрофорезу в ПААГ-ДСН, а не на основі генних послідовностей. Електрофорез в ПААГ - ДСН не є аналітично точним.

Приклад 5. Частина В

Характеристика пептидних компонентів токсину

Нову N-кінцеву послідовність, SEQ ID №15, Ala Gln Asp Gly Asn Gln Asp Thr Phe Ser Gly Asn Thr, одержували шляхом подальшого N-кінцевого секвенування пептидів, які було виділено з нативного очищеного за допомогою ВЕРХ токсину, як описано вище у Частині А Прикладу 5. Цей пептид кодується геном tsaA. Пептид, що називається TsaA_{ij}, починається у позиції 254 та сягає до позиції 491, де починається пептид TsaA_{iii}, SEQ ID №4. Розмір цього пептиду, який визначено на основі генної послідовності, дорівнює 25240Да.

Приклад 6

Характеристика пептидних компонентів токсину

Ще в одному аналізі білковий комплекс токсину повторно ізолювали з середовища для вирощування Photorhabdus luminescens (після культивування без Твіну) шляхом осадження сульфатом амонію (10% - 80%) з наступною стадією іонообмінної хроматографії (Моно Q) та двома стадіями гель-фільтрування. Ці умови подібні до умов, що використовувалися в Прикладі 1. Впродовж першої стадії визначення розміру молекул (гель-фільтрування) було виявлено другий біологічно активний пік приблизно 100 ±10кДа. На основі вимірів білка, ця фракція була в 20 - 50 разів менш активна, ніж великий (або первинний) активний пік приблизно 860±100кДа (нативний). Під час цього експерименту по виділенню було також отримано активний пік меншого розміру приблизно 325±50кДа, що зберігав значну частину початкової біологічної активності. Припускалося, що пік 325кДа стосується до піку 860кДа або утворюється з нього.

В цьому аналізі було отримано білок 56кДа. N-кінцеву послідовність цього білка наведено у SEQ ID №6. Заслугує на увагу те, що цей білок поділяє значну ідентичність та консервативність з SEQ ID №5 при N-кінці, що дозволяє припустити, що ці два білки можуть кодуватись різними членами сімейств генів та що білки, які продукуються кожним геном, є достатньо схожими, щоб бути активними в інсектицидному комплексі токсину.

Другий помітний білок 185кДа був стійко наявним у кількості, порівняної з кількістю білка 3 з Таблиці 9, та може бути тим самим білком або білковим фрагментом. N-кінцева послідовність білка 185кДа показано в SEQ ID №7.

Додаткові дані щодо N-кінцевих амінокислотних послідовностей було також отримано з виділених білків. Жодна з визначених N-кінцевих послідовностей, очевидно, не є ідентичною білку, який було ідентифіковано в таблиці 9. У виділеному препараті були наявні й інші білки. Один з таких білків має визначену молекулярну масу 108кДа та його N-кінцеву послідовність показано в SEQ ID №8. Другий такий білок має визначену молекулярну масу 80кДа, та його N-кінцеву послідовність показано в SEQ ID №9.

При аналізі за розміром білкового матеріалу в активному піку приблизно 325кДа спостерігали смуги приблизно 51, 31, 28 та 22кДа. Як й в усіх випадках, коли молекулярну масу визначали аналізом електрофоретичної рухливості, ці молекулярні маси піддавались ефектам помилок, які були пов'язані з відмінностями іонної сили буферів, різними потужностями при електрофорезі тощо. Фахівцю у цій галузі зрозуміло, що остаточні величини молекулярних мас не можна визначити цими стандартними способами та що кожна молекулярна маса зазнавала варіацій. Було висловлено гіпотезу, що білки цих розмірів являють собою продукти деградації великих молекул білків (з розміром приблизно 200кДа), що їх спостерігали у більшому первинному комплексі токсину.

Нарешті, деякі препарати містили білок, що мав N-кінцеву послідовність, яку показано в SEQ ID №10. Ця послідовність мала сильну гомологію з відомими білками-"наставниками", допоміжними білками, що, як відомо беруть участь у збиранні великих білкових комплексів.

Хоча заявники не змогли б приписати таку функцію збирання білку, який ідентифіковано в SEQ ID №10, це погоджується з існуванням описаного білкового комплексу токсину, так що такий білок-"наставник" міг би брати участь в його збиранні. Крім того, хоча прямо не припускалося, що такі білки мають токсичну активність, цей білок міг би бути важливим для визначення загальної структурної природи білкового токсину та, отже, міг би

сприяти токсичній активності або тривалості існування цього комплексу in vivo після пероральної доставки.

Було використано такий аналіз стабільності білкового комплексу токсину до протеїнази К. Було показано, що після 24 годин інкубації комплексу в присутності 10-кратного молярного надлишку протеїнази К активність по суті зникла (смертність при пероральному введенні зменшувалась до приблизно 5%). Ці дані підтверджують білкову природу токсину.

Токсична активність утримувалася також діалізною мембраною, що також підтверджує великий розмір нативного комплексу токсину.

Приклад 7

Виділення, характеристика та часткове амінокислотне секвенування токсинів *Photobacterium*

Виділення та секвенування N-кінцевих амінокислот:

У ряді експериментів, які проводили паралельно з Прикладами 5 та 6, осадження сульфатом амонію білків *Photobacterium* проводили шляхом доведення бульйону *Photobacterium*, як правило 2 - 3 літрів, до кінцевої концентрації 10% або 20% повільним доданням кристалів сульфату амонію. Після перемішування протягом 1 години при 4°C матеріал центрифугували при 12000×g протягом 30 хвилин. Супернатант доводили до 80% сульфату амонію, перемішували при 4°C протягом 1 години та центрифугували при 12000×g протягом 60 хвилин. Осад повторно суспендували в 1/10 об'єму 10мМ Na₂PO₄, pH7,0 та діалізували проти того самого фосфатного буферу протягом ночі при 4°C. Матеріал, який було діалізовано, центрифугували при 12000×g протягом 1 години перед іонообмінною хроматографією.

Аніонообмінну колонку HR 16/50 Q Sepharose (Pharmacia) врівноважували 10мМ Na₂PO₄, pH7, 0. Осад, який було центрифуговано та діалізовано, отримано осадженням сульфатом амонію, наносили на колонку Q Sepharose при швидкості 1,5мл/хвил. та промивали ретельно при 3,0мл/хвил. буфером для врівноваження до тих пір, доки оптична густина (O.D. 280) не сягала величини меншої 0,100. Потім колонку елюювали або 60-хвилинним градієнтом NaCl (від 0 до 0,5М при 3мл/хвил.), або проводили ступінчасту елюцію з використанням 0,1М, 0,4М та нарешті 1,0М NaCl протягом 60 хвилин (кожний). Фракції об'єднували та концентрували за допомогою Centrprep 100. Альтернативно, білки можна елюювати одним промиванням 0,4М NaCl без попередньої елюції 0,1М NaCl.

Аліквоти по 2мл проб, які було концентровано, з Q Sepharose наносили при швидкості 0,5мл/хвил. на гель-фільтраційну колонку HR 16,50 Superose 12 (Pharmacia), урівноважену 10мМ Na₂PO₄, pH7,0. Колонку промивали тим самим буфером протягом 240 хвилин при 0,5мл/хвил. та збирали 2-хвилинні проби. Матеріал вільного об'єму збирали та концентрували за допомогою Centrprep 100. Аліквоти по 2мл проб, що концентрувалися, з Superose 12 наносили при швидкості 0,5мл/хвил. на гель-фільтраційну колонку HR 16/50 Sepharose 48-CL (Pharmacia), урівноважену 10мМ Na₂PO₄, pH7,0. Колонку промивали тим самим буфером протягом 240 хвилин при 0,5мл/хвил. та збирали 2-хвилинні проби.

Білковий пік, який було виділено, піддавали другому фракціонуванню шляхом нанесення на гель-фільтраційну колонку з Sepharose CL-4B, що розділяє білки в діапазоні від ~30кДа до 1000кДа. Цю фракцію розділяли на колонці на два піки: міnorний пік при вільному об'ємі (>1000кДа) та основний пік, що елюювався при середній мол. масі приблизно 860кДа. Протягом періоду одного тижня наступні проби, що їх піддавали гель-фільтруванню, виявили поступову появу третього піку (приблизно 325кДа), який, напевно, виникав з основного піку, можливо, внаслідок обмеженого протеолізу. Біотести, які було проведено на цих трьох піках, показали, що пік вільного об'єму не мав активності, тоді як фракція 860кДа комплексу токсину була високоактивною, а пік 325кДа був менш активним, хоча й цілком активним. Аналіз піків комплексу токсину з різних ферментаційних процесів, отриманих з використанням Sepharose CL-4B, за допомогою електрофорезу в ПААГ-ДСН виявив два різних розподіла пептидів, які позначено як "P" та "S". Ці два розподіли пептидів мали помітні відмінності в молекулярних масах та концентраціях пептидних компонентів в їх фракціях. Розподіл "S", який одержують частіше за все, мав 4 високомолекулярних пептиду (>150кДа), тоді як розподіл "P" мав 3 високомолекулярних пептиду.

Крім того, було виявлено, що пептидна фракція "S" має в 2 - 3 рази більшу активність проти точильника зернового. Це зміщення може бути пов'язане з модифікаціями в експресії білка, пов'язаними з віком інокуляту та/або з іншими факторами, що ґрунтуються на параметрах росту культур, які старіють.

Кількість близько мг фракцій комплексу токсину, що їх визначено як розподіл пептидів "P" або "S", піддавали препаративному електрофорезу в ПААГ-ДСН та транс-блотингу з Трис-гліцином (Seprabuff™ to PVDF membranes (ProBlott™, Applied Biosystems) протягом 3 - 4 годин. Блоти посиляли для амінокислотного аналізу та N-кінцевого секвенування амінокислот до Harvard MicroChem та Cambridge ProChem, відповідно. Три пептиди в розподілі "S" мали унікальні N-кінцеві амінокислотні послідовності порівняно з послідовностями, що ідентифікувалися в попередньому прикладі. Пептид 201кДа (TcdA_{ii}), який наведено у вигляді SEQ ID №13 має 33% ідентичних амінокислот та 50% схожих амінокислот з SEQ ID №1 (TcbA_{ii}) (Таблиця 10, де вертикальні лінії означають амінокислотну ідентичність, а двокрапки вказують заміни консервативних амінокислот). Другий пептид 197кДа, SEQ ID №14 (TcdB), мав 42% ідентичність та 58% гомологію з SEQ ID №2 (TsaC). Третій пептид 205кДа було позначено як TcdA_{ii}. Крім того, обмежена N-кінцева амінокислотна послідовність, SEQ ID №16 (TcbA), пептиду принаймні 235кДа була ідентичною в гомології з амінокислотною послідовністю SEQ ID №12, яку було розшифровано з клонованого гену (tcbA), SEQ ID №11, та яка містила розшифровану амінокислотну послідовність, що відповідала SEQ ID №1 (TcbA_{ii}). Це вказує на те, що більш великий пептид 235кДа протеолітично перетворювався на пептид 201кДа, (TcbA_{ii}), (SEQ ID №1) під час ферментації, можливо, спричиняючи активацію молекули. У випадку ще однієї послідовності, послідовності, яку спочатку було розкрито як SEQ ID №5 (TsaB_{ii}) в Прикладі 5 вище, було виявлено, що вона містить залишок аспарагінової кислоти (Asp)

у позиції 3, а не гліцин (Gly) та дві додаткові амінокислоти Gly та Asp в позиціях 8 та 9, відповідно. У випадку ще двох інших послідовностей, SEQ ID №2 (TcaC) та SEQ ID №3 (TcaB_i), було одержано додаткові амінокислоти. Денситометричні визначення виконували, застосовуючи пробу, що була ідентичною препарату "S", який було відіслано на аналіз N-кінцевої послідовності. Цей аналіз показав, що пептиди 201кДа та 197кДа складають 7,0% та 7,2%, відповідно, загального білка в розподілі "S", що його забарвлюють Кумасі (Coomassie brilliant blue), та наявні в кількостях, подібних до інших пептидів, що утворюють більші кількості. Обговорюється, що ці пептиди можуть бути білковими гомологами, аналогічно до ситуації, виявленої для інших бактеріальних токсинів, таких як різноманітні токсини Grl Bt. Ці білки дають варіації від 40% до 90% гомології у їх N-кінцевій амінокислотній послідовності, що включає їх токсичний фрагмент.

Секвенування внутрішньої амінокислотної послідовності:

Для полегшення клонування генів пептидів токсину внутрішні амінокислотні послідовності вибраних пептидів одержували так. Кількість близько кількох міліграмів фракцій піка 2A, що їх визначено як пептидні розподіли "P" або "S", піддавали препаративному електрофорезу в ПААГ-ДЧ та транс-блотингу з Трис-гліцином (Seprabuff™ на PVDF-мембранах (ProBlott™, Applied Biosystems) протягом 3 - 4 годин. Блоти посиляли на амінокислотний аналіз та N-кінцеве секвенування амінокислот до Harvard MicroChem та Cambridge ProChem, відповідно. Три пептиди, які називають TcbA_{ii} (що містить SEQ ID №1), TcdA_{ii}, та TcaBt (що містить SEQ ID №3), піддавали розщепленню трипсином в Harvard MicroChem з подальшою BEPX для розділення індивідуальних пептидів.

Аналіз N-кінцевих амінокислот виконували на вибраних пептидних фрагментах, отриманих при розщепленні трипсином. Два внутрішніх пептиду секвенували для пептиду TcaB_i (пептиду 205кДа), які названо TcaB_i-PTIII (SEQ ID №17) та TcaB_i-PT79 (SEQ ID №18). Два внутрішніх пептиду секвенували для пептиду TcaB_i (пептиду 68кДа), які названо TcaB_i-PT158 (SEQ ID №19) та TcaB_i-PT108 (SEQ ID №20). Чотири внутрішніх пептиду секвенували для пептиду TcbA_{ii} (пептиду 201кДа), які названо TcbA_{ii}-PT103 (SEQ ID №21), TcbA_{ii}-PT56 (SEQ ID №22), TcbA_{ii}-PT81 (a) (SEQ ID №23) та TcbA_{ii}-PT81 (b) (SEQ ID №24).

Таблиця 10

N-кінцеві амінокислотні послідовності

201 кДа (33% ідентичність та 50% схожість з SEQ ID № 1)

L I G Y N N Q F S G * A SEQ ID № 13

: | | | :

FI Q G Y S D L F G N - A SEQ ID № 1

197 кДа (42% ідентичність та 58% схожість з SEQ ID № 2)

M Q N S Q T F S V G E L SEQ ID № 14

| | : | | : : |

M Q D S P E V S I T T L SEQ ID № 2

Приклад 8

Конструювання космидної бібліотеки геномної ДНК Photorhabdus luminescens W-14 та її скринінг для виділення генів, що кодують пептиди, які містяться в препараті токсичного білка

Необхідною умовою для отримання токсичних для комах білків Photorhabdus в гетеро логічних живителях та для інших застосувань є виділення та характеристика генів, що кодують ці пептиди. Цим завданням займалися паралельно. Один підхід, який буде описано нижче, ґрунтується на використанні моноклональних та поліклональних антитіл, індукованих проти очищеного токсину, які потім використовували для виділення клонів з експресійної бібліотеки. Інший підхід, що його описано в цьому прикладі, ґрунтується на використанні даних про N-кінцеву та внутрішню амінокислотну послідовність для конструювання вироджених олігонуклеотидів для застосування їх в ПЦР-ампліфікації. Кожен спосіб можна використовувати для ідентифікації клонів ДНК, що містять кодуючі пептид гени, таким чином, щоб зробити можливим виділення відповідних генів та визначення послідовності основ їх ДНК.

Виділення геномної ДНК:

Штам W-14 Photorhabdus luminescens (ATCC accession number 55397) вирощували на 2% агарі Proteoso

Pepton №3 (Difco Laboratories, Detroit, MI) та компетентність інсектицидного токсину підтримували шляхом біотестів, що повторювалися, після пасажу, за допомогою засобу, який описано в Прикладі 1 та вище. 50мл культури, що її струшували, одержували в колбах з перегородками на 175мл в середовищі з 2% Proteose Pepton №3, вирощування проводили при 28 °C та 150об./хвил. протягом приблизно 24 годин. 15мл цієї культури осаджували та заморожували в її середовищі при -20°C до того часу, коли її розморожували для виділення ДНК. Культуру, яку було розморожено, центрифугували (700×g, 30хвил.) та оранжевий мукополісахаридний матеріал, що плавав, видаляли. Клітинний матеріал, що залишався, центрифугували (25000 ×g, 15хвил.) для осадження бактеріальних клітин та середовище видаляли й викидали.

Геномну ДНК виділяли адаптацією способу ЦТАБ, який описано в розділі 2.4.1 Current Protocols in Molecular Biology [Ausubel et al., eds John Wiley and Sons, 1994], що модифікувався для включення в нього сольового шоку та із збільшенням всіх об'ємів в 10 разів. Осаджені бактеріальні клітини повторно суспендували в ТЕ-буфері (10мМ Трис-НCl, 1мМ ЕДТК, pH8,0) до кінцевого об'єму 10мл, потім додавали 12мл 5М NaCl; цю суміш центрифугували протягом 20 хвилин при 15000 ×g. Осад повторно суспендували в 5,7мл ТЕ та до суспензії додавали 300мл 10% ДСН та 60мл протехнази К 20мг/мл (Gibco BRL Products, Grand Island, NY; в стерильній дистильованій воді). Цю суміш інкубували при 37°C протягом 1 години; після цього додавали приблизно 10мг лізозиму (Worthington Biochemical Corp., Freehold, NJ). Ще за 45 хвилин додавали 1мл 5М NaCl та 800мл розчину ЦТАБ/NaCl (10% мас./об. ЦТАБ, 0,7М NaCl). Цей препарат інкубували 10 хвилин при 65°C, потім обережно перемішували та інкубували далі та перемішували протягом приблизно 20 хвилин для виведення клітинного матеріалу. Додавали рівний об'єм розчину хлороформ/ізоаміловий спирт (24:1, об./об.), обережно перемішували та центрифугували. Після двох екстракцій рівним об'ємом ФХІ (фенол/хлороформ/ізоаміловий спирт; 50:49:1, об/об/об, зрівноваження 1 М Трис-НCl, pH8,0; Intermountain Scientific Corporation, Kaysville, UT), ДНК осаджували 0,6 об'ємами ізопропанолу. Осад ДНК обережно видаляли склянню паличкою, промивали двічі 70% етанолом, сушили та розчиняли в 2мл СТЕ (10мМ Трис-НCl, pH8,0, 10мМ NaCl, 1мМ ЕДТК). Цей препарат містив 2,5мг/мл ДНК, як визначено за оптичною густиною при 260нм (тобто OD₂₆₀).

Діапазон молекулярних розмірів виділеної геномної ДНК оцінювали на придатність для конструювання бібліотеки. CHEF-аналіз на гелі виконували в 1,5% агарозних (Seakem® LE, FMS BioProducts, Rockland, ME) гелях з 0,5хТБЕ-буфером (44,5мМ Трис-НCl, pH8,0, 44,5мМ H₃BO₃, 1мМ ЕДТК) на пристрої BioRad CHEF-DR 11 з Pulsewave 760 Switcher (BioRad Laboratories, Inc., Richmond, CA). Параметри переміщення в гелі: початковий А час, 3сек.; кінцевий А час, 12 сек.; 200 вольт; температура 4 - 18°C; час форецу, 16,5 годин. Забарвлення етидйбромидом та дослідження гелю під УФ світлом показали ДНК в діапазоні розмірів 30-250т.п.н.

Конструювання бібліотеки:

Частковий гідролізат Sau3A 1 було отримано з цього препарату геномної ДНК Photorhabdus. Спосіб ґрунтується на розділі 3.1.3 Ausubel (supra). Адаптації включали проведення реакцій меншого масштабу за різних умов до тих пор, доки не досягали майже оптимальних результатів. Проводили декілька широкомасштабних реакцій з різними умовами, результати аналізували в CHEF-гелях та вибирали тільки найкращий широкомасштабний спосіб проведення реакції. В такому оптимальному випадку 200мкг геномної ДНК Photorhabdus інкубували з 1,5 одиницями Sau3A I (New England Biolabs, "NEB", Beverly, MA) протягом 15 хвилин при 37°C в 2мл загального об'єму їх NEB 4-буферу (що поставляється у вигляді 10х виробником). Реакцію зупиняли доданням 2мл ФХІ та центрифугуванням при 8000 ×g протягом 10 хвилин. До супернатанту додавали 200мкл 5М NaCl та 6мл охолодженого на льоду етанолу. Цей препарат охолоджували протягом 30 хвилин при -20°C, після цього центрифугували при 12000 ×g протягом 15 хвилин. Супернатант видаляли та осад сушили у вакуумній сушильній шафі при 40°C, після цього повторно суспендували в 400мкл СТЕ. Спектрофотометричний аналіз показав приблизно 40% здобування введеної ДНК. Розщеплену ДНК фракціонували за розміром на сахарозному градієнті відповідно до розділу 5.3.2 CPMB (цит. вище). Лінійний сахарозний градієнт 10% - 40% готували з пристроєм для приготування градієнтів у пробірках Ultra-Clear (Beckman Instruments, Inc., Palo Alto, CA) та пробу ДНК нашаровували на градієнт. Після центрифугування (26000об./хвил., 17 годин, ротор SW41 Beckman, 20°C) фракції (приблизно 750мкл) відтягували з верхньої частини градієнта та аналізували за допомогою електрофорезу в CHEF-гелі (як описано раніше). Фракції, що містять Sau3A 1-фрагменти в діапазоні розмірів 20 - 40т.п.н., відбирали та ДНК осаджували модифікацією (кількості усіх розчинів збільшували приблизно у 6,3 разів) способом в розділі 5.3.3 Ausubel (supra). Після осадження протягом ночі ДНК збирали центрифугуванням (17000 ×g, 15хвил.), сушили, повторно розчиняли в ТЕ, об'єднували в кінцевий об'єм 80мкл та повторно осаджували доданням 8мкл 3М ацетату натрію та 220мкл етанолу. Осад, який було зібрано центрифугуванням, як описано вище, повторно суспендували в 12мкл ТЕ. Концентрацію ДНК визначали, застосовуючи флуорометрію з барвником Hoechst 33258 (Polysciences, Inc., Warrington, PA) у флуорометрі Hoefer TK 100 (Hoefer Scientific Instruments, San Francisco, CA). Здобували приблизно 2,5мкг фракціонованої за розміром ДНК.

Тридцять мкг ДНК космиди pWE15 (Stratagene, La Jolla, CA) розщепляли повністю з 100 одиницями рестриктази BamH I (NEB) в буфері виробника (кінцевий об'єм 200мкл, 37°C, 1 година). Реакційну суміш екстрагували 100мкл ФХІ та ДНК осаджували з водної фази доданням 20мкл 3М ацетату натрію та 550мкл охолодженого до 20°C абсолютного етанолу. Після 20 хвилин при -70 °C ДНК збирали центрифугуванням (17000×g, 15 хвилин), сушили під вакуумом та розчиняли в 180мкл 10мМ Трис-НCl, pH8,0. До розчину додавали 20мкл 10х СІР-буферу (100мМ Трис-НCl, pH8,3, 10мМ ZnCl₂, 10мМ MgCl₂) та 1мкл (0,25 одиниць) розбавленої 1:4 лужної фосфатази кишечнику теляти (Boeringer Mannheim Corporation, Indianapolis, IN). Після 30 хвилин при

37°C робили такі додання: 2мкл 0,5М ЕДТК, рН8,0; 10мкл 10% ДСН; 0,5мкл 20мг/мл протеїнази К (як зазначено вище) з подальшим інкубуванням при 55°C протягом 30 хвилин. Після послідовних екстракцій 100мкл ФХІ та 100мкл фенолу (Intermountain Scientific Corporation), зрівноваженого 1М Трис-НСІ, рН8,0, дефосфорильовану ДНК осаджували доданням 72мкл 7,5М ацетату амонію та 550мкл етанолу (-20°C), інкубуванням на льоду протягом 30 хвилин та центрифугуванням, як описано вище. Осаджену ДНК промивали один раз 500мкл 70% етанолу (-70°C), сушили під вакуумом та розчиняли в 20мкл ТЕ-буферу.

Літування фракціонованих за розміром Sau3A 1-фрагментів з Ваш 1, що його розщеплено, та обробленим фосфатазою вектором рWE15 виконували з застосуванням лігази Т4 (NEB) шляхом модифікації (тобто з застосуванням попередньо змішаного 10х буферу для літування, що поставляється виробником) протоколу у розділі 3.3.3 Ausubel. Лігування проводили протягом ночі в загальному об'ємі 20мкл при 15°C, з подальшим зберіганням при -20°C.

4мкл реакційної суміші для лігування космидної ДНК, що містять приблизно 1мкг ДНК, упаковували в бактеріофаг лямбда з використанням комерційного екстракту для упаковки (Gigapack[®] III Gold Packaging Extract/Stratagene) відповідно до вказівок виробника. Препарат, який було упаковано, зберігали при 4°C до застосування. Препарат космиди, який було упаковано, використовували для інфікування клітин Escherichia coli XL1 Blue MR (Stratagene) відповідно до протоколів Gigapack III Gold ("Titering the Cosmid Library") так. Клітини XL1 Blue MR вирощували в LB-середовищі (г/л: Бактотриптон, 10; Бакто-дріжджовий екстракт, 5; Бактоагар, 15; NaCl, 5; [Difco Laboratories, Detroit, MI]), що містить 0,2% (м/об) мальтозу плюс 10мМ MgSO₄, при 37°C. Після 5 годин зростання клітини осаджували при 700×g (15хвил.) та повторно суспендували в 6мл 10мМ MgSO₄. Густина культури доводили до OD₆₀₀=0,5 10мМ MgSO₄. Космидну бібліотеку, яку було упаковано, розбавляли 1:10 або 1:20 стерильним SM-середовищем (0,1М NaCl, 10мМ MgSO₄, 50мМ Трис-НСІ, рН7,5, 0,01% (м/об) желатин) та 25мкл розбавленого препарату змішували з 25мкл розбавлених клітин XL1 Blue MR. Суміш інкубували при 25°C протягом 30 хвилин (без качання), після цього додавали 200мкл LB-бульйону та інкубування продовжували протягом приблизно 1 години з рідким обережним струшуванням. Аліквоти (20 - 40мкл) цієї культури розподіляли по чашках з LB-агаром, що містили 100мг/л ампіциліну (тобто LB-Amp₁₀₀) та інкубували протягом ночі при 37°C. Для зберігання бібліотеки без ампліфікації окремі колонії вискрібали та інокулювали в індивідуальні лунки стерильних 96-луночних мікротитраційних планшетів; кожна лунка містила 75мкл Terrific Broth (ТВ - середовища: 12г/л Бактотриптону, 24г/л Бакто-дріжджового екстракту, 0,4% (об/об) гліцерин, 17мМ KH₂PO₄, 72мМ K₂HPO₄) плюс 100мг/л ампіциліну (тобто ТВ-Amp₁₀₀) та інкубували (без качання) протягом ночі при 37°C. Після реплікації 96-луночного планшета у планшет-копію до планшета додавали 75мкл/лунку стерилізованої за допомогою стерилізуючого фільтру суміші ТВ: гліцерин (1:1, об./об.; з 100мг/л ампіциліну або без нього), планшет качали недовго при 100об./хвил., 37°C та після цього закривали парафільмом (Parafilm[®], American National Can, Greenwich, CT) та приміщували в холодильник при 70°C для зберігання. Планшети-копії вирощували та обробляли так само, як і основні планшети. Всього одержували 40 таких основних планшетів (та їх копії).

Скринінг бібліотеки радіоактивно міченими зондами ДНК

Для приготування фільтрів з колоніями для зондування радіоактивно міченими зондами 10 96-луночних планшетів бібліотеки розморожували при 25°C (на столі при кімнатній температурі). Інструмент для способу відбитків (реплік) з 96 зубцями використовували для інокулювання свіжого 96-луночного планшета-копії, що містить 75мкл на лунку ТВ-Amp₁₀₀. Планшет-копію вирощували протягом ночі (стаціонарно) при 37°C, після цього качали приблизно 30 хвилин при 100об./хвил. при 37°C. Всього 800 колоній було представлено в цих планшетах-копіях, оскільки деякі ізоляти не росли. Інструмент для отримання реплік використовували для інокулювання дублікатних відбитків 96-луночних рядів на нейлонові мембрани Magna NT (MSI, Westboro, MA) (0,45 мікрон, 220×250мм), які приміщували на тверді середовища LB-Amp₁₀₀ (100мл/чашку) в пластикові чашки BioAssay (Nunc, 243×243×18мм; Curtin Mathison Scientific, Inc., Wood Dale, IL). Колонії вирощували на цих мембранах при 37°C протягом приблизно 3 годин.

Колонію позитивного контролю (бактеріальний клон, що містив вставку послідовності GZ4, див. нижче) вирощували на окремій мембрані Magna NT (Nunc, 0,45 мікрон, 82мм в діаметрі) на середовищі LB з доданням 35мг/л хлорамфеніколу (тобто LB-Cam₄₅) та обробляли разом з мембранами з колоніями бібліотеки. Бактеріальні колонії на мембранах лізували та ДНК денатурували та нейтралізували відповідно до протоколу, який було взято з версії 2,0 посібника для користувача Genius[™] System (Boehringer Mannheim, Indianapolis, IN). Мембрани приміщували на фільтраційний папір, який було просочено 0,5н NaOH плюс 1,5М NaCl, на 15 хвилин для денатурації та нейтралізували на фільтраційному папері, який було просочено 1М Трис-НСІ, рН8,0, 1,5М NaCl, протягом 15 хвилин. Після УФ-зшивання з використанням Stratagene UV Stratalinker set on auto crosslink мембрани зберігали сухими при 25°C до використання. Мембрани нарізали на смужки, що містили точно відтворені відбитки одного 96-луночного планшета, після цього промивали ретельно відповідно до способу розділу 6.4.1 в CPMB (цит. раніше): 3 години при 25°C в 3х SSC, 0,1% (м/об) ДСН, після цього протягом 1 години при 65°C в тому самому розчині, після цього промивали в 2х SSC при підготуванні до стадії гібридизації (20х SSC=3М NaCl, 0,3М цитрат натрію, рН7,0).

Ампліфікація специфічного геномного фрагменту гена tsaC

На основі N-кінцевої амінокислотної послідовності, визначеної для очищеної пептидної фракції TsaC (яку наведено тут у вигляді SEQ ID №2), синтезували пул вироджених олігонуклеотидів (пул S4Psh) у стандартний хімічний спосіб з використанням В-ціаноетилювання на синтезаторі ДНК/PHK Applied BioSystem AB1394 (Perkin

Elmer, Foster City, CA). Захисні групи олігонуклеотидів видаляли протягом 8 годин при 55 °С, після цього олігонуклеотиди розчиняли у воді, визначали їх кількість спектрофотометричним вимірюванням та розбавляли для використання. Цей пул відповідає визначеній N-кінцевій амінокислотній послідовності пептиду TcaC. Визначену амінокислотну послідовність та відповідну вироджену послідовність ДНК подано нижче, де А, С, G та Т являють собою стандартні основи ДНК, а І означає інозит:

інозин:

Аміно-

кислота	Met	Gln	Asp	Ser	Pro	Glu	Val
S4Psh	5' ATG	CA (A/G)	GA (T/C)	(T/A)(C/G)(T/A)	CCI	GA (A/G)	GT 3'

Синтезували інший набір вироджених олігонуклеотидів (пул P2.3.5R), який являв собою комплемент кодуючого ланцюга для визначеної амінокислотної послідовності SEQ ID №17:

Амінокис-

лота	Ala	Phe	Asn	Ile	Asp	Asp	Val
Кодони	5' GCN	TT (T/C)	AA (T/C)	AT (A/T/C)	GA (T/C)	GA (T/C)	GT 3'
P2.3.5R	3' CG(A/C/G/T)	AA (A/G)	TT (A/G)	TA (T/A/G)	CT (A/G)	CT (A/G)	CA 5'

Ці олігонуклеотиди використовували як праймери в Полімеразних Ланцюгових Реакціях (PCR[®], Roche Molecular Systems, Branchburg, NJ) для ампліфікації ДНК-фрагменту з геномної ДНК, яку було отримано з штаму W-14 Photorhabdus (див. вище). Типова реакція (50мкл) містила 125пмоль кожного пула праймерів P2Psh та P2.3.5R, 253нг геномної матричної ДНК, 10нмоль кожного з dATP, dCTP, dGTP та dTTP, буфер ПЦР IX GeneAmp[®] та 2,5 одиниць ДНК-полімерази AmpliTaq[®] (Roche Molecular Systems; 10x GeneAmp[®] буфер містить 100мМ Трис-HCl, pH8,3, 500мМ KCl, 0,01% м/об желатин). Ампліфікації проводили в Термоциклері для ДНК Perkin Elmer Cetus (Perkin Elmer, Foster City, CA) з використанням 35 циклів 94°C (1хвил.), 55°C (2хвил.), 72°C (3хвил.), з подальшим періодом подовження 7хвил. при 72°C. Продукти ампліфікації аналізували за допомогою електрофорезу на 2% (м/об) NuSieve[®] 3:1 агарозі (FMC BioProducts) у TEA-буфері (40мМ Трис-ацетат, 2мМ ЕДТК, pH8,0). Спостерігали специфічний продукт визначеного розміру 250п.н. серед численних інших продуктів ампліфікації при забарвленні зтидйбромідом (0,5мкг/мл) гелю та дослідженні під УФ-світлом.

Зону гелю, що містила продукт приблизно 250п.н., вирізали та видаляли з неї невелику пробку (діаметр 0,5мм) і використовували для подавання матриці для ПЦР-ампліфікації (40 циклів). Реакційна суміш (50мкл) містила ті самі компоненти, що описано вище, але без геномної матричної ДНК. Після ампліфікації кінці фрагментів затупляли та фосфорилювали шляхом інкубування при 25 °С протягом 20 хвилин з 1 одиницею ДНК-полімерази T4 (NEB), 1нмоль АТФ та 2,15 одиницями кінази T4 (Pharmacia Biotech Inc., Piscataway, NJ).

Фрагменти ДНК відділяли від решти праймерів за допомогою електрофорезу на 1% (м/об) GTG[®] agarose (FMC) в TEA. Вирізали шматок гелю, що містив фрагменти з розміром 250п.н., та ДНК екстрагували, застосовуючи набір Qiaex (Qiagen Inc., Chatsworth, CA).

Фрагменти ДНК, які було екстраговано, лігували з плазмидним вектором pBS KS(+) (Stratagene), який було розщеплено повністю рестриктазою Sma 1, та екстрагували у спосіб, подібний до описаного для ДНК pWE15 вище. Типова реакційна суміш для літування (16,3мкл) містила 100нг розщепленої ДНК pBS (+), 70нг фрагменту ДНК 250п.н., 1нмоль [Co(NH₃)₆]Cl₃ та 3,9 одиниць Weiss ДНК-лігази T4 (Collaborative Biomedical Products, Bedford, MA) в IX буфері для літування (50мМ Трис-HCl, pH7,4; 10мМ MgCl₂; 10мМ дітіотреїтол; 1мМ спермідин, 1мМ АТФ, 100мг/мл бичачого сироваткового альбуміну). Після інкубування протягом ночі при 14 °С ліговані продукти трансформували в заморожені, компетентні клітини DH5A Escherichia coli (Gibco BRL) відповідно до рекомендацій виробника та висіювали на чашки з LB-Sam₃₅, що містили ІПТГ (119мкг/мл) та X-гал (50мкг/мл). Незалежні білі колонії вискрібали та плазмидну ДНК одержували у модифікаційний спосіб, застосовуючи лужний лізис/осадження ПЕГ (PRISM Ready Reaction DyeDeoxy[™] Terminator Cycle Sequencing Kit Protocols; ABI/Perkin Elmer). Визначали нуклеотидну послідовність обох ланцюгів ДНК, яку було вставлено, застосовуючи праймери T7 [основи 601 - 623 pBS KS (+): TAAAACGACGGCCAGTGAGCGCG] та LacZ праймери [основи 792 - 816 pBS KS(+): ATGACCATGATTACGCCAAGCGCGC] й протоколи, що поставляються з набором для секвенування PRISM[™] (ABI/Perkin Elmer). Комплекси барвник-термінуючі дідезоксирибонуклеотиди, які не було включено, видаляли шляхом пропускання через колонки Centry-Sep 100 (Princeton Separations, Inc., Adelphia, NJ) відповідно до інструкцій виробника. Послідовність ДНК одержували шляхом аналізу проб на ДНК-секвенаторі ABI Model 373A (ABI/Perkin Elmer). Було виявлено, що послідовності ДНК двох ізолятів, GZ4 та HB14, були такими, як показано на Фіг.1.

Ця послідовність ілюструє такі ознаки:

і) основи 1-20 представляють одну з 64 можливих послідовностей вироджених олігонуклеотидів S4Psh,

ii) послідовність амінокислот 1 - 3 та 6 - 12 відповідає точно послідовності, визначеної для N-кінця TsaC (яку представлено у вигляді SEQ ID №2),

iii) четверта амінокислота, що її кодують, скоріше є залишком цистеїну, ніж залишком серину; цю відмінність заковано у виродженості для серинових кодонів (див. вище),

iv) п'ята амінокислота, яку кодують, є проліном, що відповідає N-кінцевій послідовності TsaC, яку представлено у вигляді SEQ ID №2,

N) основи 257 - 276 кодують одну з 192 можливих послідовностей, які було сконструйовано у виродженому пулі,

vi) термінуючий кодон TGA, введений при основах 268 - 270, є результатом комплементарності виродженої частини, вбудованої в пул олігонуклеотидів у відповідній позиції, та не свідчить про скороченість рамки зчитування для відповідного гена.

Мічення зонда, специфічного для гена пептиду TsaC.

Фрагменти ДНК, що відповідають наведеним вище 276 основам, було ампліфіковано (35 циклів) за допомогою ПЦР в реакційному об'ємі 100мкл з використанням 100пмоль кожного з праймерів P2Psh та P2.3.5R, 10нг плазмид GZ4 або HB14 як матриць, 20нмоль кожного з dATP, dCTP, dGTP та dTTP, 5 одиниць ДНК-полімерази AmpliTaq® та буферу GeneAmp в концентрації їх за температурних режимів, які описано вище. Продукти ампліфікації екстрагували з 1% GTG-агарозного гелю з використанням набору Qiaex та визначали їх кількість за допомогою флюорометрії.

Екстраговані продукти ампліфікації з плазмиди HB14 як матриці (приблизно 400нг) ділили на п'ять аліквот та мітили ³²P-dCTP з використанням суміші High Prime Labeling Mix (Boehringer Mannheim) відповідно до інструкцій виробника. Радіоізотопи, що виявились не включеними, видаляли шляхом пропускання через колонки для очищення зондів NucTrap (Stratagene) відповідно до інструкцій постачальника. Питому активність міченого ДНК-продукту визначали за допомогою сцинтиляційного лічильника, та вона дорівнювала $3,11 \times 10^8$ розпадів на хвилину на мкг. Цю мічену ДНК використовували для зондування мембран, які було приготовано з 800 членів геномної бібліотеки.

Скринінг з зондом, специфічним для гена пептиду TsaC.

Радіоактивно мічений зонд HB14 кип'ятили приблизно 10 хвилин, після цього додавали до розчину "мінімальний гіб.". [Примітка: спосіб "мінімальний гіб." Взято з протоколу CERES; "Restriction Fragment Length Polymorphism Laboratory Manual version 4,0", розділи 4 - 40 та 4 - 47; CERES/NPI, Salt Lake City, Ut. Зараз NPI не функціонує, його наступники діють як Linkage Genetics]. Розчин "мінімальної гіб." містить 10% м/об. ПЕГ (поліетиленгліколь, мол. маса приблизно 8000), 7% м/об. ДСН; 0,6×SSC, 10мМ буфер з фосфатом натрію (з 1М вихідного розчину, що містить 95г/л NaH₂PO₄·H₂O та 84,5г/л NaH₂PO₄·7H₂O), 5мМ ЕДТК та 100мг/мл денатурованої ДНК сперми лосося. Мембрани обсушували короточасним промакуванням та після цього, без попередньої гібридизації, 5 смужок мембрани приміщували в кожний з 2 пластикових боксів, які містили 75мл розчину "мінімальної гіб." та 2,6нг/мл радіоактивно міченого зонда HB14. Їх інкубували протягом ночі при повільному струшуванні (50об/хвил) при 60°C. Фільтри промивали три рази протягом приблизно 10хвил. (кожний) при 25°C в "промивному розчині мінімальної гіб." (0,25×SSC, 0,2% ДСН) з подальшими двома 30-хвилинними промиваннями з повільним струшуванням при 60°C в тому самому розчині. Фільтри приміщували на папір, покритий Saran Wrap® (Dow Brands, Indianapolis, IN) в світлонепроникній авторадіографічній касеті та експонували з рентгенівською плівкою X-Omat (Kodak, Rochester, NY) з двома підсилюючими екранами Du Pont Cronex Lightning-Plus CI (Sigma Chemical Co., St. Louis, MO) протягом 4 годин при -70°C. При проявленні (стандартні процедури фотографії) значні сигнали було видно в обох копіях серед високого тла більш слабких, більш нерівномірних сигналів. Фільтри знову промивали протягом приблизно 4 годин при 68°C в "промивному розчині мінімальної гіб." і потім приміщували знов в ці касети та плівку експонували протягом ночі при -70°C. 12 можливих позитивних відбитків ідентифікували внаслідок сильних сигналів на обох дублікатах відбитків 96-луночних колоній. Сигнали не було виявлено з мембранами негативного контролю (колонії клітин XL1 Blue MR, що містили pWE15), та дуже сильний сигнал спостерігали з мембранами позитивного контролю (клітини DH5A, що містили ізолят GZ4 продукту ПЦР), які обробляли паралельно з експериментальними пробами.

Дванадцять можливих гібридизаційно-позитивних колоній здобували з заморожених 96-луночних планшетів з бібліотекою та вирощували протягом ночі при 37°C на твердому середовищі LB-Amp₁₀₀. Потім їх приміщували на тверде середовище LB-Amp₁₀₀ (3 на чашку плюс 3 негативних контролю: клітини XL1 Blue MR, що містять вектор pWE15). Два набору мембран (Magna NT nylon 0,45 мікрон) готували для гібридизації. Перший набір готували, приміщуючи фільтр безпосередньо на колонії на чашці з "петчами" ("осередками") та витягаючи його разом з бактеріальними клітинами, що поприлипали, з подальшою обробкою, як описано нижче. Фільтри другого набору приміщували на чашки, що містили середовище LB-Amp₁₀₀, яку було інокульовано потім шляхом перенесення клітинами з чашок з "петчами" на фільтри. Після росту протягом ночі при 37°C фільтри видаляли з чашок та обробляли.

Бактеріальні клітини на фільтрах лізували та ДНК денатуровували шляхом приміщення кожного фільтру (колоніями догори) на пул (калюжку) (1,0мл) 0,5н NaOH в пластиковій чашці на 3 хвилини. Фільтри обсушували промакуванням на паперовому рушнику, потім процес повторювали зі свіжим розчином 0,5н NaOH. Після промакування фільтри нейтралізували шляхом приміщення кожного на пул 1,0мл 1М Трис-HCl, pH7,5, на три хвилини, промакували для обсушування та повторно нейтралізували свіжим буфером. Після цього проводили два подібних просочування (по 5 хвилин) на калюжках з 0,5М Трис-HCl, pH7,5, плюс 1,5М NaCl. Після промакування для обсушування ДНК зшивали ультрафіолетом з фільтром (як описано вище) та фільтри

промивали (25°C, 100об/хвил) приблизно в 100мл 3х SSC плюс 0,1% (м/об) ДСН (4 рази по 30 хвилин зі свіжим розчином для кожного промивання). Потім їх приміщували в мінімальний об'єм передгібризаційного розчину [6х SSC плюс 1% (м/об) (кожний) Ficoll400 (Pharmacia), полівінілпіролідон (середня мол. маса 360000; Sigma) та бичачий сироватковий альбумін (фракція V) (Sigma)] протягом 2 годин при 65 °С, 50об/хвил. Передгібризаційний розчин видаляли та замінювали ³²Р-міченим зондом HB14, що його було збережено з попередньої гібридизації мембран з бібліотекою та який денатурували при 95 °С протягом 5 хвилин. Гібридизацію проводили при 60°C протягом 16 годин з качанням при 50об/хвил.

Після видалення розчину міченого зонда мембрани промивали 3 рази при 25 °С (50об/хвил, 15 хвилин) в 3х SSC (приблизно по 150мл для кожного промивання). Потім їх промивали протягом 3 годин при 68°C (50об/хвил) в 0,25х SSC плюс 0,2% ДСН (промивний розчин мінімальної гіб.) та експонували на рентгенівській плівці, як описано вище, протягом 1,5 години при 25°C (без підсилюючих екранів). Це експонування виявило дуже сильні сигнали гібридизації з ізолятами космид 22G12, 25A10, 26A5 та 26B10 та дуже слабкий сигнал з ізолятом космиди 8B10. Не спостерігали сигналу з колоніями негативного контролю (pWE15), та дуже сильний сигнал спостерігали з мембранами позитивного контролю (клітини DH5A, що містять ізолят GZ4 продукту ПЛР), які обробляли паралельно з експериментальними пробами.

Ампліфікація специфічного геномного фрагменту гена tsaB

На основі N-кінцевої амінокислотної послідовності, яку було визначено для очищеної фракції пептиду TsaB_i (що її описано тут у вигляді SEQ ID №3), синтезували пул вироджених олігонуклеотидів (пул P8F), як описано для пептиду TsaC. Визначену амінокислотну послідовність та відповідну вироджену послідовність ДНК наведено нижче, де A, C, G та T є стандартними основами ДНК, а I означає інозин:

Аміно-

кислота Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg

P8F 5' TTT ACT CA(A/G) ACT (C/T)TI AAA GAA GCI (A/C)G3'

Синтезували інший набір вироджених олігонуклеотидів (пул P8.108.3R), який являв собою комплемент кодуючого ланцюга для визначеної амінокислотної послідовності внутрішнього пептиду TsaB_i-PT108 (яку описано тут як SEQ ID №20):

Амінокислоти Met Tyr Tyr Ile Gln Ala Gln Gln

Кодони ATG TA(T/C) TA(T/C) AT(T/C/A) CA(A/G) GCA/C/G/T CA(A/G) CA(A/G)

P8.108.3R 3' TAC AT(A/G) AT(A/G) TA(A/G/T) GT(T/C) CGI GT(T/C) GT 5'

Ці олігонуклеотиди використовували як праймери для ПЛР з застосуванням пробірок HotStart 50 Tubes (Molecular Bioproducts, Inc., San Diego, CA) для ампліфікації специфічного фрагменту ДНК з геномної ДНК, отриманої з штаму W-14 Photorhabdus (див. вище). Типова реакційна суміш (50мкл) містила (нижній шар) 25пмоль кожного пула праймерів, P8F та P8.108.3R з 2пмоль кожного з dATP, dCTP, dGTP та dTTP в буфері ПЛР 1х GeneAmp® та (верхній шар) 230нг геномної матричної ДНК, 8нмоль кожного з dATP, dCTP, dGTP та dTTP та 2,5 одиниць ДНК-полімерази AmpliTaq в буфері їх GeneAmp®. Ампліфікації проводили шляхом 35 циклів, як описано для пептиду TsaC. Продукти ампліфікації аналізували за допомогою електрофорезу на 0,7% (м/об) SeaKem®LE агарозі (FEM) в ТЕА-буфері. Спостерігали специфічний продукт з визначеним розміром 1600п.н.

Чотири таких реакції об'єднували та ДНК, яку було ампліфіковано, екстрагували з 1,0% SeaKem LE гелю з використанням набору Qiaex, як описано для пептиду TsaC. ДНК, яку було екстраговано, використовували безпосередньо у вигляді матриці для визначення послідовності (набір для секвенування PRISM) з використанням пулів праймерів P8F та P8.108.3R. Кожна реакція містила приблизно 100нг матричної ДНК та 25пмоль одного з пулів праймерів та проводилася відповідно до стандартного протоколу, як описано для пептиду TsaC. Аналіз послідовності, отриманої з подовження праймерів P8F, виявив коротку послідовність ДНК (та амінокислотну послідовність, яку кодують):

GAT GCA TTG NTT GCT
Asp Ala Leu (Val) Ala

яка відповідає частині N-кінцевої пептидної послідовності, описаної у вигляді SEQ ID №3 (TsaB_i).

Мічення зонда, специфічного для гена пептиду TsaB_i

Приблизно 50нг очищеного на гелі фрагменту ДНК TsaB_i мітили ³²Р-dCTP, як описано вище, та радіоізотопи,

що не було включено, видаляли шляхом пропускання через колонку NICK (Pharmacia). Було визначено, що питома активність міченої ДНК дорівнювала 6×10^8 розпадів на хвилину на мкг. Цю мічену ДНК використовували для зондування мембран з колоніями, отриманими з членів геномної бібліотеки, які гібридизувались з зондом, специфічним для пептиду TsaC.

Мембрани, що містили ці 12 колоній, які було ідентифіковано в скринінгу бібліотеки TsaC-зондом (див. вище), звільняли від радіоактивної, TsaC-специфічної мітки шляхом кип'ятіння 2 рази по 30 хвилин в 1 літрі $0,1 \times$ SSC плюс $0,1\%$ ДСН. Видалення радіоактивної мітки перевіряли з 6-годинною експозицією плівки. Потім звільнені від мітки мембрани інкубували з отриманим, як описано вище, зондом, специфічним для пептиду TsaB_i. Мічену ДНК денатурували кип'ятінням протягом 10 хвилин та потім додавали до фільтрів, які після цього інкубували протягом 1 години в 100мл розчину "мінімального гіб." при 60°C . Після гібридизації протягом ночі при цій температурі розчин зонда видаляли та фільтри промивали так (всі в $0,3 \times$ SSC+ $0,1\%$ ДСН): один раз протягом 5 хвилин при 25°C , один раз протягом 1 години при 60°C в свіжому розчині та один раз протягом 1 години в свіжому розчині. Після 1,5-годинної експозиції на рентгенівській плівці відповідно до стандартних процедур спостерігали 4 колонії, що сильно гібридизувалися. Це були, як і з зондом, специфічним для TsaC, ізоляти 22G12, 25A10, 26A5 та 26B10.

Той самий розчин зонда розбавляли рівним об'ємом (приблизно 100мл) розчину "мінімальний гіб." та потім використовували для скринінга мембран, що містять 800 членів геномної бібліотеки. Після гібридизації, промивання та експозиції з рентгенівською плівкою, як описано вище, було виявлено, що тільки чотири космидних клони 22G12, 25A10, 26A5 та 26B10 сильно гібридизувалися з цим зондом.

Виділення субклонів, що містять гени, які кодують пептиди TsaC та TsaB_i, та визначення послідовності основ їх ДНК:

Три гібридизаційно-позитивні космиди в штамі XLI Blue MR вирощували з качанням протягом ночі (200об./хвил.) при 30°C в 100мл середовища TB-Amp₁₀₀. Після збирання клітин центрифугуванням, космидну ДНК одержували за допомогою комерційно доступного набору (BIGprep™, 5 Prime 3 Prime, Inc., Boulder, CO) відповідно до протоколів виробника. Тільки одну космиду, 26A5, було виділено успішно у цей спосіб. При розщепленні рестриктазою EcoR 1 (NEB) та аналізі за допомогою гель-електрофорезу було виявлено фрагменти з приблизними розмірами 14, 10, 8 (вектор), 5, 3,3, 2,9 та 1,5т.п.н. В другій спробі виділити космидну ДНК з тих самих трьох штамів (культура в 8мл; TB-Amp₁₀₀, 30°C) використовували miniprep-спосіб з кип'ятінням [Evans G. And G. Wahl., 1987, "Cosmid vectors for genomic walking and rapid restriction mapping" in Guide to Molecular Cloning Techniques Meth. Enzymology, vol.152, S. Berger and Kimmel, eds., pgs. 604-610]. У цей спосіб успішно виділили лише одну космиду. При розщепленні рестриктазою EcoR 1 та аналізі за допомогою гель - електрофорезу ця космида виявила розподіл фрагментів, що є ідентичним розподілу, який було виявлено з космидою 26A5.

0,15мкг проби ДНК космиди 26A5 використовували для трансформації 50мл клітин DH5A E. Coli (Gibco BRL) відповідно до протоколів виробника. Ізолят однієї колонії цього штаму інокулювали в 4мл середовища TB-Amp₁₀₀ та вирощували протягом 8 годин при 37°C . Додавали хлорамфенікол аж до кінцевої концентрації 225мкг/мл, інкубацію продовжували ще протягом 24 годин, потім клітини збирали центрифугуванням та заморожували при -20°C . ДНК космиди 26A5 виділяли шляхом стандартного лужного лізису (miniprep) [Maniatis et al., цит. вище, р.382] з модифікацією, що полягає в збільшенні всіх об'ємів на 50%, та з перемішуванням або обережним змішуванням, а не за допомогою вортексу, на кожній стадії. Після промивання осаду ДНК в 70% етанолі її розчиняли в TE, що містив 25мкг/мл рибонуклеази А (Boehringer Mannheim).

Ідентифікація фрагментів EcoR 1. що гібридизуються з отриманими з GZ4 та TsaB_i-зондів.

Приблизно 0,4мкг космиди 25A10 (з клітин XLI Blue MR) та приблизно 0,5мкг космиди 26A5 (з ампліфікованих з хлорамфеніколом клітин DH5A) розщепляли окремо приблизно з 15 одиницями EcoR I (NEB) протягом 85 хвилин, заморожували протягом ночі, потім нагрівали при 65°C протягом 5 хвилин та піддавали електрофорезу в 0,7% агарозному гелі (Seakem® LE, $1 \times$ TEA, 80 вольт, 90 хвилин). ДНК забарвлювали етилійбромідом, як описано вище, та фотографували під УФ-світлом. За цих умов продукт розщеплення EcoR 1 космиди 25A10 було розщеплено повністю, але пробу космиди 26A5 було лише частково розщеплено. Агарозні гелі, що містили фрагменти ДНК, піддавали депуринізації, денатурації та нейтралізації з подальшим блотингом за Саузерном на нейлоновій мембрані Magna NT з використанням протоколу з високою концентрацією солі ($20 \times$ SSC), всі ці засоби описано [в розділі 2.9 Ausubel et al. (CPMB, цит. вище)]. Потім перенесену ДНК зшивали УФ-світлом з нейлоновою мембраною, як описано раніше.

Фрагмент ДНК, специфічний для пептиду TsaC, який відповідав вставці плазмидного ізоляту GZ4, ампліфікували за допомогою ПЛР в реакційному об'ємі 100мл, як описано вище. Продукти ампліфікації з трьох таких реакцій об'єднували та екстрагували з 1% GTG®-агарозного гелю з використанням набору Qiaex, як описано вище, та кількісно визначали за допомогою флуориметрії. Очищену на гелі ДНК (100нг) мітили ^{32}P -dCTP з використанням суміші для мічення High Prime Labeling Mix (Boehringer Mannheim), як описано вище, аж до питомої активності $6,34 \times 10^8$ розпадів на хвилину на мкг.

^{32}P -мічений зонд GZ4 кип'ятили 10 хвилин, потім додавали до буферу "мінімальний гіб." (при 1нг/мл), додавали мембрану з блотом за Саузерном, що містила розщеплені фрагменти ДНК космиди, та інкубували протягом 4 годин при 60°C з обережним струшуванням при 50об./хвил. Потім мембрану промивали 3 рази при 25°C протягом приблизно 5 хвилин в кожному випадку (промивним розчином "мінімальної гіб. "), з подальшими двома промиваннями протягом 20 хвилин (кожен) при 60°C . Блот експонували на плівці (з підсилюючими

екранами) протягом приблизно 30хвил. при -70°C. Зонд GZ4 сильно гібридизувався з фрагментом EсоR 1 5,0т.п.н. (видимий розмір) обох плазмид, 26A5 та 25A10.

Мембрану визволяли від радіоактивності кип'ятінням протягом 30 хвилин в 0,1х SSC+0,1% ДСН та відсутність радіоактивної мітки перевіряли експозицією на плівці. Потім її гібридизували при 60°C протягом 3,5 годин з денатурованим зондом ТсаВ_i в буфері "мінімальної гіб.", що використовувався раніше для скринінгу мембран з колоніями (див. вище), промивали, як описано раніше, та експонували на плівці протягом 40 хвилин при -70°C з двома підсилюючими екранами. У випадку обох плазмид, зонд ТсаВ_i гібридизувався слабо з фрагментом EсоR 1 5,0т.п.н. та сильно з фрагментом приблизно 2,9т.п.н.

Пробу ДНК космиди 26A5, яку було описано раніше, (з клітин DH5A) використовували як джерело ДНК, з якого субклонували цільові смуги. Цю ДНК (2,5мкг) переварювали з приблизно 3 одиницями EсоR 1 (NEB) в загальному об'ємі 30мкл протягом 1,5 годин з отриманням продукту часткового розщеплення, що було підтверджено гель-електрофорезом. Десять мкг ДНК рВC KS (+) (Stratagene) розщепляли протягом 1,5 годин з 20 одиницями EсоR 1 в загальному об'ємі 20 мкл, що спричинювало повне розщеплення, як було підтверджено електрофорезом. Обидва препарати розрізаної EсоR 1 ДНК розбавляли до 50мкл водою, до кожного додавали рівний об'єм ФХІ, суспензію обережно змішували, відкручували у мікроцентрифузі та водний супернатант збирали. ДНК осаджували 150мкл етанолу та суміш приміщували при -20°C на ніч. Після центрифугування та висушування розщеплену EсоR 1 плазмиду рBS KS (+) розчиняли в 100мкл ТЕ; частково розщеплену 26A5 розчиняли в 20мкл ТЕ. Здобування ДНК контролювали флуориметрією.

В окремих реакціях приблизно 60нг ДНК рBS KS (+), розщепленої EсоR 1, лігували приблизно з 180нг або 270нг частково розщепленою ДНК космиди 26A5. Лігування проводили в об'ємі 20мкл при 15°C протягом 5 годин з використанням лігази T4 та буферу з New England BioLabs. Суміш для лігування, розбавлену до 100мкл стерильним ТЕ, використовували для трансформації заморожених, компетентних клітин DH5A E. Coli (Gibco BRL) відповідно до інструкцій виробника. Різноманітні кількості (25 - 200мкл) клітин, що їх було трансформовано, висіювали на свіжоприготовлене тверде середовище LB-Cam₃₅ з 1мМ ІПГТ та 5мг/л X-гал. Чашки інкубували при 37°C протягом 20 годин, потім охолоджували в темряві протягом приблизно 3 годин для підсилення забарвлення для відбирання інсертів. Білі колонії вискрібали та переносили на чашки з "петчами" того самого складу та інкубували протягом ночі при 37°C.

Два "ліфти" колоній кожної з вибраних "петч"-чашок готували так. Після перенесення білих колоній на свіжі чашки круглі нейлонові мембрани Magna NT притискали на "петч"-чашки, мембрану піднімали та піддавали денатурації, нейтралізації та зшиванню під впливом УФ, як описане вище для мембран з колоніями бібліотеки. Піддані зшиванню "ліфти" колоній інтенсивно промивали, обережно змиваючи надлишок залишків клітин тканиною. Набір гібридизували з розчином зонда GZ4 (ТсаC), описаного раніше, а інший набір гібридизували з розчином зонда ТсаВ[±], який було описано вище, відповідно до протоколу "мінімальної гіб.", з подальшим промиванням та експозицією на плівці, як описано для мембран з колоніями бібліотеки.

Колонії, що виявляли сигнали гібридизації, або тільки з зондом GZ4, з обома зондами GZ4 та ТсаВ_i, або тільки з зондом ТсаВ_i, відбирали для подальшої роботи та клітини висіювали штрихом для виділення окремої колонії на середовища LB-Cam₃₅ з ІПГТ та X-гал, як описано раніше. Приблизно 35 окремих колоній, з 16 ізолятів, переносили до рідких середовищ LB-Cam₃₅ та вирощували протягом ночі при 37°C; клітини збирали центрифугуванням та плазмидну ДНК виділяли за допомогою стандартного лужного лізису (mini-prep) відповідно до Maniatis et al. (цит. вище, р.368). Осади ДНК розчиняли в ТЕ+25мкг/мл рибонуклеази А та концентрацію ДНК визначали флуориметрією. Розподіл продуктів розщеплення EсоR 1 аналізували гель-електрофорезом. Наступні ізоляти здобували у звичайний спосіб. Ізолят А17.2 містить тільки повторно ліговану рВC KS (+) та його використовували для (негативного) контролю. Ізоляти D38,3 та C44,1 (кожний) містять лише 2,9т.п.н., фрагмент EсоR 1, що гібридизується з ТсаВ_i, який вбудовано в рВC KS (+). Ці плаزمиди, які названо рDAB2000 та рDAB2001, відповідно, показано на Фіг.2.

Ізолят А35,3 містить лише приблизно 5 т.п.н., фрагмент EсоR 1, що гібридизується з GZ4, який вбудовано в рВC KS (+). Цю плазмиду назвали рDAB2002 (також Фіг.2). Ці ізоляти забезпечили матриці для секвенування ДНК.

Плазмиди рDAB2000 та рDAB2001 одержували з використанням набору BIGprep™, як описано вище. Культури (30мл) вирощували протягом ночі в середовищі TB-Cam₃₅ аж до OD₆₀₀2, потім плазмиду виділяли відповідно до інструкцій виробника. Осад ДНК перерозчиняли в 100мкл ТЕ (кожну) та цілісність проб перевіряли розщепленням EсоR 1 та аналізом за допомогою гель-електрофорезу.

Реакції секвенування проводили в двох повторностях, з однією реплікою, що використовує як матрицю ДНК рDAB2000, та з іншою реплікою, що використовує як матрицю ДНК рDAB2001. Реакції проводили, застосовуючи спосіб циклічного секвенування з комплексом дідезокси-барвник як термінатором, який описано вище для секвенування ДНК GZ4/HB14. Початкові раунди секвенування використовували як праймери праймери LacZ та T7, описані вище, плюс праймери, основані на визначеній послідовності ПЛР-ампліфікованого продукту ТсаВ_i,

(TH1=ATTGCAGACTGCCAATCGCTTCGC,

TH12=GAGAGTAGTATCCAGACCGCGGATGATCT).

Після аналізу первинної структури та підготовки кожного секвенуючого продукту, кожен укорочували до 250-350 основ, залежно від повноти хроматографічних даних, що інтерпретувалися програмним забезпеченням Perkin Elmer Applied Biosystems Division SeqEd675. Подальші "стадії" секвенування виконували шляхом вибору підходящої послідовності для нових праймерів. За невеликими винятками, праймери (синтезовані, як описано вище) мали довжину 24 основи зі складом, що мав 50% G+C. Секвенування у цей спосіб проводили на обох ланцюгах фрагменту EcoR 1 з величиною приблизно 2,9т.п.н.

Для подальшого використання як матриці для секвенування ДНК одержували плазмидну ДНК з ізоляту рDAB2002, використовуючи набір BIGrep. Реакції секвенування проводили та аналізували, як описано вище. Спочатку праймер T3 (основи 774-796 рBC KS (+): CGCGCAATTAACCCCTCACTAAAG) та праймер T7 (основи 621-643 рBC KS (+): GCGCGTAATACGACTCACTATAG) використовували для праймування реакцій секвенування з фланкуючих вектор послідовностей зі зчитуванням у вставлення ДНК. Інший набір праймерів

(GZ4F: GTATCGATTACAACGCTGTCACTTCCC;

TH13: GCGAAGTGACAGCGTTGTAATCGATAC;

TH14: ATGTTGGGTGCGTCGGCTAATGGACATAAC; та

LW1-204: GCGAAGTGACAGCGTTGTAATCGATAC) застосовували для праймування с внутрішніх послідовностей, які визначали попередньо шляхом секвенування з використанням вироджених олігонуклеотидів субклонованих ПЛР-продуктів пептиду TsaC. На основі даних, отриманих під час початкових раундів секвенування, конструювали нові набори праймерів та використовували їх для "прогулянки" по всій довжині фрагменту ~5т.п.н. Всього використовували 55 оліго-праймерів, що дозволило ідентифікувати в цілому 4832п.н. неперервної послідовності.

При об'єднанні послідовності ДНК інсерту фрагменту EcoR 1 рDAB2002 з частиною визначеної послідовності ізолятів рDAB2000/рDAB2001 одержували спільну послідовність суміжних нуклеотидів довжиною 6005п.н. (яку наведено тут у вигляді SEQ ID №25). При трансляції довгих відкритих рамок зчитування у відповідні амінокислоти ця послідовність ясно виявляє N-кінцевий пептид (який описано у вигляді SEQ ID №3), що його кодують основами 19-75, безпосередньо після залишку метіоніну (стартова точка трансляції). Далі по ходу транскрипції, в тій самій рамці зчитування, при основах 1738 - 1773 знаходиться послідовність, що кодує внутрішній пептид TsaB_i-PT108 (який описано у вигляді SEQ ID №20). Також в тій самій рамці зчитування, при основах 1897-1923, кодується N-кінцевий пептид TsaB_i (який описано тут у вигляді SEQ ID №5) та рамка зчитування продовжується без переривання до кодона термінації транскрипції при нуклеотидах 3586 - 3588.

Відсутність всередині рамки зчитування стоп-кодону між кінцем послідовності, що кодує TsaB_i-PT108, та початком кодуєного району TsaB_i та відсутність помітного сайту зв'язування рибосом безпосередньо проти ходу транскрипції від кодуєного району TsaB_{ii} вказує на те, що пептиди TsaB_{ii} та TsaB_i кодуються однією відкритою рамкою зчитування з 3567п.н., що починається при парі основ 16 в SEQ ID №25, та найбільш ймовірно, вони утворюються з єдиного первинного продукту гена з 1189 амінокислот (131586Да; який описано тут у вигляді SEQ ID №26) в результаті посттрансляційного розщеплення. Якщо амінокислота, яка безпосередньо передує N-кінцевому пептиду TsaB_i, являє собою C-кінцеву амінокислоту пептиду TsaB_i, то прогнозована маса TsaB_{ii} (627 амінокислот) дорівнює 70814Да (яку описано тут у вигляді SEQ ID №28), що дещо більше, ніж розмір, який спостерігався за допомогою електрофорезу в ПААГ-ДСН (68кДа). Цей пептид кодувався б неперервним відрізком з 1881п.н. (який описано тут у вигляді SEQ ID №27). Можна думати, що нативний C-кінець TsaB_i лежить дещо ближче до C-кінця TsaB_i-PT108. Молекулярна маса PT108 [3,438кДа; визначена під час аналізу N-кінцевої амінокислотної послідовності цього пептиду] передбачає розмір з 30 амінокислот. Використовуючи цей розмір пептиду для визначення C-кінця кодуєного району TsaB_i [Glu у позиції 604 SEQ ID №28], отриманий розмір TsaB_i можна визначити як 604 амінокислоти або 68463Да, що більше погоджується з експериментальними спостереженнями.

Трансляція кодуєного району з 1686 пар основ пептиду TsaB_{ii} (який описано тут у вигляді SEQ ID №29) дає білок з 562 амінокислот (який описано тут у вигляді SEQ ID №30) з прогнозованою масою 60789 Да, що добре погоджується з розміром 61кДа, який спостерігають.

Потенційний сайт зв'язування рибосом (основи 3633 - 3638) було виявлено на відстані 48п.н. по ходу транскрипції від стоп-кодону для відкритої рамки зчитування tsaB. При основах 3645 - 3677 виявлено послідовність, що кодує N-кінець пептиду TsaC, (яку описано тут у вигляді SEQ ID №21). Відкрита рамка зчитування, що починається цим N-кінцевим пептидом, триває без перерви до основи 6005 (2361п.н., що її представлено тут у вигляді перших 2361п.н. SEQ ID №31). Ген (tsaC), що кодує весь пептид TsaC, (середній розмір ~165т.п.н.; ~1500 амінокислот), має містити приблизно 4500п.н.

Інший ізолят, що містив клоновані фрагменти EcoR 1 космиди 26A5, E20.6, також ідентифікували за його гомологією з названими раніше зондами GZ4 та TsaB_i. Аналіз в агарозному гелі продуктів розщеплення EcoR 1 ДНК цієї плаزمиди, що її утримує цей штам (рDAB2004, Fir.2), виявив вставлені фрагменти визначених розмірів

2,9, 5 та 3,3т.п.н. Аналіз послідовності ДНК, що ініціюється від праймерів, які було сконструйовано з послідовності плаزمиди pDAB2002, виявив, що фрагмент EcoR 1 3,3т.п.н. pDAB2004 лежить поруч з фрагментом EcoR 1 5т.п.н., представленим в pDAB2002. Відкрита рамка зчитування з 2361п.н., яку виявлено в pDAB2002, продовжується без переривання на наступні 2094п.н. в pDAB2004 [яку описано тут у вигляді пар основ 2362 - 4458 SEQ ID №31]. Аналіз послідовності ДНК з використанням ДНК батьківської космиди 26A5 як матриці підтвердив неперервність цієї відкритої рамки зчитування. В цілому, ця відкрита рамка зчитування (TsaC SEQ ID №31) містить 4455п.н. та кодує білок (TsaC) з 1485 амінокислот [який представлено тут у вигляді SEQ ID №32]. Розрахований молекулярний розмір 166214Да погоджується з визначеним розміром пептиду TsaC (165кДа) та вироблена амінокислотна послідовність співпадає точно з послідовністю, яку описано для N-кінцевої послідовності TsaC [SEQ ID №2].

Відсутність амінокислотної послідовності, що відповідає SEQ ID №17, яку використано для конструювання пула вироджених олігонуклеотидних праймерів, у виявленій послідовності вказує на те, що утворення продуктів ПЛР, виявлених в ізолятах GZ4 та HB14, які було використано як зонди в початковому скринінгу бібліотеки, здійснювалося шляхом випадкового праймування зворотного ланцюга одним з праймерів в виродженому пулі. Крім того, утворена білкова послідовність не містить внутрішнього фрагменту, який описано тут у вигляді SEQ ID №18. Ці послідовності показують, що плазмида pDAB2004 містить повний кодуючий район для пептиду TsaC.

Приклад 9

Скринінг геномної бібліотеки Photorhabdus на гени, що кодують пептид TsbA_{ij}

Цей приклад описує спосіб, що застосовується для ідентифікації клонів ДНК, які містять гени, кодуючі пептид TsbA_{ij}, виділення цього гена та визначення його часткової послідовності основ ДНК.

Праймери та ПЛР-реакції

Поліпептид TsbA₁₁, активного проти комах препарату має мол. масу ~206кДа. Амінокислотну послідовність N-кінця цього пептиду описано у вигляді SEQ ID №1. Синтезували 4 пула вироджених олігонуклеотидних праймерів ("Прямі праймери": TH-4, TH-5, TH-6 та TH-7) для кодування цієї амінокислотної послідовності, як описано в Прикладі 8. їх показано нижче.

Таблиця 11

Аміно- кислота	Phe	Ile	Gln	Gly	Tyr	Ser	Asp	Leu	Phe
TH-4	5-TT(T/C)	ATT	CA(A/G)	GGI	TA(T/C)	TCT	GA(T/C)	CTT	TT-3'
TH-5	5-TT(T/C)	ATT	CA(A/G)	GGI	TA(T/C)	AG(T/C)	GA(T/C)	CTT	TT-3'
TH-6	5-TT(T/C)	ATT	CA(A/G)	GGI	TA(T/C)	TCT	GA(T/C)	TT(A/G)	TT-3'
TH-7	5-TT(T/C)	ATT	CA(A/G)	GGI	TA(T/C)	AG(T/C)	GA(T/C)	TT(A/G)	TT-3'

Крім того, було визначено первинну ("a") та вторинну ("b") послідовність препарату внутрішнього пептиду (TsbA_{ij}-PT81), що описано тут у вигляді SEQ ID №23 та SEQ ID №24, відповідно.

У подібний спосіб було сконструйовано та синтезовано 4 пула вироджених олігонуклеотидів ("Зворотні праймери": TH-8, TH-9, TH-10 та TH-11) для кодування зворотного комплементу послідовностей, що кодують частину пептидів SEQ ID №23, яку показано нижче.

Таблиця 12

5	Амів	Thr	Tyr	Leu	Thr	Ser	Phe	Glu	Gln	Val	Ala	Asn
	окис											
	лота											
10	ТН-8	3'TGI	АТА(Г)	ГАІ	TGI	АГІ	АА(А/Г)	СТ(Т/С)	СТ(Т/С)	САІ	СГІ	ТГА(А/Г)
	ТН-9	3'TGI	АТА(Г)	ТТА(Г)	TGI	АГІ	АА(А/Г)	СТ(Т/С)	СТ(Т/С)	САІ	СГІ	ТГА(А/Г)
15	ТН-10	3'TGI	АТА(Г)	ГАІ	TGI	ТГА(А)	АА(А/Г)	СТ(Т/С)	СТ(Т/С)	САІ	СГІ	ТГА(А/Г)
	ТН-11	3'TGI	АТА(Г)	ТТА(Г)	TGI	ТГА(А)	АА(А/Г)	СТ(Т/С)	СТ(Т/С)	САІ	СГІ	ТГА(А/Г)

Набори цих праймерів використовували в реакціях ПЛР для ампліфікації фрагментів гена, що кодує TcbA_{ij}, з геномної ДНК штаму W-14 Photorhabdus luminescens, яку було отримано в Прикладі 6. Всі реакції ПЛР виконували за способом "Hot Start" з використанням AmpliWax™ gems та інших реагентів та протоколів Perkin Elmer. Як правило, суміш (загальний об'єм 11мкл) MgCL2, dNTP, 10х буферу ІІ для ПЛР GeneAmp® та праймери додавали до пробірок, що містили одну воскову бусину, [10х GeneAmp® PCR Buffer ІІ складається з 100мМ Трис-НСІ, рН8,3, та 500мМ КСІ]. Пробірки нагрівали до 80°С протягом 2 хвилин та давали їм охолонути. На верхню поверхню воскових затворів, що утворилися, додавали розчин, що містив 10х GeneAmp® PCR буфер ІІ, ДНК-матрицю та ДНК-полімеразу AmpliTaq®. Після розплавлення воскової пробки та змішування компонентів шляхом термоциклювання кінцеві умови реакції (об'єм 50мкл) були такими: 10мМ Трис-НСІ, рН 8,3; 50мМ КСІ; 2,5 MgCl₂, 200мкМ кожного з dATP, dCTP, dGTP, dTTP; 1,25мМ в одному пулі Прямих праймерів; 1,25мкМ в одному пулі Зворотних праймерів, 1,25 одиниць ДНК-полімерази AmpliTaq та 170нг матричної ДНК. Реакційні суміші приміщували до термоциклеру (як у Прикладі 8) та проводили ПЛР за такою програмою:

Таблиця 13		
Температура	Час	Повторення циклів
94°С	2 хвилини	1Х
94°С	15 секунд	30Х
55-56°С	30 секунд	30Х
72°С	1 хвилина	30Х
72°С	7 хвилин	1Х
15°С	Постійна	

Серію ампліфікацій проводили при трьох різних температурах відпала (55°, 60°, 65°С) з використанням пулів вироджених олігонуклеотидів. Реакції з відпалом при 65°С не давали продуктів ампліфікації, видимих після електрофорезу в агарозних гелях. Реакції, що мали режим відпала з 60°С та містили праймери ТН-5+ТН-10, давали продукт ампліфікації, що мав рухливість, яка відповідала 2,9т.п.н. Менша кількість продукту 2,9т.п.н. продукувалася за цих умов з праймерами ТН-7+ТН-10. При відпалі реакцій при 55 °С ці пари праймерів продукували більше продукту 2,9т.п.н., та цей продукт одержували також з парами праймерів ТН-5+ТН-8 та ТН-5+ТН-11. Додаткові дуже слабкі смуги 2,9т.п.н. було видно в доріжках, що містили продукти ампліфікації від пар праймерів ТН-7+ТН-8, ТН-9, ТН-10 та ТН-11. Для отримання достатньої кількості продукту ПЛР-ампліфікації для клонування та визначення послідовності ДНК готували 10 окремих ПЛР-реакцій з праймерами ТН-5+ТН-10 та їх проводили за описаних вище умов з температурою відпала 55°С. Всі реакційні суміші об'єднували та продукт 2,9т.п.н. очищували екстракцією по Qіаех з агарозного гелю, як описано вище.

Додаткові послідовності, визначені для внутрішніх пептидів ТСbA_{ij}, описано тут у вигляді SEQ ID №21 та SEQ ID №22. Як і раніше, було приготовано вироджені олігонуклеотиди (Зворотні праймери ТН-17 та ТН-18), що відповідають зворотному комплементу послідовностей, які кодують частину амінокислотної послідовності цих пептидів.

Таблиця 14

3 SEQ ID № 21

Аміно- кислота	Met	Glu	Thr	Gln	Asn	Ile	Gln	Glu	Pro
ТН-17	3'-TAC	CTT/C	TGI	CTT/C	TTA/G	TAI	GTT/C	GTT/C	GG-5'

Таблиця 15

3 SEQ ID № 22

Аміно- кислота	Asn	Pro	Ile	Asn	Ile	Asn	Thr	Gly	Ile	Asp
ТН-18	3-TT(A/G)	GGI	TAI	TT(A/G)	TAI	TT(A/G)	TGI	CCI	TAI	CTAG-5'

Вироджені олігонуклеотиди ТН-18 та ТН-17 використовували в експерименті по ампліфікації з ДНК *Photorhabdus luminescens* W-14 як матрицю та праймерів ТН-4, ТН-5, ТН-6 або ТН-7 як 5'-(Прямі) праймери. Ці реакції ампліфікували продукти приблизно 4т.п.н. та 4, 5т.п.н., відповідно. Ці ДНК переносили з агарозних гелів на нейлонові мембрани та гібридизували з Р-міченим зондом (як описано вище), який було приготовано з продукту 2,9т.п.н., ампліфікованим за допомогою пари праймерів ТН-5+ТН-10. Обидва продукти ампліфікації 4т.п.н. та 4,5т.п.н. сильно гібридизувалися з зондом 2,9т.п.н. Ці результати використовували для побудови карти, що розташовує внутрішні пептидні послідовності ТсbA_{ij}, як показане на Фіг.3.

Послідовність фрагменту, що кодує ТсbA_{ij} 2,9т.п.н.

Приблизно 200нг очищеного фрагменту 2,9т.п.н. (отриманого, як описано вище) осаджували етанолом та розчиняли в 17мл води. Одну половину використовували як матрицю з 25 пмоль пула ТН-5 як праймерів, а іншу половину використовували як матрицю з праймером ТН-10. Реакції секвенування проводили, як описано в Прикладі 8. З пулом праймерів ТН-10 не було отримано вірогідної послідовності; однак реакції з пулом праймерів ТН-5 продукували описану нижче послідовність:

```

1   AATCGTGTTC ATCCCTATGC CGNCCCGGGT TCGGTGGAAT CGATGTCTTC ACCGGGGGTT
51  TATTNGAGGC ANTNGTCCCG TCAGGCCAAA AANTGGAATG AAAGAAGTTC AATTINTTAC
121 CTAGATAAAC GTCGCCCGGN TTTAGAAAGN TTANTGNTCA GCCAGAAAAT TTTGGTTGAC
131 GAAATTCCAC CGNTGGTTCT CTCTATTGAT TNGGGCCTCG CCGGGTTCCA ANNAAAACNA
241 GGAATATCAC AAGTTGAGGT GATGGNTTTC TNGCNANCTT NTCGTTTAGG TCGCGAGAAA
301 CCTINTCANC ACGNTTNTGA AACTGTCCGG GAAATCGTCC ATGANCGTCA MCCAGGNTTN
361 CGCCATTGG

```

На основі цієї послідовності було сконструйовано секвенуючий праймер (ТН-21, 5'-CCGGGCGACGTTTATCTAGG-3') для основ зворотного комплементу 120-139 та ініціації полімеризації у напрямі 5'-кінця (тобто ТН-5-кінця) очищеного на гелі ПЛР-фрагменту 2,9т.п.н. ТсbA_{ij}. Визначену послідовність показано нижче та її порівняно з біохімічно визначеною N-кінцевою послідовністю пептиду ТсbA_{ij} SEQ ID №1.

Підтвердження послідовності ПЛР-фрагменту 2,9т.п.н. ТсbA_{ij}

(Підкреслені амінокислоти= послідовності, що їх кодують вироджені олігонуклеотиди)

5

F I Q G Y S D L F G - - A

2,9 GC ATG CAG GGG TAT AGT GAC CTG TTT GGT AAT CGT GCT

10

Т.П.Н.

M Q G Y S D L F G N R A

15

З гомології виробленої амінокислотної послідовності з послідовністю, визначеною біохімічно, впливає, що ПЛР-фрагмент 2, 9т.п.н. являє собою район, що кодує TcbA. Цей фрагмент 2,9т.п.н. використовували потім як гібридизаційний зонд для скринінгу геномної бібліотеки Photorhabdus W-14, отриманої в Прикладі 8, на космиди, що містять ген, який кодує TcbA_{ij}.

20

Скринінг космидної бібліотеки Photorhabdus

25

Очищений на гелі фрагмент ПЛР 2,9т.п.н. мітили ³²P з використанням набору для мічення Boehringer Mannheim High Prime, як описано в Прикладі 8. Фільтри, що містили залишки приблизно 800 колоній з космидної бібліотеки, скринювали, як описано раніше (Приклад 8), та позитивні клони висіювали штрихом для отримання окремих колоній та повторно піддавали скринінгу. Три клони (8A11, 25G8 та 26DI) давали позитивні результати протягом декількох стадій скринінгу та характеристики. Гібридизації TcbA_{ij}-специфічного зонда ніколи не одержували з будь-якою з чотирьох космид, що ідентифікувалися в Прикладі 8, які містять гени tcaB та tcaC. ДНК з космид 8A11, 25G8 та 26DI розщепляли рестриктазами Big 2, EcoR 1 або Hind 3 (окремо або в комбінації одна з одною), та фрагменти розділяли на агарозному гелі та переносили на нейлонову мембрану, як описано в Прикладі 8. Мембрану гібридизували з ³²P-міченим зондом, отриманим з фрагменту 4,5т.п.н. (отриманого шляхом ампліфікації геномної ДНК Photorhabdus з праймерами TH-5+TH-17). Розподіли (малюнки смуг), отримані з космидних ДНК 8A11 та 26DI, були ідентичними з розподілами, отриманими з геномного ДНК, що її розрізають у подібний спосіб, на тій самій мембрані. Було зроблено висновок, що космиди 8A11 та 26DI є точними відображеннями кодуючого локусу геномного TcbA_{ij}. Однак, космида 25G8 має один фрагмент Bgl 2, що є дещо більшим за геномну ДНК. Це може бути результатом розташування інсерту всередині вектору.

35

Послідовність ДНК гена, що кодує TcbA

40

Гібридизаційний аналіз на мембрані космиди 26DI виявив зонд 4,5т.п.н., який було гібридизовано з одним великим фрагментом EcoR 1 (більшим, ніж 9т.п.н.). Цей фрагмент очищали на гелі та лігували в сайт EcoR 1 pBC KS (+), як описано в Прикладі 8, для створення плазмиди pBC-S1/R1. Часткову послідовність ДНК інсерту ДНК цієї плазмиди визначали "прогулянкою праймера" від фланкуючої послідовності вектору з використанням процедур, описаних в Прикладі 8. Подальшу послідовність генерували шляхом подовження від нових олігонуклеотидів, що їх було сконструйовано з раніше визначеної послідовності. При порівнянні з визначеною послідовністю ДНК для гена tcbA, яку ідентифікували іншими способами (яку описано тут у вигляді SEQ ID №11, як описано в Прикладі 12 нижче) було виявлена повну гомологію з нуклеотидами 1-272, 319-826, 2578-3036 та 3068-3540 (загальна кількість основ = 1712). Було зроблено висновок, що обидва підходи може бути використано для ідентифікації фрагментів ДНК, які кодують пептид TcbA_{ij}.

45

Аналіз отриманої з гена tcbA амінокислотної послідовності

50

Послідовність фрагменту ДНК, яку ідентифіковано як SEQ ID №11, кодує білок, амінокислотну послідовність якого описано тут як SEQ ID №12. Деякі ознаки підтверджують ідентичність цього гена як гена, що кодує білок TcbA_{ij}. N-кінцевий пептид (SEQ ID №1: Phe lie Gin Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala) кодує амінокислоти 88-100. Внутрішній пептид TcbA_{ij} TcbA_{ij}-PT81 (a) (SEQ ID №23) кодує амінокислоти 1065-1077, а TcbA_{ij}-PT81 (b) (SEQ ID №24) кодує амінокислоти 1571-1592. Крім того, внутрішній пептид TcbA_{ij}-PT56 (SEQ ID №22) кодує амінокислоти 1474-1488, а внутрішній пептид TcbA_{ij}-PT103 (SEQ ID №24) кодує амінокислоти 1614-1639. Очевидно, що цей ген є аутентичним клоном, що кодує пептид TcbA_{ij}, який виділено з препаратів інсектицидного білка штаму W-14 Photorhabdus luminescens. Білок, який виділено у вигляді пептиду TcbA_{ij}, утворено внаслідок розщеплення більш довгого пептиду. Доводить це той факт, що нуклеотиди, які кодують N-кінцевий пептид SEQ ID №1 TcbA_{ij}, мають перед ними 261 основу (кодуючі 87 найближчих N-кінцевих амінокислот) більш довгої відкритої рамки зчитування (SEQ ID №11). Ця рамка зчитування починається нуклеотидами, що кодують амінокислотну послідовність Met Gin Asn Ser Leu, яка відповідає N-кінцевій послідовності великого пептиду TcbA, та яку описано тут як SEQ ID №16. Можна думати, що TcbA є білком-попередником TcbA_{ij}.

60

Спорідненість генів tcbA, tcaB та tcaC

65

Гени tcaB та tcaC є тісно зв'язаними та можуть транскрибуватись у вигляді однієї мРНК (Приклад 8). Ген tcbA знаходиться на космидах, які, очевидно, не перекриваються з космидами, що несуть кластер tcaB та tcaC, тому що відповідні скринінги геномної бібліотеки ідентифікували різні космиди. Однак, порівняння амінокислотних послідовностей, що їх кодують гени tcaB та tcaC, з геном tcbA, виявило значну міру гомології. Консервативність амінокислот (Protein Alignment Mode of MacVector™ Sequence Analysis Software, scoring

matrix pam250, hash value=2; Kodak Scientific Imaging Systems, Rochester, NY) показано на Фіг.4. На лінії оцінки кожної частини в Фіг.4 направлені догори значки (^) вказують гомологію або зміни консервативних амінокислот, а направлені донизу значки (N) вказують відсутність гомології.

Цей аналіз показує, що амінокислотна послідовність пептиду TcbA від залишку 1739 до залишку 1894 є високо гомологічною амінокислотам 441-603 пептиду TsaB_i (162 з 627 амінокислот P8; SEQ ID №28). Крім того, послідовність амінокислот TcbA 1932-2459 є високо гомологічною амінокислотам 12-531 пептиду TsaB_{ii} (520 з 562 амінокислот; SEQ ID №30). Зважаючи на те, що пептид TcbA (SEQ ID №12) містить 2505 амінокислот, всього 684 амінокислоти (27%) при С-проксимальному кінці його є гомологічними пептидам TsaB_i або TsaB_{ii}, та гомології розміщені колінеарно з розташуванням можливого препротейну TsaB (SEQ ID №26). Розмір проміжку в гомології TcbA співпадає з місцем сполучення між частинами TsaB_i та TsaB_{ii} препротейну TsaB. Ясно, що генні продукти TcbA та TsaB є еволюційно спорідненими, та припускається, що вони мають загальну функцію (функції) в *Photorhabdus*.

Приклад 10

Характеристика цинк-металопротеаз в бульйоні *Photorhabdus*: Інгібування протеази, класифікація та очищення

Аналіз інгібування протеази та класифікаційні аналізи: Тести на протеазу виконували з застосуванням ФІТЦ-казеїну, який було розчинено у воді, як субстрату (кінцева концентрація в тесті 0,08%). Реакції протеолізу проводили при 25°C протягом 1 години в підходящому буфері з 25мкл бульйону *Photorhabdus* (загальний реакційний об'єм 150мкл). Проби також аналізували на присутність та відсутність дітіотреїтолу. Після інкубації додавали рівний об'єм 12% трихлороцтової кислоти до осаду нерозщепленого білка. Після осадження протягом 0,5 години та подальшого центрифугування 100мкл супернатанту приміщували в 96-луночний мікротитраційний планшет та рН розчину доводили доданням рівного об'єму 4н NaOH. Потім протеоліз визначали кількісно, застосовуючи флуориметричний планшет-рідер Fluoroscan II за довжин хвиль збудження та емісії 485 та 538нм, відповідно. Протеазну активність вимірювали в діапазоні рН5,0 - 10 з приростами 0,5 одиниць. Використовували такі буфери при остаточній концентрації 50мМ: ацетат натрію (рН5,0 - 6,5); Трис-HCl (рН7,0 - 8,0) та біс-Трис-пропан (рН8,5 - 10,0). Для ідентифікації класу протеази (протеаз), що її спостерігали, вихідний бульйон обробляли різними інгібіторами протеаз (кінцева концентрація 0,5мкг/мкл) та потім випробовували на протеазну активність при рН8,0 з застосуванням описаного вище субстрату. Інгібітори протеаз, що їх використовували, включали Е-64 (L-транс-епоксисукциніллейциламід[4-гуанідино]-бутан), 3,4-діхлоризокумарин, Лейпептин, пепстатин, амастатин, етилен-діамінтетраоцтову кислоту (ЕДТК) та 1,10-фенантролін.

Визначення протеази, проведені в діапазоні рН, виявили, що дійсно наявна протеаза (або протеази), що виявляють максимальну активність при рН ~8,0 (Таблиця 16). Додання ДТТ не впливало на протеазну активність. Потім вихідний бульйон обробляли різними інгібіторами протеаз (Таблиця 17). Обробка вихідного бульйону описаними вище інгібіторами виявила, що 1,10-фенантролін викликав повне інгібування всієї протеазної активності при доданні при остаточній концентрації 50мкг, з IC₅₀=5мкг в 100мкл вихідного розчину бульйону 2мг/мл. Ці дані вказують на те, що наявною у найбільшій кількості протеазою (протеазами) в бульйоні *Photorhabdus* є ферменти класу цинкметалопротеаз.

Таблиця 16

Вплив рН на протеазну активність, виявлену в продуктах 1-го дня *Photorhabdus luminescens* (штам W-14)

рН	Флуор. одиниці ^а Активність ^б	Відсоток
5,0	3013±78	17
5,5	7994±448	45
6,0	12965±483	74
6,5	14390±1291	82
7,0	14386±1287	82
7,5	14135±198	80
8,0	17582±831	100
8,5	16183±953	92
9,0	16795±760	96
9,5	16279±1022	93
10,0	15225±210	87

^а Флуор. одиниці = Одиниці флуоресценції (Максимум =~28000; тло =~2200).

^б Активність у відсотках відносно до максимуму при рН8,0

Таблиця 17

Вплив різних. Інгібіторів протеаз на протеазну активність при рН8, виявлену в продуктах 1-го дня *Photorhabdus luminescens* (штам W-14)

Інгібітор	Флуор. од. ^а , які було скоректовано	% інгібування ^б
Контроль	13053	0

Е-64	14259	0
1,10-Фенантролін ^с	15	99
3,4-Діхлорізокумарин ^д	7956	39
Лейпептин	13074	0
Пепстатин ^с	13441	0
Амастатин	12474	4
Контроль з ДМСО	12005	8
Контроль з метанолом	12125	7
^а Флуор. одиниці, що було скоректовано = Одиниці флуоресценції - тло (2200 флуор. одиниць) ^б Відсоток інгібування відносно протеазної активності при рН 8,0 ^с Інгібітори розчиняли в метанолі ^д Інгібітори розчиняли в ДМСО		

Виділення цинк-металопротеази виконували шляхом нанесення діалізованого отриманого осадженням 10 - 80% сульфатом амонію осаду на колонку Q Sepharose, зрівноважену 50мМ Na₂PO₄, рН7,0, як описано в Прикладі 5 для токсину Photorhabdus. Після інтенсивного промивання використовували градієнт 0 - 0,5М NaCl для елюції білка токсину. Більша частина біологічної активності та білка елюювалася в діапазоні 0,15 - 0,45М NaCl. Однак, було виявлено, що основна протеолітична активність була наявною у фракції 0,25 - 0,35М NaCl з деякою активністю в фракції 0,15 - 0,25М NaCl. Аналіз з застосуванням електрофорезу в ПААГ-ДСН фракції 0,25 - 0,35М NaCl показав основну пептидну смугу приблизно 60кДа. Фракція 0,15 - 0,25М NaCl містила схожу смугу 60кДа, але при більш низькій відносній концентрації білка. Подальше гель-фільтрування цієї фракції з застосуванням колонки Superose 12 HR 16/50 дало основний пік, що мігрував при 57,5кДа, який містив переважну (>90% всього забарвленого білка) смугу 58,5кДа при аналізі з застосуванням електрофорезу в ПААГ-ДСН. Додатковий аналіз цієї фракції з застосуванням різноманітних інгібіторів протеаз, описаних вище, показав, що ця протеаза була цинкметалопротеазою. Майже уся протеазна активність, наявна в бульйоні Photorhabdus при 1-ому дні ферментації, відповідала цинк-металопротеазі ~58кДа.

В другому виділенні цинк-металопротеази (протеаз) брали штам W-14 Photorhabdus luminescens, що зростає протягом 3 днів, та протеазну активність візуалізували за допомогою гель-електрофорезу в поліакриламідному гелі з додецилсульфатом натрію (ПААГ-ДСН), з доданням желатина, як описано [Schmidt, t. m., Bleakley, B. and Neelson, K. M. 1988]. Розділяючи гелі, що використовувалися з ДСН (5,5×8см), готували з 12,5% поліакриламідом (40% вихідний розчин акриламід/бісакриламід; Sigma Chemical 3i., St. Louis MO), до якого включали желатин при кінцевій концентрації 1% (Biorad EIA grade reagent; Richmond CA) при розчиненні у воді. Концентруючі ДСН-гелі (1,0×8см) готували з 5% поліакриламідом, також з включенням 0,1% желатина. Як правило, 2,5мкг білка, який піддавали тестуванню, розбавляли в 0,03мл буферу для нанесення для електрофорезу в ПААГ-ДСН без дітіотрехтолу (ДТТ) та наносили на гель. Білки розділяли в ДСН-буфері для розділення (Laemmli, U.K. 1970. Nature 227, 680) при 0°С та при 8мА. Після завершення електрофорезу гель промивали протягом 2 годин в 2,5% (об./об.) Тритоні Х-100. Потім гелі інкубували протягом 1 години при 37 °С в 0,1М гліцині (рН8,0). Після інкубації гелі фіксували та забарвлювали протягом ночі 0,1% амідом чорним в суміші метанол-оцтова кислота - вода (30:10:60, об./об./об.; Sigma Chemical Co.). Протеазну активність візуалізували у вигляді світлих зон проти темного, забарвленого амідом чорним тла, зумовлених протеолізом та подальшою дифузією включеного желатина. Спостерігали принаймні три окремі смуги, що продукувались протеолітичною активністю при 58, 41 та 38кДа.

Аналізи активності різноманітних протеаз в культуральному бульйоні штаму W-14 на третій день ферментації проводили з застосуванням ФІТЦ-казеїну, розчиненого у воді, як субстрату (кінцева концентрація в тесті 0,02%). Експерименти з протеолізу проводили при 37°С протягом 0 - 0,5г в Трис-НСІ (рН8,0) з різними білковими фракціями в загальному об'ємі 0,15мл. Реакції зупиняли доданням рівного об'єму 12% трихлороцтової кислоти (ТХК), розчиненої у воді. Після інкубації при кімнатній температурі протягом 0,25г проби центрифугували при 10000×g протягом 0,25г та брали аліквоти по 0,10мл й приміщували в 96-луночні мікротитраційні планшети. Потім розчин нейтралізували доданням рівного об'єму 2н гідроксиду натрію та кількісні вимірювання виконували з використанням флуориметричного планшет-рідера з довжинами хвиль збудження та емісії 485 та 538нм, відповідно. Вимірювання активності проводили з ФІТЦ-казеїном з різними концентраціями протеази при 37 °С протягом 0-10 хвилин. Одиницю активності довільно встановлювали як кількість ферменту, необхідну для продукування 1000 одиниць флуоресценції на хвилину, а питому активність визначали як одиниці/мг протеази. Дослідження з інгібування проводили з застосуванням двох інгібіторів цинк-металопротеаз: 1,10-фенантроліну та N-(а-рамнопіранозилоксигідроксифосфініл)-Leu-Trp (фосфорамідону) з вихідними розчинами інгібіторів, розчинених в 100% етанолі та воді, відповідно. Вихідні концентрації порівнювали, як правило, 10мг/мл та 5мг/мл для 1,10-фенантроліну та фосфорамідону, відповідно, з кінцевими концентраціями інгібітору 0,5-1,0мг/мл реакційної суміші. Обробка триденного вихідного бульйону W-14 1,10-фенантроліном, інгібітором усіх цинк-металопротеаз, призводила до повної елімінації всієї протеазної активності, тоді як обробка фосфорамідонм, інгібітором термолізін-подібних протеаз [Weaver, L.H., Kester, W.R., and Matthews, B.W. 1977, J.Mol. Biol. 114, 119-132], призводила до ~56% зменшення протеазної активності. Протеолітична активність, що залишалась, не зменшувалась далі при додатковому додаванні фосфорамідону. Протеази триденного бульйону W-14 Photorhabdus очищували так: 4,0 літри бульйону концентрували за допомогою спірального ультрафільтраційного картриджа Amicon типу S1Y100, який

приєднано до фільтраційного пристрою Amicon M-12. Матеріал, що проходив, який мав нативні білки розміром менш 100кДа (3,8л) концентрували до 0,375л за допомогою означеного ультрафільтраційного картриджа. Матеріал, що залишався на фільтрі, містив білки в межах розмірів 10 - 100кДа. Цей матеріал наносили на колонку HR16/10 Pharmacia, яку упаковували сильним аніонообмінним матеріалом PerSeptive Biosystem (Framington, MA) Porose® HQ, який зрівноважували в 10мМ буфері з фосфату натрію (pH7,0). Білки наносили на цю колонку при швидкості потоку 5мл/хвил., потім змивали білок, що не зв'язався, буфером до $A_{280}=0,00$. Потім білки елюювали за допомогою градієнта 0 - 1,0М NaCl протягом 40 хвилин при швидкості потоку 7,5мл/хвил. Фракції аналізували на протеазну активність, як зазначено вище, та активні фракції об'єднували. Протеолітично активні фракції розбавляли в 2 рази (за об'ємом) 10мМ фосфатним буфером (pH7,0) та наносили на колонку 10/10 Mono Q Pharmacia, зрівноважену в 10мМ фосфаті натрію. Після промивання колонки буфером до $A_{280}=0,0$ білки елюювали за допомогою градієнта 0 - 0,5М NaCl протягом 1 години при швидкості потоку 2,0мл/хвил. Фракції аналізували на протеазну активність. Фракції, що мали найбільшу кількість фосфорамідон-чутливої протеазної активності, яку було зумовлено протеазою 41/38кДа, *infra*, об'єднували. Було виявлено, що ці фракції елюювались в межах 0,15 - 0,25М NaCl. Фракції, що містили переважну кількість фосфорамідон-нечутливої протеазної активності, протеази 58кДа, також об'єднували. Було виявлено, що ці фракції елюються в межах 0,25 - 0,35М NaCl. Потім фракції фосфорамідон-чутливої протеази концентрували до кінцевого об'єму 0,75мл за допомогою мембрани Biomax-5K NMWL центрифугального фільтруючого пристрою Millipore Ultra-free®. Цей матеріал наносили при швидкості потоку 0,5мл/хвил., на колонку Pharmacia HR 10/30, яку було упаковано Sephadex G-50 (Pharmacia), зрівноважений в фосфатному буфері (pH7,0) (10мМ фосфат натрію-0,1М NaCl). Потім фракції, що мали максимальну фосфорамідон-чутливу активність, об'єднували та центрифугували через мембрану Biomax-50K NMWL центрифугального фільтруючого пристрою Millipore Ultrafree®-15. Аналіз протеолітичної активності, *supra*, показав, що цей матеріал мав тільки фосфорамідон-чутливу протеазну активність. Після об'єднання фосфорамідон-нечутливої протеази, білка 58кДа, її концентрували в тому самому пристрої та далі розділяли на колонці Superdex-75 (Pharmacia). Фракції, що містили цю протеазу, об'єднували.

Аналіз очищених протеаз 58кДа та 41/38кДа виявив, що, у той час як обидва типи протеази повністю інгібувались 1,10-фенантроліном, тільки протеаза 41/38кДа інгібувалась фосфорамідонем. Подальший аналіз вихідного бульйону показав, що протеазна активність одностенного бульйону W-14 мала 23% загальної протеазної активності, зумовленої протеазою 41/38кДа, з збільшенням до 44% в триденному бульйоні W-14. Стандартний аналіз за допомогою електрофорезу в ПААГ-ДСН для тесту чистоти білка та отримання аміно-кінцевої послідовності, проводили з використанням 4-20% градієнтних гелів (MiniPlus SeptraGels), придбаних в Integrated Separation Systems (Natick, MA). Білки, які треба було піддати аміно-кінцевому секвенуванню, приміщували блотингом на мембрану PVDF після очищення, *infra*, (ProDlott™ Membranes; Applied Biosystems, Foster City, CA), візуалізували за допомогою амідно чорного, вирізали та посиляли до Cambridge Prochem; Cambridge, MA, для секвенування.

Розшифрована аміно-кінцева послідовність протеаз 58кДа (SEQ ID №45) та 41/38кДа (SEQ ID №44) з бульйону триденного W-14 була: DV-GSEKANEKCLK (SEQ ID №45) та DSGDDDKVTNTDIHR (SEQ ID №44), відповідно. Секвенування протеази 41/38кДа виявило декілька аміно-кінців, кожен з яких мав додаткову амінокислоту, яку видаляв протеолізом. Випробування первинної, вторинної, третинної та четвертинної послідовностей для поліпептидів 38 та 41кДа дозволило розшифрувати наведену вище послідовність та виявило, що ці дві протеази є гомологічними.

Приклад 11, Частина А

Скринінг геномної бібліотеки Photorhabdus з використанням антитіл проти генів, що кодуєть пептид TcbA Паралельно до описаного вище секвенування відповідне зондування та секвенування проводили на пептиді TcbA_{ii} (SEQ ID №1). Це секвенування проводили шляхом отримання бактеріальних культуральних бульйонів та очистки токсину, як описано в Прикладах 1 та 2 вище. Геномну ДНК виділяли з штаму W-14 Photorhabdus luminescens, що зростає на середовищі Грейса для культури тканини комах. Бактерії вирощували в 5мл культурального середовища в колбі Ерленмейера на 250мл при 28°C та 250об./хвил., протягом приблизно 24 годин. Бактеріальні клітини з 100мл культурального середовища осаджували при 5000×g протягом 10 хвилин. Супернатант викидали і потім осад клітин використовували для виділення геномної ДНК.

Геномну ДНК виділяли за допомогою модифікації способу ЦТАБ, який описано в Розділі 2.4.3 Ausubel (*supra*). Розділ, що має заголовок "Large Scale CsCl prep of bacterial genomic DNA", використовували до стадії 6 (включно). В цьому пункті проводили додаткову екстракцію ДНК сумішшю хлороформ/ізоаміловий спирт (24:1) з наступною стадією екстракції сумішшю фенол/хлороформ/ізоаміловий спирт (25:24:1) та кінцевою стадією екстракції сумішшю хлороформ/ізоаміловий спирт (24:1). ДНК осаджували додаванням 0,6 об'єму ізопропанолу. Осаджену ДНК зачіпляли та намотували на кінець зігнутої скляної палички, занурювали ненадовго в 70% етанол як кінцеве промивання та розчиняли в 3мл буферу TE.

Концентрація ДНК, яку було визначено за оптичною густиною при 280/260нм, дорівнювала приблизно 2мг/мл. З використанням цієї геномної ДНК одержували геномну бібліотеку. Приблизно 50мкг геномної ДНК частково розщепляли рестриктазою Sau3 A1. Потім використовували центрифугування в градієнті NaCl для фракціонування за розміром частково розщеплених фрагментів ДНК. Фракції, що містили фрагменти ДНК зі середнім розміром 12т.п.н. або більше, як було визначено за допомогою електрофорезу в агарозному гелі, лігували в плазмиду Bluescript, Stratagene, La Jolla, California, та трансформували в штам DH5a або DHB10E. coli.

Окремо, очищені аліквоти білка посиляли до центру біотехнології гібридом при Вісконсинському університеті

(Madison) для отримання моноклональних антитіл до цих білків. Матеріал, що було послано, являв собою очищену за допомогою ВЕРХ фракцію, яка містила нативні смуги 1 та 2, що їх було денатуровано при 65°C, та 20мкг кожної ін'єкували в кожну з чотирьох мишей. Стабільні клітинні лінії, що продукують моноклональні антитіла одержували після зливання клітин селезінки з неімунізованих мишей зі стабільною лінією мієломних клітин. Моноклональні антитіла здобували з гібридом.

Окремо, одержували поліклональні антитіла, беручи білок смуги 1, очищений на нативному агарозному гелі (див. Приклад 1), який потім використовували для імунізації Новозеландського білого кролика. Білок одержували вирізанням смуги з нативних агарозних гелів, недовгим нагріванням шматочків гелю до 65°C для розплавлення агарози та негайним емульгуванням з ад'ювантом. Повний ад'ювант Фрейда використовували для первинних імунізацій та неповний ад'ювант Фрейнда використовували для 3 додаткових ін'єкцій з місячними інтервалами. Для кожного введення, приблизно 0,2мл емульгованої смуги 1, що містить 50 - 100мкг білка, доставляли шляхом численних підшкірних ін'єкцій в спину кролика. Сироватку одержували за 10 днів після останньої ін'єкції та додаткові кровопускання виконували з тижневими інтервалами протягом 3 тижнів. Сироватковий комплемент інактивували нагріванням до 56°C протягом 15 хвилин та потім зберігали при -20°C. Потім моноклональні та поліклональні антитіла використовували для скринінгу геномної бібліотеки на експресію антигенів, які могли б детектуватися за епітопом. Позитивні клони виявляли на нітроцелюлозних фільтрах, що застосовувались для підйому колоній. Було вжито імуноблотингу позитивних клонів.

Аналіз клонів, що визначалися як імуноблотингом, так і аналізом за Саузерном, спричинив експериментальну ідентифікацію п'яти класів клонів. В клоні першого класу був ген, який кодує пептид, що називається тут TcbA_{ij}. Було отримано повну послідовність ДНК цього гена (TcbA). Її представлено як SEQ ID №11. Підтвердження того, що ця послідовність кодує внутрішню послідовність SEQ ID №1, демонструється наявністю SEQ ID №1 у позиції амінокислот 88 від розшифрованої амінокислотної послідовності, яку створює відкрита рамка зчитування SEQ ID №11. Це може бути підтверджено посиланням на SEQ ID №12, що є розшифрованою амінокислотою послідовністю, що створюється SEQ ID №11. Другий клас пептидів токсину включає сегменти, що називаються вище як TsaB_i, TsaB_{ij} та TsaC. Після скринінг у бібліотеки поліклональними антисироватками цей другий клас генів токсину ідентифікували у вигляді кількох клонів, що продукували різні за розміром білки, які всі перехресно реагували з поліклональними антитілами на імуноблоті та, як було виявлено, мали гомологію ДНК на блоті в аналізі за Саузерном. Порівняння послідовностей виявило, що вони відносяться до комплексу генів, що називається вище TsaB та TsaC. Три інші класи клонів також виділили при поліклональному скринінгу. Ці класи продукували білки, що перехресно реагували з поліклональними антитілами та також мали гомологію ДНК з цими класами, як було визначено блотингом за Саузерном.

Ці класи було названо Клас III, Клас IV та Клас V. Також можна виявити моноклональні антитіла, що перехресно реагують з Класами I, II, III та IV. Це дозволяє припустити, що всі вони мають високу білкову гомологію. Таким чином, очевидно, гени екстрацелюлярних білків *Phototribdus luminescens* являють собою сімейство генів, що є еволюційно спорідненими.

Для подальшої розробки концепції, що можуть існувати еволюційно споріднені варіації в пептидах токсину, які містяться в цьому організмі, було вжито два підходи для дослідження інших штамів *Phototribdus luminescens* на присутність споріднених білків. Це було зроблено за допомогою ПЛР-ампліфікації геномної ДНК та за допомогою імуноблотингу з використанням поліклональних та моноклональних антитіл.

Ці результати показують, що споріднені білки продукуються штамми *Phototribdus luminescens* WX-2, WX-3, WX-4, WX-5, WX-6, WX-7, WX-8, WX-11, WX-12, WX-15 та WX-14.

Приклад 11. Частина В

Секвенування та аналіз клонів токсину Класу III-tcc

Подальше секвенування ДНК проводили на плазмидях, які було виділено з клонів *E. coli* Класу III, описаних в Прикладі 11, Чащині А. Було показано, що нуклеотидна послідовність являла собою три тісно зв'язані відкриті рамки зчитування в цьому генному локусі. Цей локус було названо tcc, а три відкриті рамки зчитування було названо tccA SEQ ID №56, tccB SEQ ID №58 та tccC SEQ ID №60 (Фіг.6В). Розшифрована амінокислотна послідовність з відкритої рамки зчитування tccA показує, що цей ген кодує білок 105459Да. Цей білок було названо TccA. Перші 12 амінокислот цього білка збігаються з N-кінцевою послідовністю, яку було одержано з білка 108кДа SEQ ID №7, що її було ідентифіковано раніше як частину комплексу токсину.

Розшифрована амінокислотна послідовність з відкритої рамки зчитування tccB показує, що цей ген кодує білок 175716Да. Цей білок було названо TccB. Перші 11 амінокислот цього білка збігаються з N-кінцевою послідовністю, яку було отримано з білка з визначеною молекулярною масою 185кДа SEQ ID №8.

Розшифрована амінокислотна послідовність tccC показує, що ця відкрита рамка зчитування кодує білок 111694Да й цей білковий продукт було названо TccC.

Приклад 12

Характеристика штамів *Phototribdus*

Для підтвердження того, що колекція, яку описано тут, складається з штамів *Phototribdus*, ці штами оцінювали за мікробіологічними ознаками, які є характерними для *Phototribdus* та відрізняють його від інших видів *Enterobacteriaceae* та від видів *Xenorhabdus* [Farmer, J.J. 1984. *Bergey's Manual of Systemic Bacteriology*, vol. I, pp.510 - 511. (ed. Kreig N.R. and Holt, J.G.). Williams and Wilkins, Baltimore.; Akhurst and Boemare, 1988, Boemare et al., 1993]. Цими характерними ознаками є: грамнегативні палички, розмір організму 0,5 - 2мкМ завширшки та 2 - 10мкМ завдовжки, червона/жовта пігментація колоній, наявність кристалічних тіл включення, наявність каталази, неспроможність відновлювати нітрати, наявність біolumінесценції, спроможність поглинати барвник з середовищ для вирощування, позитивна реакція на

5 продукування протеази, температурна межа зростання нижче 37 °С, виживання за анаеробних умов та рухливість (Таблиця 18). Стандартні штами *Escherichia coli*, *Xenorhabdus* та *Photorhabdus* включали до всіх тестів для порівняння. Всі результати погоджуються з тим, що всі штами належать до родини *Enterobacteriaceae* та роду *Photorhabdus*.

Люменометр використовували для виявлення біоломінесценції кожного штаму та забезпечення кількісного та відносного вимірювання утворення світла. Для вимірювання відносних одиниць емісії світла бульйони від кожного штаму (клітини та середовища) вимірювали при трьох часових інтервалах після інокулювання в рідкій культурі (6, 12 та 24г) та порівнювали з фоновою освітленістю (не інокульовані середовища та вода). Перед вимірюванням емісії світла з різних бульйонів густину клітин визначали шляхом вимірювання поглинання світла (560нм) в спектрофотометрі Gilford Systems (Oberlin, OH) з використанням sipper-елементу. Потім робили підхожі розведення (для нормалізації оптичної густини до 1,0 одиниці) перед вимірюванням освітленості. Потім аліквоти бульйонів, які було розведено, приміщували в кювети (300мкл кожна) та знімали показання люменометру Bio-Orbit 1251 (Bio-Orbit Oy, Twiki, Finland). Період інтеграції для кожної проби дорівнював 45 секунд. Проби безупинно перемішували (крутили в кюветах з перегородками) під час зняття показань для забезпечення доступу кисню. Позитивну відповідь було визначено як ≥ 5 -разову фонову люмінесценцію (~5 - 10 одиниць). Крім того, освітленість колоній виявляли шляхом накладання на фотографічні плівки та візуально, після адаптації в темній камері. Характеристики забарвлення за Грамом кожного штаму визначали з комерційним набором для забарвлення за Грамом (BBL, Cockeysville, MD), що використовувався разом з контрольними слайдами забарвлення за Грамом (Fischer Scientific, Pittsburg, PA). Потім проводили мікроскопічну оцінку з використанням мікроскопу Цейса (Carl Zeiss, Germany) з 100х масляним імерсійним об'єктивом (з збільшенням окуляру 10х та збільшенням корпусу 2х). Дослідження під мікроскопом розмірів окремих штамів організму, опис клітин та тіл включення (останнє - після логарифмічної фази зростання) виконували з застосуванням вологих мікроскопічних препаратів (wet mount slides) (10х збільшення окуляру, 2х збільшення корпусу та 40х збільшення об'єктиву) за допомогою мікроскопії з масляною імерсією та фазово-контрасної мікроскопії з мікрометром [Akhurst, R.J. and Boemare, N.E. 1990. Entomopathogenic Nematodes in Biological Control (ed. Gaugler, R. and Kaya, H.). pp.75-90. CRS Press, Boca Raton, USA; Baghdiguian S., Boyer-Giglio M.H., Thaler, J.O., Bonnot G., Boemare M. 1993. Biol. Cell 79, 177-185]. Пігментацію колоній спостерігали після інокулювання на бактоагар (Bacto nutrient agar, Difco Laboratories, Detroit, MI), який було приготовано відповідно до інструкцій, що додавалися. Інкубування відбувалося при 28 °С, та описи робили після 5 - 7 днів. Для тестування на наявність ферменту каталази колонію організму, який тестували, видаляли на маленькій пробці з чашки з живильним агаром та приміщували на дно скляної тест-пробірки. Один мл розчину пероксиду водню, приготованого в лабораторії, обережно додавали по боковій стінці пробірки. Позитивну реакцію реєстрували, коли відразу ж або в межах 5 секунд з'являлися бульки газу (як припускалось - кисню). Досліджували також контролю з неінокульованим живильним агаром та розчином пероксиду водню. Для визначення відновлення нітратів кожен культуру інокулювали в 10мл Бакто-нітратного Бульйону (Difco Laboratories, Detroit, MI). Після 24 годин інкубації при 28°C визначали утворення нітритів шляхом додання двох крапель реагенту з сульфаніловою кислотою та двох крапель альфа-нафтиламінового реагенту [див. Difco Manual, 10th edition, Difco Laboratories, Detroit, MI, 1984]. Утворення чіткого рожевого або червоного забарвлення вказує на утворення нітриту з нітрату. Спроможність кожного штаму поглинати барвник з середовища для вирощування визначали з бактоагаром MacConkey, що містив барвник нейтральний червоний; з бактоагаром Tergitol-7, що містив барвник бромтімол синій, та бактоагаром EMB, що містив барвник еозин-Y (агари з Difco Laboratories, Detroit, MI, всі було приготовано відповідно до інструкцій, що додавались). Після інокулювання на ці середовища поглинання барвника реєстрували після інкубування при 28 °С протягом 5 днів. Зростання на цих середовищах є характеристикою для членів родини *Enterobacteriaceae*. Рухливість кожного штаму визначали з використанням Тест-середовища для визначення рухливості (Bacto Motility Test Medium (Difco Laboratories, Detroit, MI), яку було приготовано відповідно до інструкцій, що додавались. Проводили уколочне інокулюванням кожним штамом та рухливість оцінювали макроскопічно за дифузною зоною зростання, що розповсюджується від лінії інокулята. В багатьох випадках рухливість також спостерігали під мікроскопом з рідкої культури на вологих препаратах (wet mount slides). Біохімічну живильну оцінку для кожного штаму виконували з застосуванням BBL Enterotube II (Benton, Dickinson, Germany). Дотримувались рекомендацій, що супроводжують цей продукт, за винятком того, що інкубацію проводили при 28°C протягом 5 днів. Результати погоджувались з раніше наведеними даними для *Photorhabdus*. Утворення протеази визначали спостереженням гідролізу желатина з використанням бактожелатину (Difco Laboratories, Detroit, MI), чашки готували відповідно до інструкцій, що додавались. Культури інокулювали та чашки інкубували при 28°C протягом 5 днів. Для оцінки зростання при різноманітних температурах чашки з агаром [2% proteose pepton №3 з 2Z бактоагаром (Difco Laboratories, Detroit, MI) в деіонізованій воді] засіювали штрихом з звичайного джерела інокулята. Чашки герметизували Nesco® film та інкубували при 20, 28 та 37°C протягом періоду до трьох тижнів. Чашки, що не виявляли росту при 37 °С, не виявляли життєздатності після їх перенесення до термостату на 28 °С протягом одного тижня. Потреба в кисні для штамів *Photorhabdus* тестували так. Проводили уколочне інокулювання у рідке бульйонне середовище, що містило тіогліколат (Difco Laboratories, Detroit, MI). Трубки інкубували при кімнатній температурі протягом тижня та потім культури випробовували на тип та ступінь зростання. Індикатор резазурин демонструє рівень помірного окислення або зону аеробіозу (Difco Manual, 10-th edition, Difco Laboratories, Detroit, MI). Результати щодо зони зростання, отримані для штамів *Photorhabdus*, які було тестовано, відповідають результатам для факультативного анаеробного мікроорганізму.

5

Таксономічні ознаки *Photorhabdus luminescens*

Ознаки, що їх оцінюють*

10

Штам	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
W-14	+	±	±	rd S	±	=	±	±	±	Q	±	±	±	±	±	±	=

15

A= граммпозитивний штам, B= кристалічні тіла включення, C= біолюмінесценція, D= форма клітини, E= рухливість, F= відновлення нітратів, G= наявність каталази, H= гідроліз желатина, I= поглинання барвника, J= пігментація, K= зростання на EMB-агарі, L= зростання на агарі MacConkey, M= зростання на агарі Tergitol-7, N= факультативний анаероб, O= зростання при 20°C, P= зростання при 28°C, Q= зростання при 37°C;

20

25

30

35

40

45

50

55

60

65

U A 8 2 4 8 5 C 2

C 2 8 2 4 8 5 U A

Итам	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
WX-1	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-2	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-3	-	+	+	rd S	+	-	+	+	+	YT	+	+	+	+	+	+	-
WX-4	-	+	+	rd S	+	-	+	+	+	YT	+	+	+	+	+	+	-
WX-5	-	+	+	rd S	+	-	+	+	+	LO	+	+	+	+	+	+	-
WX-6	-	+	+	rd S	+	-	+	+	+	LY	+	+	+	+	+	+	-
WX-7	-	+	+	rd S	+	-	+	+	+	R	+	+	+	+	+	+	-
WX-8	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-9	-	+	+	rd S	+	-	+	+	+	YT	+	+	+	+	+	+	-
WX-10	-	+	+	rd S	+	-	+	+	+	RO	+	+	+	+	+	+	-
WX-11	-	+	+	rd S	+	-	+	+	+	RO	+	+	+	+	+	+	-
WX-12	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
WX-14	-	+	+	rd S	+	-	+	+	+	LR	+	+	+	+	+	+	-
WX-15	-	+	+	rd S	+	-	+	+	+	LR	+	+	+	+	+	+	-
H9	-	+	+	rd S	+	-	+	+	+	LY	+	+	+	+	+	+	-
Hb	-	+	+	rd S	+	-	+	+	+	YT	+	+	+	+	+	+	-
Hm	-	+	+	rd S	+	-	+	+	+	TY	+	+	+	+	+	+	-
HP88	-	+	+	rd S	+	-	+	+	+	LY	+	+	+	+	+	+	-
NC-1	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
W30	-	+	+	rd S	+	-	+	+	+	YT	+	+	+	+	+	+	-
WIR	-	+	+	rd S	+	-	+	+	+	RO	+	+	+	+	+	+	-
B2	-	+	+	rd S	+	-	+	+	+	R	+	+	+	+	+	+	-
4398	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-

+/-= позитивний чи негативний щодо ознаки, rd= паличка, S= підходить за розміром до цього роду, RO= червоно-оранжева, LR= ясно-червона, R= червона, O= оранжева, Y= жовта, T= рудувато-коричнева, LY= ясно-жовта, YT= жовто-рудувато-коричнева, LO= ясно-оранжева.

Итам	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
4399	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
4390	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
4391	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-
4392	-	+	+	rd S	+	-	+	+	+	O	+	+	+	+	+	+	-

Аналіз жирних кислот клітини є визнаним інструментом бактеріальної характеристики на рівні роду та виду [Tornabene T.G. 1985. Lipid Analysis and Relationship to Chemotaxonomy in Methods in Microbiology. Vol.18, 209-224; Goodfellow, M. and O'Donnel, A.G. 1993. Roots of Bacterial Systematics (ed. Goodfellow, M. and O'Daniel, A.G.) pp.3-54. London: Academic Press Ltd.], що їх включено тут як посилання та використано для підтвердження того, що наша колекція має спорідненість на рівні роду. Культури було відправлено до іншої, контрактної лабораторії для аналізу метилових ефірів жирних кислот (FAME) з застосуванням Системи ідентифікації мікробів (MIS) з Microbial ID (MIDI, Newark, USA). Система MIS складається з газового хроматографу Hewlett Packard HP5890A з колонкою з сплавлених 5% метилфенілсиліконом капілярів, що складаються з діоксиду кремнію (25мм×0,2мм). Як газ-носії використовують водень та детектор на основі плазмової іонізації працює разом з автоматичним пробовідбірником, інтегратором та комп'ютером. Комп'ютер порівнює метилові ефіри жирних кислот проб з мікробною бібліотекою жирних кислот та порівнює з калібровальною сумішшю відомих жирних кислот. За розсудом контрактної лабораторії штами росли протягом 24 годин при 28°C на триптиказному соєвому агарі перед аналізом. Екстракцію проб в контрактній лабораторії проводили за стандартною методологією FAME.

Не було прямої ідентифікації штамів відносно до будь-якої люмінесцентної бактеріальної групи, іншої, ніж *Photorhabdus*. При виконанні кластерного аналізу, що порівнює профілі жирних кислот групи ізолятів, профілі жирних кислот штамів були спорідненими на рівні роду.

Еволюційна різноманітність штамів *Photorhabdus* в нашій колекції вимірювали шляхом аналізу опосередкованого ПЛР (Полімеразною Ланцюговою Реакцією) геномного фінгерпрінтинга ("відбитку пальців") з використанням геномної ДНК з кожного штаму. Цей спосіб засновано на родинах послідовностей ДНК, що повторюються, наявних по всьому геному бактеріальних видів, які відрізняються [див. огляд Versalovic, J., Schneider, M., De Bruijn, F.J. and Lupski, J.R. 1994. Methods Mol. Cell. Biol., 5, 25-40]. Вважають, що три з них, екстрагенна паліндромна послідовність, що повторюється, (REP), ентеробактеріальний міжгенний консенсус, що повторюється, (ERIC) та BOX-елемент, віграють важливу роль в організації бактеріального геному. Вважається, що організація геному створюється відбором, та диференційна дисперсія (розкид) цих елементів в геномі близькоспоріднених бактеріальних штамів, може бути використано для відрізнєння цих штамів [наприклад, Louws, F.J., Fulbright, D.W., Stephens, C.T. and DE Bruijn, F.J. 1994. Appl. Environ. Micro. 60, 2286-2295]. Rep-PCR використовує олігонуклеотидні праймери, комплементарні до цих послідовностей, що повторюються, для ампліфікації ДНК-фрагментів з розмірами, які відрізняються, що лежать між ними. Отримані продукти розділяють електрофорезом для встановлення "фінгерпринту" ДНК для кожного штаму.

Для виділення геномної ДНК з наших штамів осад клітин повторно суспендували в буфері TE (10мМ Трис-НСІ, 1мМ ЕДТК, рН8,0) до кінцевого об'єму 10мл та потім додавали 5М NaCl. Цю суміш центрифугували 20 хвилин при 15000×g. Отриманий осад повторно суспендували в 5,7мл TE та додавали 300мкл 10% ДСН та 60мкл 20мг/мл протеїнази К (Gibco BRL Products, Grand Island, NY). Суміш інкубували при 37°C протягом 1г, потім додавали приблизно 10 мг л іозиму та суміш інкубували ще протягом 45 хвилин. Потім додавали один мілілітр 5М NaCl та 800мкл розчину ЦТАБ/NaCl (10% м./об. ЦТАБ, 0,7М NaCl) та суміш інкубували та перемішували ще протягом 20 хвилин для покращання виведення клітинного матеріалу. Додавали рівний об'єм розчину хлороформ/ізоаміловий спирт (24:1об./об.), змішували обережно та потім центрифугували. Потім проводили три екстракції рівним об'ємом суміші фенол/хлороформ/ізоаміловий спирт (50:49:1). Геномну ДНК осаджували 0,6 об'єму ізопропанолу. Осаджену ДНК видаляли скляною паличкою, промивали двічі 70% етанолом, сушили та розчиняли в 2мл СТЕ (10мМ Трис-НСІ, рН8,0, 10мМ NaCl, 1мМ ЕДТК). Потім ДНК кількісно визначали за оптичною густиною при 260нм. Для виконання Rep-PCR-аналізу геномної ДНК *Photorhabdus* використовували такі праймери, REPIR-1: 5'-IIICGICGICATCIGGC-3' та REP2-1: 5'-ICGICTTATCIGGCCTAC-3' ПЛР виконували з використанням об'єму реакції 25мкл: 7,75мкл води, 2,5мкл 10X буферу LA (PanVera Corp., Madison, WI) 16мкл суміші dNTP (2,5мМ кожний), 1мкл кожного праймера при 50пМ/мкл, 1мкл ДМСО, 1,5мкл геномної ДНК (з концентрацією в діапазоні 0,075 - 0,480мкг/мкл) та 0,25мкл TaKaRa EX Taq (PanVera Corp., Madison, WI). ПЛР-ампліфікацію проводили в ДНК-термоциклері Perkin Elmer (Norfolk, CT) з використанням таких умов: 95°C/7 хвилин, потім 35 циклів: 94°C/1хвил., 65°C/8хвил., потім 15хвил. при 65°C. Після проведення циклів реакційну суміш 25мкл додавали до 5мкл 6X буферу для нанесення на гель (0,25% бромфеноловий синій, 40% м./об. сахароза в Н₂О). Потім проводили електрофорез в 1% агарозному гелі 15×20см в буфері ТБЕ (0,09М Трис-НСІ, 0,002М ЕДТК) з використанням 8мкл кожної ПСР-реакції. Електрофорез проводили протягом приблизно 16 годин при 45В. Потім гелі забарвлювали в 20мкг/мл етидйброміда протягом 1 години та обезбарвлювали в буфері ТБЕ протягом приблизно 3 годин. Потім гелі фотографували Полярідом при освітленні ультрафіолетовим світлом.

Наявність або відсутність смуг при специфічних для кожного штаму розмірах оцінювали з цих фотографій та вводили як матриці подібності в цифрову програму таксономії комп'ютера, NTSYS-pc (Exeter Software, Setauket, NY). Контролі, штам РИ101 *E. Coli* та *Xenorhabdus oryzae* pv. *oryzae*, що аналізувалися одночасно, давали "фінгерпринти" ПСР, які відповідали опублікованим даним [Versalovic, J., Koeuth, T. and Lupski, J.R. 1991. Nucleic Acids Res. 19, 6823-6831; Vera Cruz, CM., Halda-Alija, L., Louws, F., Skinner, D.Z., George, M.L., Nelson, R.J., DE Bruijn, F.J., Rice, C. and Leach, J.E. 1995. Int. Rice Res. Notes, 20, 23-24; Vera Cruz, CM., Ardales, E.Y., Skinner, D.Z., Talag, J., Nelson, R.J., Louws, F.J., Leung, H., Mew, T.W. and Leach, J.E. 1996. Phytopathology (in press, respectively)]. Потім дані для штамів *Photorhabdus* аналізували з серією програм в NTSYS-pc; SIMQUAL (Подібність для кількісних даних) для отримання матриці коефіцієнтів подібності (з використанням коефіцієнту Жаккарда) та утворення SAHN-кластерів (Sequential, Agglomerative, Hierarchy)

and Nested) [з застосуванням UPGMA (Unweight Pair-Group Method with Arithmetic Averages)-способу], що групують споріднені штами та можуть бути виражені у вигляді фенограми (Fig.5). Програми SOPH (ко-фенетичних величин) та MXCOMP (порівняння матриць) використовували для отримання матриці ко-фенетичних величин та порівняння кореляції між нею та вихідною матрицею, на якій засновано утворення кластерів. Було отримано результуючу нормалізовану статистику Мантеля (r), що є мірою погодження для кластерного аналізу ($r=0,8-0,9$ означає дуже добре погодження). В нашому випадку $r=0,919$. Таким чином, наша колекція складається з різноманітної групи штамів, які легко розрізнити та які є характерними для роду *Photorhabdus*.

Приклад 13

Придатність як інсектицидів токсину (токсинів), що продукуються штамами *Photorhabdus*

Вихідні культури для посіву різних штамів *Photorhabdus* одержували інокулюванням 175мл 2% Proteose Pepton №3 (PP3) (Difco Laboratories, Detroit, MI) рідкого середовища первинним варіантним субклоном в колбі з трьома перегородками на 500мл з горлом Делонга, закритій Kaput. Інокулят для кожної культури для посіву одержували з залитих маслом культур на скошеному агарі або з культур на чашках. Після інокулювання ці колби інкубували протягом 16 г при 28 °C на ротаційному шейкері при 150об./хвил. Ці культури для посіву використовували потім як однорідні джерела інокулятів для конкретної ферментації кожного штаму. Крім того, покриття культури для посіву пост-логарифмічної фази росту стерильним мінеральним маслом, додання стерильного магнітного стержня для подальшого повторного суспендування та зберігання культури в темряві при кімнатній температурі забезпечували тривале консервування інокулята в токсин-компетентному стані. Продукційні бульйони інокулювали додаванням 1% активно зростаючої посівної культури до свіжого 2% PP3-середовища (наприклад, 1,75мл на 175мл свіжого середовища). Приготування бульйонів відбувалось або в колбах з трьома перегородками на 500мл (див. вище), або в колбах з випуклим дном на 2800мл (об'єм 500мл), закритих пробками з кремнійорганічного пенопласту. Продукційні колби інкубували протягом 24 - 48 годин за описаних вище умов. Після інкубації бульйони розподіляли в стерильні поліетиленові бутлі на 1л, відкручували при 2600xg протягом 1 години при 10°C та декантували з осаду клітин та залишків клітин. Потім рідкий бульйон фільтрували під вакуумом через фільтри Whatman GF/D (2,7мкМ утримання) та скляні фільтри GF/B (1,0мкМ утримання) для вилучення клітинних залишків. Подальше освітлення бульйону досягали за допомогою тангенціального проточного мікрофільтраційного пристрою з відкритими каналами 0,5мкм. При необхідності додаткового освітлення можна досягнути шляхом охолодження бульйону (до 4°C) та центрифугування протягом декількох годин при 2600xg. Після цих процедур бульйон пропускали крізь стерилізуючий фільтр, що використовує нітроцелюлозну мембрану 0,2мкм. Потім стерильні бульйони використовували безпосередньо для біологічного аналізу, біохімічного аналізу або концентрували (максимально в 15 разів) з застосуванням відсікаючого мол. масу 10000 ультрафільтраційного приладу M12 (Amicon, Beverly MA) або центрифугальних концентраторів (Millipore, Bedford, MA та Pall Filtron, Northborough, MA) з розміром пор 10000 (мол. маса). У випадку центрифугальних концентраторів бульйон відкручували при 2000xg протягом приблизно 2 годин. Матеріал, що проходив, (пермеат) з мол. масою 10000 додавали до відповідного матеріалу, що його утримував фільтр, (ретентат) для досягнення бажаної концентрації компонентів з мол. масою більшою за 10000. Інактивацію нагріванням проб обробленого бульйону досягали шляхом нагрівання проб при 100 °C в заповненому піском блоці нагрівання протягом 10 хвилин.

Бульйон (бульйони) та комплекс (комплекси) токсину з різних штамів *Photorhabdus* є застосовними для зменшення популяцій комах та були використані в способі інгібування популяції комах, що передбачає нанесення в осередок комах ефективного для інактивації комах кількості описаного активного компонента. Демонстрацію широти інсектицидної активності, що спостерігається при застосуванні бульйонів вибраної групи штамів *Photorhabdus*, що їх ферментовано, як описано вище, показано в Таблиці 19. Можливо, що можна було б виявити додаткові інсектицидні активності з цими штамами шляхом збільшення концентрації бульйону або шляхом застосування різних способів ферментації. Відповідно до активності, зв'язаної з білком, інсектицидна активність всіх випробуваних штамів була термолабільною (див. вище). Культуральний бульйон (бульйони) різноманітних штамів *Photorhabdus* виявляють диференційну інсектицидну активність (смертність та/або інгібування росту, зменшене вилуплення дорослих особин) у відношенні до личинок блішки довговусої та личинок довгоноса бавовняного, що є членами ряду комах Coleoptera. Інші члени Coleoptera включають дротяників, квітогриза рапсового та колорадського жука. Також спостерігали активність проти цикадок айстр, що є членами ряду Homoptera. Інші члени Homoptera включають цикад, листоблішку грушеву, медяницю звичайну, виноградних червеців та щитівок, білокрилок, spittle-жуків, а також численні специфічні у відношенні до живителя види тлі. Бульйони та очищені комплекси токсину активні також проти совки та точильника зернового, що є членами ряду Lepidoptera. Іншими типовими членами цього ряду є совка мала, совка капустяна, совка іпсилон, совка бавовняна, плодоярка яблунева, справжня міль, вогнівка амбарна південна, метелик-листовійки, мішечниця поденкоподібна, коконопряд кільчастий американський, лучний метелик та совка трав'яна. Активність також спостерігали проти личинок плодової мушки звичайної та личинок комарів, що є членами ряду Diptera. Іншими членами ряду Diptera є галіця горохова, муха морквяна, муха капустяна весняна, муха ріпи, муха цибульна, довгоніг та муха кімнатна й різні види комарів. Активність бульйонів та комплексів токсину спостерігали також проти кліща двопламистого павутинного, що є членом ряду Acarina, який включає також кліщів сунічних, кліщів, кліщика червоного цитрусового, кліщика павутинного волохатого, кліщика грушевого та кліщика-руйнівника.

Активність проти личинок блішки довговусої визначали так. Культуральні бульйони *Photorhabdus* (сконцентровані в 0 - 15 разів, які було пропущено через стерилізуючий фільтр), 2% Proteose Pepton №3,

очищені комплекси токсину [0,23мг/мл] або буфер, що містить 10мМ фосфату натрію, рН7,0, наносили безпосередньо на поверхню (приблизно 1,5см²) штучного корму [Rose, R.I. and McCabe, J.M. (1973). J. Econ. Entomol. 66, (398-400)] у вигляді аліквот по 40мкл. Комплекс токсину розбавляли в 10мМ фосфатному буфері, рН7,0. Чашки з кормом сушили на повітрі в стерильному ламінарному боксі та лунки заражали окремими новонародженими *Diabrotica undecimpunctata howardi* (блішка довговуса південна, SCR), що вилупилися з стерилізованих на поверхні яєць. Чашки герметизували, приміщували в зволожену камеру для вирощування та підтримували при 27°C протягом підходящого періоду часу (3 - 5 днів). Потім оцінювали смертність та масу личинок. Як правило, в усіх дослідженнях використовували 16 комах на варіант. Смертність у контролі як правило була меншою за 5%.

Активність проти довгоносика бавовняного (*Anthonomus grandis*) тестували так. Концентровані (в 1 - 10 разів) бульйони *Photobacterium*, контрольне середовище (2% *Proteose Pepton №3*), очищені комплекси токсину [0,23мг/мл] або 10мМ фосфатний буфер, рН7,0, наносили у вигляді аліквот по 60мкл на поверхню 0,35г штучного корму (корму лускокрилих, білянки з жовтим забарвленням) та давали висохнути. Одну 12 - 24-годинну личинку довгоносика бавовняного приміщували на корм та лунки герметизували та витримували при 25 °С та 50% відносній вологості протягом 5 днів. Потім оцінювали смертність та масу личинок. Смертність у контролі була в діапазоні 0 - 13%.

Активність проти личинок комарів тестували так. Кожна лунка містила 200мкл водного розчину (концентрованих в 10 разів) культуральних бульйонів *Photobacterium*, контрольного середовища (2% *Proteose Pepton №3*), 10мМ фосфатного буферу, комплексів токсину (0,23мг/мл) або води та приблизно одноденні личинки (*Aedes aegypti*). Використовували 6 лунк на варіант. Результати реєстрували при 3 - 4 днях після зараження. Смертність у контролі була між 0 та 20%.

Активність проти плодової мушки звичайної тестували так. Придбане середовище для *Drosophila melanogaster* готували з застосуванням 50% сухого середовища та 50% рідини, такої як вода, контрольне середовище (2% *Proteose Pepton №3*), концентрований в 10 разів культуральний бульйон *Photobacterium*, очищений комплекс токсину [0,23мг/мл] або 10мМ фосфатний буфер, рН7,0. Це виконували приміщенням 4,0мл сухого середовища в кожний з 3 флаконів для вирощування на варіант та додаванням 4,0мл відповідної рідини. Потім 10 личинок *Drosophila melanogaster* пізньої вікової стадії додавали до кожного флакону на 25мл. Флакони витримували на лабораторному столі при кімнатній температурі під світлом флуоресцентних ламп на стелі. Підрахунок ляльок або дорослих особин проводили після 15 днів експозиції. Вилуплення дорослих особин порівнювали з варіантами з водою та контрольним середовищем (0 - 16% зменшення).

Активність проти дорослих особин цикадок айстр (*Macrostelus severini*) та німф цикади кукурудзяної (*Peregrinus maidis*) тестували за допомогою тесту з зараженням через рот, призначеним для проковтування активного інгредієнту без іншого зовнішнього контакту. Резервуар для розчину активного інгредієнта "корма" готували, зробивши 2 отвори в центрі нижньої частини чашки Петрі розміром 35х10мм. Квадрат з парафілму (*Parafilm*®) (2 дюйми) приміщували на верхній частині чашки та закріплювали його "0"-кільцем. Потім пластикову чашку на 1 унцію заражали приблизно 7 цикадами, та резервуар приміщували на цю чашку парафілмом донизу. Потім додавали тест-розчин до резервуару через отвір. В тестах з використанням культурального бульйону (бульйонів), який було сконцентровано в 10 разів, бульйон та контрольне середовище (2% *Proteose Pepton №3*) діалізували проти 10мМ буферу, що містив фосфат натрію, рН7,0, та додавали сахарозу (до 5%) до отриманого розчину для зниження смертності у контролі. Очищені комплекси токсину [0,23мг/мл] або 10мМ фосфатний буфер, рН7,0, також випробовували. Смертність реєстрували на 3-й день. Аналіз проводили в термостаті при 28°C, 70% відносній вологості з фотоперіодом 16/8. Оцінку смертності реєстрували при 72 годинах. Смертність у контролі була меншою за 6%.

Активність проти личинок лускокрилих комах тестували так. Сконцентрований (в 10 разів) культуральний бульйон (бульйони) *Photobacterium*, контрольне середовище (2% *Proteose Pepton №3*), очищений комплекс токсину [0,23мг/мл] або 10мМ буфер, що містить фосфат натрію, рН7,0, наносили безпосередньо на поверхню (~1,5см) стандартного штучного корму для лускокрилих (корм *Stoneville Yellow*) у вигляді аліквот по 40мкл. Планшетам з кормом давали висохнути на повітрі в стерильному ламінарі та кожну лунку заражали однією новонародженою личинкою. Яйця тоїчильника зернового (*Ostrinia nubilalis*) та бражника (*Manduca sexta*) одержували з комерційних джерел та вилуплення проводили в лабораторії, тоді як личинок совки одержували самі. Після зараження личинкою планшети з кормом герметизували, приміщували в зволожену камеру для вирощування та тримали в темряві при 27°C протягом підходящого періоду. Смертність та масу оцінювали на 5-й день. Як правило, в усіх дослідженнях використовували 16 комах на варіант. Смертність у контролі була між 4 та 12,5% для контрольного середовища та менше 10% для фосфатного буферу.

Активність проти кліща двоплямистого павутинного (*Tetranychus urticae*) визначали так. Молоді рослини гарбузу обрізали до однієї сім'ядолі та обприскували до стікання сконцентрованим в 10 разів бульйоном, контрольним середовищем (2% *Proteose pepton №3*), очищеним комплексом токсину [0,23мг/мл] або 10мМ фосфатним буфером, рН7,0. Після висушування рослини заражали змішаною популяцією павутинних кліщів та витримували при температурі та вологості лабораторії протягом 72 годин. Потім робили підрахунок живих кліщів для визначення рівня контролю.

Таблица 19	
Інсектицидний спектр бульйонів від різноманітних штамів <i>Photobacterium</i> , що його спостерігали	
Штам <i>Photobacterium</i>	Види комах, що є чутливими*

WX-1	3**,4,5,6,7,8
WX-2	2,4
WX-3	1,4
WX-4	1,4

*=≥25% смертність та/або інгібування росту порівняно з контролем
 **=1: совка; 2: точильник зерновий; 3: бражник; 4: блішка довговуса; 5: довгоносик бавовняний; 6: комар; 7: плодова мушка звичайна (дрозофіла); 8: цикада айстр; 9: цикада кукурудзяна; 10: кліщ двоплямистий павутинний

Штам Photorhabdus	Види комах, що є чутливими*
WX-5	4
WX-6	4
WX-7	3,4,5,6,7,8
WX-8	1,2,4
WX-9	1,2,4
WX-10	4
WX-11	1,2,4
WX-12	2,4,5,6,7,8
WX-14	1,2,4
WX-15	1,2,4
W30	3,4,5,8
NC-1	1,2,3,4,5,6,7,8,9
WIR	2,3,5,6,7,8
HP88	1,3,4,5,7,8
Hb	3,4,5,7,8
Hm	1,2,3,4,5,7,8
H9	1,2,3,4,5,6,7,8
W-14	1,2,3,4,5,6,7,8,10
ATCC 43948	4
ATCC 43949	4
ATCC 43950	4
ATCC 43951	4
ATCC 43952	4

Приклад 14

Штами Photorhabdus, що не є W-14:

Очищення, характеристика та спектр активності

Очищення

Протокол, що його викладено далі, є подібним до розробленого для очищення W-14 та заснований на очищенні фракцій, які мають найбільшу активність проти блішки довговусої південної (SCR), як визначено в біотестах (див. Приклад 13). Як правило одержували 4 - 20л бульйону, який було профільтровано, як описано в Прикладі 13, та концентрували його за допомогою спірального ультрафільтраційного картриджа типу SIYIOO (Amicon), який було прикріплено до фільтраційного пристрою Amicon M-12. Матеріал, що утримувався, містив нативні білки з мол. масами більше 100кДа, тоді як матеріал, що проходив, містив нативні білки розміром менше 100кДа. Більша частина активності проти SCR містилася в матеріалі 100кДа, що утримувався. Потім матеріал, що утримувався, безупинно діалізували з 10мМ фосфатним буфером (pH=7,0) до тих пір, поки фільтрат не досягав $A_{280} < 0,100$. Якщо немає інших вказівок, всі процедури до цього моменту виконували в буфері, такому як 10мМ буфер, що містить фосфат натрію (pH7,9). Потім матеріал, що утримувався, концентрували до кінцевого об'єму приблизно 0,20л та фільтрували за допомогою стерилізуючого фільтруючого елементу 0,45мм Nalgene Filterware. Матеріал, який було відфільтровано, наносили при швидкості 7,5мл/хвил. на колонку Pharmacia HR16/10, що була упакована сильним аніонообмінним матриксом PerSeptive Biosystem Porose® HQ, зрівноваженим в буфері, за допомогою системи BEPX PerSeptive Biosystem Sprint®. Після нанесення колонку промивали буфером до тих пір, доки не досягали $A_{280} 0,100$. Потім білки елюювали з колонки при швидкості 2,5мл/хвил. за допомогою буферу з 0,4М NaCl протягом 20 хвилин для загального об'єму 50мл. Потім колонку промивали буфером з 1,0М NaCl при тій самій швидкості потоку ще протягом 20 хвилин (кінцевий об'єм =50мл). Білки, які було елюювано 0,4М та 1,0М NaCl, приміщували в окремі діалізні мішочки та давали діалізу проходити впродовж ночі при 4°C в 12л буферу. Більша частина активності проти SCR містилася в фракції 0,4М. Фракцію 0,4М очищали далі нанесенням 20мл на колонку 25/100 Pharmacia, що була упакована Sepharose CL4B (Pharmacia), з застосуванням швидкості потоку 0,75мл/хвил. Фракції об'єднували на основі профілю піків A_{280} та концентрували до кінцевого об'єму 0,75мл за допомогою мембрани Biomax-50K NMWL центрифугального фільтруючого пристрою Millipore Ultrafree®. Концентрації білка визначали, застосовуючи набір Biorad Protein Assay з бичачим гама-глобуліном як стандартом.

Характеристика

Нативну молекулярну масу комплексу токсину проти SCR визначали за допомогою колонки Pharmacia 16/50, що її було упаковано попередньо Sepharose CL4B в буфері. Потім колонку калібрували за допомогою білків

відомого молекулярного розміру, що дозволяло розрахувати приблизний нативний молекулярний розмір. Як показано в таблиці 20, молекулярний розмір комплексу токсину знаходиться в межах від 777кДа для штаму Hb до 1900кДа для штаму WX-12. Вихід комплексу токсину також варіював, від 0,8мг/л для штаму WX-12 до 7,0мг/л для штаму Hb.

Білки, виявлені в комплексі токсину, випробовували на розмір індивідуальних поліпептидів за допомогою аналізу електрофорезом в ПААГ-ДСН. Як правило, 20мг білка комплексу токсину з кожного штаму наносили на 2 - 15% поліакриламідний гель (Integrated Separation Systems) та піддавали електрофорезу при 20мА в буфері для електрофорезу Biorad. По закінченні електрофорезу гелі забарвлювали протягом ночі в Biorad Coomassie blue R-250 (0,2% в суміші метанол:оцтова кислота:вода; 40:10:40 об./об./об.). Потім барвник видаляли з гелів в суміші метанол:оцтова кислота:вода; 40:10:40. Потім гелі промивали водою протягом 15 хвилин та сканували за допомогою персонального лазерного денситометру Molecular Dynamics. Доріжки оцінювали кількісно та розраховували молекулярні розміри в порівнянні з високомолекулярними стандартами Biorad, що були в діапазоні від 200 до 45кДа.

Розміри індивідуальних поліпептидів, що містять комплекс токсину проти SCR з кожного штаму, наведено в таблиці 21. Розміри окремих поліпептидів знаходяться в межах від 230кДа у штаму WX-1 до 16кДа, як видно у випадку штаму WX-7. Кожен штам, за винятком штаму Hb, мав поліпептиди в комплексі токсину, що були в діапазоні 160-230кДа, 100-160кДа та 50-80кДа. Ці дані вказують на те, що комплекс токсину може варіювати в пептидному складі та компонентах у різних штаммах, однак, в усіх випадках токсин, очевидно, складається з великого олігомерного білкового комплексу.

Таблиця 20		
Характеристика комплексу токсинів з штамів Photorhabdus, які не є W-14		
Штам	Прибл. нативна мол. маса ^a	Вихід Активна фракція ^b (мг/л)
H9	972.000	1,8
Hb	777.000	7,0
Hm	1.400.000	1,1
HP88	813.000	2,5
NCI	1.092.000	3,3
WIR	979.000	1,0
WX-1	973.000	0,8
WX-2	951.000	2,2
WX-7	1.000.000	1,5
WX-12	898.000	0,4
WX-14	1.900.000	1,9
W-14	860.000	7,5
^a Нативна молекулярна маса, яку визначено з застосуванням колонки 16/50 Pharmacia HR, яку набито Сефарозою CL4B		
^b Кількість комплексу токсинів, що їх здобувають з культурального бульйону		

Спектр активності

Як показано в таблиці 21, комплекси токсину, очищені з штамів Hm та H9, випробовували на активність проти численних комах, з комплексом токсину з штаму W-14 для порівняння. Тести проводили, як описано в Прикладі 13. Комплекс токсину з всіх трьох штамів виявив активність проти совки, точильника зернового та цикадки айстр. Крім того, комплекс токсину з штамів Hm та W-14 виявляв активність проти кліща двоплямистого павутинного. Крім того, комплекс токсину з W-14 виявив активність проти личинок комарів. Ці дані свідчать про те, що комплекс токсину, що мав схожість в активностях щодо деяких рядів комах, може також виявляти різні активності проти інших рядів комах.

Таблиця 21											
Приблизні розміри (в кДа) пептидів в очищеному комплексі токсинів з Photorhabdus, який не є W-14											
H9	Hb	Hm	HP88	NC-1	WIR	WX-1	WX-2	WX-7	WX-12	WX-14	W-14
180	150	170	170	180	170	230	200	200	180	210	190
170	140	140	160	170	160	190	170	180	160	180	180
160	139	100	140	140	120	170	150	110	140	160	170
140	130	81	130	110	110	160	120	87	139	120	160
120	120	72	129	44	89	110	110	75	130	110	150
98	100	68	110	16	79	98	82	43	110	100	130
87	98	49	100		74	76	64	33	92	95	120
84	88	46	86		62	58	37	28	87	80	110
79	81	30	81		51	53	30	26	80	69	93
72	75	22	77		40	41		23	73	49	90
68	69	20	73		39	35		22	59	41	77
60	60	19	60		37	31		21	56	33	69
57	57		58		33	28		19	51		65
52	54		45		30	24		18	37		63

46	49		39		28	22		16	33		60
40	44		35		27				32		51
37	39				25				26		46
	37				23						40
	35										39
											29

Таблиця 22	
Інсектицидний спектр очищеного комплексу токсину з штамів Photorhabdus, що його спостерігають	
Штам Photorhabdus	Види комах, що є чутливими
Комплекс токсину Hm	1**,2,3,5,6,7,8
Комплекс токсину H9	1,2,3,6,7,8
Комплекс токсину W-14	1,2,3,4,5,6,7,8
*=>25% смертність та/або інгібування росту	
**=1: совка; 2: точильник зерновий; 3: білшка довговуса південна; 4: комар; 5: кліщ двоплямистий павутинний; 6: цикадка айстр; 7: плодова мушка звичайна; 8: довгоносик бавовняний	

Приклад 15

Субфракціонування комплексу білкового токсину Photorhabdus

Комплекс білкового токсину Photorhabdus виділяли, як описано в Прикладі 14. Потім приблизно 10мг токсину наносили на колонку MonoQ 5,5, зрівноважену 20мМ Трис-НCl, рН7,0, при швидкості потоку 1мл/хвил. Колонку промивали 20мМ Трхс-НCl, рН7,0, до тих пір, поки оптична густина при 280нм не верталася до поглинання базової лінії. Білки, що зв'язувались з колонкою, елюювали лінійним градієнтом 0 - 1,0М NaCl в 20мМ Трис-НCl, рН7,0, при 1мл/хвил., протягом 30 хвилин. Збирали фракції по 1мл та піддавали їх біоаналізу з блішкою довговусою південною (SCR) (див. Приклад 13). Піки активності визначали за допомогою серій розведень кожної фракції в біотестах з SCR. Спостерігали два піки активності проти SCR, які було названо А (елюйований приблизно при 0,2 - 0,3М NaCl) та В (елюйований приблизно при 0,3 - 0,4М NaCl). Піки активності А та В зливали окремо та обидва піки очищали далі за допомогою 3-стадійної процедури, яку описано нижче.

Твердий (NH₄)₂SO₄ додавали до згаданої вище білкової фракції до кінцевої концентрації 1,7М. Потім білки наносили на колонку 5/5phenyl-Superoose, зрівноважену 1,7М (NH₄)₂SO₄ в 50мМ К-фосфатному буфері, рН7,0, при швидкості потоку 1мл/хвил. Білки, що зв'язалися з колонкою, елюювали лінійним градієнтом 1,7М (NH₄)₂SO₄, 0% етиленгліколь, 50мМ фосфат калію, рН7,0, - 25% етиленгліколь, 25мМ фосфат калію, рН7,0 (без сульфату амонію) при 0,5мл/хвил. Фракції діалізували протягом ночі проти 10мМ К-фосфатного буферу, рН7,0. Активності в кожній фракції проти SCR визначали за допомогою біотесту.

Фракції з найвищою активністю об'єднували та наносили на колонку 5/5 MonoQ, яку зрівноважували 20мМ Трис-НCl, рН7,0, при 1мл/хвил. Білки, що зв'язалися з колонкою, елюювали при 1мл/хвил лінійним градієнтом 0-1н NaCl в 20мМ Трис-НCl, рН7,0.

Для кінцевої стадії очищення найбільш активні фракції (які визначено за допомогою біотесту з SCR) об'єднували та піддавали очищенню на другій колонці phenyl-Superoose 5/5. Твердий (NH₄)₂SO₄ додавали аж до кінцевої концентрації 1,7М. Потім розчин наносили на колонку, зрівноважену 1,7М (NH₄)₂SO₄ в 50мМ К-фосфатному буфері, рН7,0, при 1мл/хвил. Білки, зв'язані з колонкою, елюювали лінійним градієнтом 1,7М (NH₄)₂SO₄, 5мМ фосфат калію, рН7,0, - 10мМ фосфат калію, рН7,0, при 0,5мл/хвил. Фракції діалізували протягом ночі проти 10мМ Na-фосфатного буферу, рН7,0. Активності в кожній фракції проти SCR визначали за допомогою біотесту.

Кінцевий білок, очищений за допомогою описаної вище 3-стадійної процедури, з піка А було названо токсином А, а кінцевий білок з піка В було названо токсином В. Характеристика та секвенування амінокислот токсину А та токсину В.

При електрофорезі в ПААГ-ДСН токсин А та токсин В містили два основних (>90% від всього забарвленого Кумасі білка) пептиду: 192кДа (що їх назвали А1 та В1, відповідно) та 58кДа (що їх назвали А2 та В2, відповідно). Токсин А та токсин В виявили тільки одну основну смугу в нативному електрофорезі, що свідчить про те, що А1 та А2 були субодинацями одного білкового комплексу, та В1 і В2 були субодинацями одного білкового комплексу. Далі, було визначено, що нативна молекулярна маса обох токсинів (А та В) дорівнює 860кДа відповідно до гель-фільтраційної хроматографії. Відносні молярні концентрації А1 до А2 було оцінено як 1екв. до 1екв., як визначене денситометричним аналізом гелів ПААГ-ДСН. Подібно, пептиди В1 та В2 наявні при одній і тій самій молярній концентрації.

Токсин А та токсин В піддавали електрофорезу в 10% ПААГ-ДСН та трансблотингу на мембрани PVDF. Блоти посилали на амінокислотний аналіз та N-кінцеве амінокислотне секвенування до Harvard Microchem та Cambridge Prochem, відповідно. Було визначено, що N-кінцева послідовність В1 є ідентичною SEQ ID №1, району TcbA_{ii} гена tcbA (SEQ ID №12, позиція 87-99). Унікальну N-кінцеву послідовність було отримано для пептиду В2 (SEQ ID №40). N-кінцева амінокислотна послідовність пептиду В2 була ідентична району TcbA_{iii} отриманої амінокислотної послідовності для гена tcbA (SEQ ID №12, позиція 1935-1945). Отже, токсин В містив здебільшого два пептиди, TcbA_{ii} та TcbA_{iii}, що були вироблені з одного ген продукту, TcbA.

N-кінцева послідовність А2 (SEQ ID №41) була унікальною порівняно з пептидом TcbA_{iii} та іншими

пептидами. Пептид A2 було названо TcdA_{iii} (див. Приклад 17). Було визначено, що SEQ ID №6 є сумішшю амінокислотних послідовностей SEQ ID №40 та 41.

Пептиди A1 та A2 піддавали далі секвенуванню внутрішньою амінокислотою послідовністю. 10мкг токсину A піддавали електрофорезу в 10% ПААГ-ДСН-електрофорезі та трансблотингу на мембрану PVDF. Після забарвлення блята амідо чорним пептиди A1 та A2, названі TcdA_{ii} та TcdA_{iii}, відповідно, вирізали з блята та посилали до Harvard Microchem та Cambridge ProChem. Пептиди піддавали розщепленню трипсином з подальшою ВЕРХ для розділення індивідуальних пептидів. Аналіз N-кінцевої амінокислотної послідовності виконували на вибраних пептидних фрагментах, отриманих при розщепленні трипсином. Було виявлено, що дві внутрішні амінокислотні послідовності пептиду A1 (TcdA_{ii}-PK71, SEQ ID №38 та TcdA_{ii}-PK44, SEQ ID №39) мають значні гомології з розшифрованими амінокислотними послідовностями району TcdA_{ii} гена tcdA (SEQ ID №12). Таким чином, N-кінцева послідовність (SEQ ID №41) та дві внутрішні послідовності пептидів A2 (TcdA_{iii}-PK57, SEQ ID №42 та TcdA_{iii}-PK20, SEQ ID №43) також виявили значну гомологію з розшифрованими амінокислотними послідовностями району TcdA_{iii} гена tcdA SEQ ID №12).

Резюмуючи наведені вище результати, можна сказати, що комплекс токсину має принаймні два активних проти SCR комплекси: токсин A та токсин B. Токсин A та токсин B є схожими за нативною молекулярною масою та молекулярною масою субодиниць, однак, їх пептидні склади відрізняються. Токсин A містить пептиди TcdA_{ii} та TcdA_{iii} як основні пептиди, а токсин B містить TcdA_{ii} та TcdA_{iii} як основні пептиди.

Приклад 16

Розщеплення та активація пептиду TcdA

В комплексі токсину B пептиди TcdA_{ii} та TcdA_{iii} походять від одного генного продукту TcdA (Приклад 15). Процесінг пептиду TcdA до TcdA_{ii} та TcdA_{iii}, як припускається, здійснюється протеазою (протеазами) Photorhabdus та, найбільш імовірно, металопротеазами, які описано в Прикладі 10. У деяких випадках відзначалось, що при обробці бульйону W-14 Photorhabdus пептид TcdA є наявним в комплексі токсину B як основний компонент, на доповнення до пептидів TcdA_{ii} та TcdA_{iii}. Ідентичні процедури, які описано для очищення комплексу токсину B (Приклад 15) використовували для збагачення пептиду TcdA з фракції комплексу токсину бульйону W-14. Кінцевий очищений матеріал аналізували в 4-20% градієнтному електрофорезі в ПААГ-ДСН та основні пептиди визначали кількісно денситометрією. Було визначено, що TcdA, TcdA_{ii} та TcdA_{iii} містили 58%, 36% та 6%, відповідно, загального білка. Ідентичності цих пептидів було підтверджено їх відповідними молекулярними розмірами в електрофорезі в ПААГ-ДСН та Вестерн-блотингом з застосуванням моноспецифічних антитіл. Було визначено, що нативна молекулярна маса цієї фракції дорівнює 860кДа.

Розщеплення TcdA оцінювали шляхом обробки очищеного, як описано вище, матеріалу металопротеазами 38кДа та 58кДа W-14 Photorhabdus (Приклад 10) та трипсином як контрольним ферментом (Sigma, MO). Стандартна реакція містила 17,5мкг очищеної фракції, 1,5 одиниць протеази та 0,1М Трис-буфер, pH8,0, в загальному об'ємі 100мкл. До контрольної реакції протеази не додавали. Реакційні суміші інкубували при 37°C протягом 90 хвилин. В кінці реакції брали 20мкл та кип'ятили з буфером для проб для електрофорезу в ПААГ-ДСН безпосередньо перед електрофорезом в 4 - 20% градієнтному гелі. Було визначено на основі електрофорезу в ПААГ з ДСН, що в обох варіантах (з протеазою 38кДа та протеазою 58кДа) кількість пептидів TcdA_{ii} та TcdA_{iii} збільшувалася приблизно в 3 рази, тоді як кількість пептиду TcdA зменшувалася пропорційно (Таблиця 23). Відносне зменшення та збільшення вибраних пептидів було підтверджено Вестерн-блотингом. Крім того, гель-фільтрування розщепленого матеріалу виявило, що нативний молекулярний розмір комплексу залишився тим самим. При обробці трипсином пептиди TcdA та TcdA_{ii} неспецифічно розщеплялись на невеликі пептиди. Це вказувало на те, що протеази 38кДа та 58кДа Photorhabdus можуть специфічно процесувати пептид TcdA в пептиди TcdA_{ii} та TcdA_{iii}. Оброблений протеазою та необроблений контроль решти - 80мкл - реакційної суміші серійно розводили 10мМ буфером з фосфату натрію, pH7,0, та аналізували за допомогою біотесту з SCR. Шляхом порівняння активності в кількох розведеннях визначили, що обробка протеазою 38кДа збільшувала інсектицидну активність у відношенні до SCR приблизно в 3 - 4 рази. Інгібування росту комах, що залишалися, при обробці протеазою було більш значним, ніж у контролі (Таблиця 23).

Таблиця 23		
Перетворення та активація пептиду TcdA в пептиди TcdA _{ii} та TcdA _{iii} внаслідок обробки протеазою		
	Контроль	Обробка протеазою 38кДа
SO (% загального білка)	58	18
S1 (% загального білка)	36	64
S9 (% загального білка)	6	18
LD50 (мкг білка)	2,1	0,52
маса SCR (мг/комаха)*	0,2	0,1

*: вказує на інгібування росту шляхом вимірювання середньої маси живої комахи після 5 днів на кормі в тесті.

Приклад 17

Скринінг бібліотеки на ген, що кодує пептид TcdA_{ii}

Було проведено клонування та характеристика гена, що кодує пептид TcdA_{ii}, який описано в SEQ ID №17 (N-кінцева послідовність внутрішнього пептиду TcdA_{ii}-PTIII), та SEQ ID №18 (N-кінцева послідовність внутрішнього пептиду TcdA_{ii}-PT79). Було синтезовано два пула вироджених олігонуклеотидів, призначених для

кодування амінокислотних послідовностей SEQ ID №17 (таблиця 24) та SEQ ID №18 (Таблиця 25) й зворотних комплементів цих послідовностей, як описано в Прикладі 8. Послідовність ДНК цих олігонуклеотидів наведено нижче:

Таблиця 24

Вироджений олігонуклеотид для SEQ ID № 17

P2-PT111	1	2	3	4	5	6	7	8
Амінокислота	Ala	Phe	Asn	Ile	Asp	Asp	Val	Ser
Кодони	5' GCN	TT(T/C)	AA(T/C)	AT(T/C/A)	GA(T/C)	GA(T/C)	GTN 3'	
P2.3.6.CB	5' GC(A/C/G/T)	TT(T/C)	AAT	ATT	GAT	GAT	GT 3'	
P2.3.5	5' GC(A/C/G/T)	TT(T/C)	AA(T/C)	AT(T/C/A)	GA(T/C)	GA(T/C)	GT 3'	
P2.3.5R	5' AC	(G/A)TC	(G/A)TC	(T/G/A)AT	(G/A)TT	(G/A)AA	(A/C/G/T)GC 3'	
P2.3.5RI	5' ACI	TCI	TCI	ATI	TTI	AAI	GC 3'	
P2.3R.CB	5' CAG	(A/G)CT	(A/C)AC	ATC	ATC	AAT	ATT	AAA 3'

Таблиця 24

Вироджений олігонуклеотид для SEQ ID № 18

P2-PT79	1	2	3	4	5	6	7	8	9	10	11	12	13
Амінокислота	Phe	Ile	Val	Tyr	thr	Ser	Leu	Gly	Val	Asn	Pro	Asn	Asn
Кодони*	5' TTY	ATH	GTN	TAY	ACN	6	6	GGN	GTN	AAY	CCN	AAY	AAY 3'
P2.79.2	5' TTY	ATY	GTK	TAT	ACY	TCI	YTR	GGY	GTK	AAT	CCR	AAT	AAT 3'
P2.79.3	5' TTT	ATT	GTK	TAT	ACY	AGY	YTR	GGY	GTK	AAT	CCR	AAT	AAT 3'
P2.79.R.1	5' ATT	ATT	YGG	ATT	MAC	RCC	YAR	RCT	RGT	ATA	MAC	AAT	AAA 3'
P2.79R.CB	5' ATT	ATT	YGG	ATT	MAC	ACC	CAG	RCT	GGT	ATA	MAC	AAT	AAA 3'

* Відповідно до IUPAC-IUB кодує нуклеотиди: Y = C або T, H = A, C або T, N = A, C або T, K = G або T, R = A або C та M = A або C

Полімеразні Ланцюгові Реакції (ПЛР) виконували по суті, як описано в Прикладі 8, з використанням як прямих праймерів P2.3.6.CB або P2.3.5 та як зворотних праймерів P2.79.R.1 або P2.79R.CB в усіх прямих/зворотних комбінаціях, з використанням геномної ДНК Photorhabdus як матриці. В іншій серії реакцій праймери P2.79.2 або P2.79.3 використовували як прямі праймери та P2.3.5R, P2.3.5R.1 та P2.3R.CB використовували як зворотні праймери в усіх прямих /зворотних комбінаціях. Тільки в реакціях, що містили P2.79.R.1 або P2.79R.CB як прямі праймери в комбінації з P2.79.R.1 або P2.79R.CB як зворотні праймери було виявлено не-артефактний продукт ампліфікації, з визначеним розміром (за рухливістю на агарозних гелях) 2500п.н. Порядок праймерів, які використовували для отримання цього продукту ампліфікації, вказує на те, що пептидний фрагмент TcdA_{II}-PTIII лежить амінопроксимально по відношенню до пептидного фрагменту TcdA_{II}-PT79.

Продукти ПЛР 2500п.н. лігували з плазмидним вектором pCRTMII (Invitrogen, San Diego, CA) відповідно до

інструкцій виробника, та послідовності ДНК, що перетинають кінці інсерційних фрагментів двох ізолятів (HS24 та HS27) визначали за допомогою рекомендованих постачальником праймерів та способів секвенування, які було описано вище. Послідовність обох ізолятів була однаковою. Нові праймери синтезували на основі визначеної послідовності та використовували для праймування додаткових реакцій секвенування для отримання в цілому 2557 основ інсерту [SEQ ID №36]. Трансляція часткового пептиду, який кодує SEQ ID №36, дає послідовність з 845 амінокислот, описану в SEQ ID №37. Аналіз білкової гомології цієї частини пептидного фрагменту TcdA_{ii} виявив значну амінокислотну гомологію (68% схожість; 53% ідентичність) з залишками 542-1390 білка TcbA [SEQ ID №12]. Таким чином, очевидно, що ген, який представлено в частині SEQ ID №36, продукує білок схожої, але не ідентичної амінокислотної послідовності з білком TcbA, причому цей білок, як видно, має подібну, але не ідентичну біологічну активність порівняно з білком TcbA.

Ще в одному випадку, було виділено ген, що кодує пептиди TcdA_{ii}-PK44 та N-кінцевий пептид 58кДа TcdA_{iii}, які описано як SEQ ID №9 (TcdA_{ii}-PK44 послідовність внутрішнього пептиду), та SEQ ID №41 (N-кінцева пептидна послідовність 58кДа TcdA_{iii}). Було синтезовано два пули вироджених олігонуклеотидів, призначених для кодування амінокислотних послідовностей, описаних як SEQ ID №39 (таблиця 27) та SEQ ID №41 (таблиця 26), та зворотних комшієментів цих послідовностей, як описано в Прикладі 8, та їх послідовності ДНК.

Таблиця 26

Вироджений олігонуклеотид для SEQ ID № 41

№ кодову	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Амінокис- лота	Leu	Arg	Ser	Ala	Asn	Thr	Leu	Thr	Asp	Leu	Phe	Leu	Pro	Gln
A2.1	5' YTR	CGY	AGY	GCI	AAT	ACY	YTR	ACY	GAT	YTR	TTT	YTR	CCR	CA 3'
A2.2				GCI	AAT	ACI	YTR	ACI	GAY	YTR	TTY	YTR	CCI	CA 3'
A2.3.R		5' TG	YGG	YAR	AAA	YAR	RTC	RGT	YAR	RGT	RTT	IGC	RCT	RCG 3'
A2.4.R				5' TG	IGG	CAG	AAA	CAG	RTC	IGT	CAG	IGT	ATT	IGC 3'

Вироджений олігонуклеотид для SEQ ID № 39

№ амі- нокис- лоти	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
№ кодо- ву	1	2	3	4	5	6	7	8	9
Аміно- кислота	Gly	Pro	Val	Glu	Ile	Asn	Thr	Ala	Ile
A1.44.1	5' GGY	CCR	GTK	GAA	ATT	AAT	ACC	GCI	AT 3'
A1.44.1R	5' ATI	GCG	GTA	TTA	ATT	TCM	ACY	GGR	CC 3'
A1.44.2	5' GGI	CCI	GTI	GAR	ATY	AAV	ACI	GCI	AT 3'
A1.44.2R	5' ATI	GCI	GTR	TTR	ATY	TCI	ACI	GGI	CC 3'

Полімеразні Ланцюгові Реакції (ПЛР) проводили по суті, як описано в Прикладі 8, з застосуванням прямих праймерів A1.44.1 або A1.44.2 та зворотних праймерів A2.3R або A2.4R в усіх прямих/зворотних комбінаціях, з використанням геномної ДНК *Photobacterium W-14* як матриці. В іншій серії реакцій праймери A2.1 або A2.2 використовували як прямі праймери та A1.44.1R та A1.44.2R використовували як зворотні праймери в усіх прямих/зворотних комбінаціях. Не-артефактний ампліфікований продукт з розміром 1400п.н. (визначеним за рухливістю на агарозних гелях) було виявлено тільки в реакціях, що містили A 1.44.1 або A 1.44.2 як прямі праймери в поєднанні з A2.3R як зворотним праймером. Порядок праймерів, що використовувалися для отримання цього ампліфікованого продукту вказує на те, що пептидний фрагмент TcdA_{II}-PK44 лежить амінопроксимально по відношенню до пептидного фрагменту 58кДа TcdA_{III}.

Продукти ПЛР 1400п.н. лігували з плазмидним вектором pCRTMII відповідно до інструкцій постачальника. Послідовності ДНК, що перетинають кінці інсерційних фрагментів чотирьох ізолятів визначали з використанням праймерів, схожих за послідовністю з рекомендованими постачальником праймерами, та з використанням способів секвенування, які було описано раніше. Послідовність нуклеїнової кислоти всіх ізолятів відрізнялася, як і очікувалося, в районах, що відповідають виродженим послідовностям праймерів, але амінокислотні послідовності, які було розшифровано з цих даних, були такими ж, що й істинні амінокислотні послідовності для цих пептидів, визначені раніше, (SEQ ID №41 та 39).

Скринінг геномної космидної бібліотеки W-14, як описано в Прикладі 8, радіоактивно міченим зондом, що складався з ДНК, отриманої вище (SEQ ID №36), ідентифікував п'ять космидних ізолятів, які гібридизувалися, а саме, 17D9, 20B10, 21D2, 27B10 та 26d1. Ці космиди відрізнялися від космид, що ідентифікувалися раніше з зондами, які відповідали генам, описаним як SEQ ID №11 або SEQ ID №25. Рестрикційний аналіз та блот-гібридизації ДНК ідентифікували три фрагменти EcoR I з приблизним розміром 3,7, 3,7 та 1,1т.п.н., що охоплюють район, який включає ДНК SEQ ID №36. Скринінг геномної космидної бібліотеки W-14 з застосуванням як зонду радіоактивно міченого фрагменту ДНК 1,4т.п.н., отриманого в цьому прикладі, ідентифікував ті ж самі п'ять космид (17D9, 20B10, 21D2, 27B10 та 26D1). Блот-гібридизація ДНК з розщепленими EcoR I космидними ДНК також показала гібридизацію з тією ж самою підсерією фрагментів EcoR I, яку спостерігали з зондом гена TcdA_{II} 2,5т.п.н., що свідчить про те, що обидва фрагменти кодуються на геномній ДНК.

Визначення послідовності ДНК клонуваних фрагментів EcoR I виявило неперервну рамку зчитування з 7551п.н. (SEQ ID №46), що кодує білок 282,9кДа з 2516 амінокислот (SEQ ID №47). Аналіз амінокислотної послідовності цього білка виявив всі очікувані внутрішні фрагменти пептидів TcdA_{II} (SEQ ID №17, 18, 37, 38 та 39) та N-кінця пептиду TcdA_{III} (SEQ ID №41) й усі внутрішні пептиди TcdA_{III} (SEQ ID №42 та SEQ ID №43). Кожен з пептидів, виділених та ідентифікованих як TcdA_{II} та TcdA_{III}, є продуктом відкритої рамки зчитування, яку позначено TcdA, описано в SEQ ID №46. Далі, SEQ ID №47 показує, починаючи з позиції 89, послідовність, яку описано у вигляді SEQ ID №13, що є N-кінцевою послідовністю пептиду з розміром приблизно 201кДа, а це вказує на те, що первинний білок, який продукується з SEQ ID №46, процесується подібно до того, як описано

раніше для SEQ ID №12. Крім того, цей білок потім розщеплюється з утворенням продукту з розміром 209,2кДа, який кодує SEQ ID №48 та який описано як SEQ ID №49 (пептид TcdA_{ii}), та продукту з розміром 63,6кДа, який кодує SEQ ID №50 та який описано як SEQ ID №51 (пептид TcdA_{iii}). Таким чином, можна вважати, що інсектицидна активність, що ідентифікувалася як токсин А (Приклад 15), вироблений з продуктів SEQ ID №46, таких як повнорозмірний білок 282,9кДа, який описано в SEQ ID №47, процесується з утворенням пептидів, описаних у вигляді SEQ ID №49 та SEQ ID №51. Вважають, що інсектицидна активність, яку ідентифіковано як токсин В (Приклад 15), виробляється з продуктів SEQ ID №11, таких як білок 280,6кДа, що його описано у вигляді SEQ ID №12. Цей білок протеолітично процесується з утворенням пептиду 207,6кДа, який описано в SEQ ID №53, що його кодує SEQ ID №52, та пептиду 62,9кДа, що має N-кінцеву послідовність, описану у вигляді SEQ ID №40 та далі описану у вигляді SEQ ID №55, що кодується SEQ ID №54.

Порівняння амінокислотних послідовностей між білками, описаними як SEQ ID №12 та SEQ ID №47, виявили, що вони мають 69% схожість та 54% ідентичність. Цей високий ступінь еволюційної спорідненості не є однорідним по всій амінокислотній послідовності цих пептидів, але він є високим у напрямку карбокси-кінця цих білків, оскільки пептиди, описані як SEQ ID №51 (вироблені з SEQ ID №47) та як SEQ ID №55 (вироблені з SEQ ID №12) мають 76% схожість та 64% ідентичність.

Приклад 18

Захист від пошкодження листя рослин кукурудзи, який індукує точильник зерновий, шляхом нанесення розпиленням бульйону

Photorhabdus (штаму W-14)

Здатність токсину (токсинів) *Photorhabdus* зменшувати пошкодження листя, що викликається личинками комах, було показано шляхом вимірювання пошкодження листя, що викликається точильником зерновим (*Ostrinia nubilalis*), оселеним на рослини кукурудзи, оброблені бульйоном *Photorhabdus*. Ферментаційний бульйон від штаму W-14 *Photorhabdus* одержували та концентрували приблизно в 10 разів за допомогою ультрафільтрування (розмір пор 10000 (мол. маса)), як описано в Прикладі 13. Отриманий концентрований бульйон стерилізували, пропускаючи крізь нітроцелюлозні мембранні фільтри 0,2 мікрон. Приготовану у такий спосіб пробу неінокульованого 2% протеозе-пептона №3 використовували для контролю. Рослини кукурудзи (інбредна лінія, маєток DowElanco) вирощували з насіння до вегетативної стадії 7 або 8 у горщиках, що містили суміш (без ґрунту), в оранжереї (27°C удень, 22°C вночі, приблизно 50% відносна вологість, тривалість світлового періоду 14 годин, полив та добрива - по необхідності). Тест-рослини розміщували у вигляді рандомізованих груп, що повторюються (3 повторності/варіант (обробку), 6 рослин/варіант)) в оранжереї з температурою приблизно 22°C удень, 18°C вночі, без штучного світла та з частковим затіненням, з відносною вологістю близько 50% та з поливами та внесенням добрив по необхідності. Обробляючи розчини (неінокульовані середовища та концентрований бульйон *Photorhabdus*) наносили обприскувачем з шприцом, по 2,0мл наносили з відстані приблизно 6 дюймів над мутовкою листя та додаткові 2,0мл наносили рухом по колу з відстані приблизно одного фута над мутовкою. Крім того, одну групу рослин не було оброблено. Після підсихання нанесених розчинів (приблизно через 30 хвилин) дванадцять новонароджених личинок точильника зернового (яйця одержували з комерційних джерел та вилуплення проводили в лабораторії) наносили безпосередньо на мутовку. Після одного тижня рослини оцінювали на пошкодження листя, застосовуючи модифіковану шкалу Гатрі [Koziel, M.G., Beland, G.L., Bowman, C., Carozzi, N.B., Crenshaw, R., Crossland, L., Dawson, J., Desai, M., Hill, M., Kadwell, S., Launis, K., Lewis, K., Maddox, D., McPherson, K., Meghji, M.Z., Merlin, E., Rhodes, R., Warren, G.W., Wright, M. and Evola, S.V. 1993] *Bio/Technology*, II, 194-195] та оцінка порівнювали статистично [Т-критерій (найменша значуща (вірогідна) різниця, LSD) $p < 0,05$ та Tukey's Studentized Range (HSD) Test $p < 0,11$]. Ці результати наведено в таблиці 28. Для посилення, оцінка 1 означає відсутність пошкодження, оцінка 2 означає дрібне пошкодження типу "віконного скла" на листі, що не згорнувся, без шпилькових проникнень та оцінка 5 означає проникнення в лист з подовженими пошкодженнями та/або поїданням середньої жилки, що спостерігаються на більш ніж 3 листах (пошкодження < 1 дюйму). Ці дані показують, що бульйон або інші фракції, що містять білок, можуть створювати захист проти специфічних комах-шкідників при нанесенні у вигляді розприскуваного препарату або при постачанні гена або його похідного, що кодує цей білок або його частину, через трансгенну рослину або мікроб.

Таблиця 28

Валив культурального бульйону <i>Photorhabdus</i> на пошкодження листя точильником зерновим на кукурудзі	
Обробка	Середній бал за Гатрі
Без обробки	5,02 ^a
Неінокульоване середовище	5,15 ^a
Бульйон <i>Photorhabdus</i>	2,24 ^b
Середні з різними літерами є статистично різними ($p < 0,05$ або $p < 0,1$).	

Приклад 19

Генетичне конструювання генів для експресії в *E. Coli*

Резюме конструювання

Ряд плазмид було сконструйовано для експресії гена *tcbA* *Photorhabdus* W-14 в *Escherichia coli*. Перелік цих плазмид наведено в Таблиці 29. Далі йде стислий опис кожної конструкції, а також резюме отриманих для

експресії E. coli даних.

5

10

Таблиця 29			
Експресійні плазмиди для гена TcbA			
Плазида	Ген	Вектор/відбір	Компартмент
pDAB634	tcbA	pBC/Chl	Всередині клітини
pAcGP67B/TcbA	tcbA	pAcGP67B/Amp	Vasculovirus, секретований
pDAB635	tcbA	pET27b/Kan	Периплазма
pET15-tcbA	tcbA	pET15-tcbA	Всередині клітини
Абревіатури: Kan=канаміцин, Chl=хлорамфенікол, Amp=ампіцилін			

Конструювання pPAB634

В Прикладі 9 описано великий фрагмент Ecor I, що гібридизується з зондом TcbA_{ij}. Цей фрагмент субклонували в pBC (Stratagene, La Jolla CA). Аналіз секвенування показує, що цей фрагмент має 8816п.н. Фрагмент кодує ген tcbA з ініціюючим кодоном ATG у позиції 571 та термінуючим кодоном TAA у позиції 8086. Отже, цей фрагмент несе 570п.н. ДНК Photorhabdus проти ходу транскрипції від ATG та 730п.н. по ходу транскрипції від TAA.

Конструювання плазмиди pAcGP67B/tcbA

Ген tcbA було ампліфіковано за допомогою ПЛР з використанням таких праймерів: 5'-праймер (SIac51) 5' TTT AAA CCA TGG GAA ACT CAT TAT CAA GCA CTA TC 3' та 3'-праймер (SIac31) 5' TTT AAA GCG GCC GCT TAA CGG ATG GTA TAA CGA ATA TG 3'. ПЛР проводили, застосовуючи набір TaKaRa LA PCR з PanVera (Madison, Wisconsin) в такій реакції: 57,5мл води, 10мл 10X LA-буферу, 16мл dNTP (2,5мМ кожного вихідного розчину), 20мл кожного з праймерів при 10пмоль/мл, 300нг плазмиди pDAB634, що містила ген tcbA W-14 та один мл полімерази Taq TaKaRa La. Умови проведення циклів були такими: 98 °/20сек., 68°C/5хвил., 72°C/10хвил. впродовж 30 циклів. Очікуваний продукт ПЛР близько 7526п.н. виділяли в 0,8% агарозному гелі в ТБЕ-буфері (100мМ Трис, 90мМ борна кислота, 1мМ ЕДТК) та очищали за допомогою набору Qiaex II з Qiagen (Chatsworth, California). Очищений ген tcbA розщепляли Nco I та Not I та лігували в бакуловірусний вектор-переносник pAcGP67B (PharMingen (San Diego, California) та трансформували у клітини DH5a E. coli. Потім ген tcbA вирізали з pAcGP67B та переносили в pET27b для створення плазмиди pDAB 635. Місценс-мутацію в гені tcbA було репаровано в pDAB635.

Репарований ген tcbA містить дві зміни порівняно з послідовністю, яку показано в SEQ ID №11: А замість G у позиції 212, що замінює аланін 71 на серин 71, та G замість А у позиції 229, що замінює аланін 77 на треонін 77. Обидві ці модифікації знаходяться проти ходу транскрипції від прогнозованого N-кінця TcbA_{ij}.

Конструювання pET15-tcbA

Кодуючий район tcbA pDAB635 переносили у вектор pET15b. Це здійснювали за допомогою шотган-лігувань, ДНК вирізали за допомогою рестриктаз Nco I та Xho I. Отриманий рекомбінант було названо pET15-tcbA.

Експресія TcbA в E. coli з плазмиди pET15-tcbA

Експресію tcbA в E. coli одержували шляхом модифікації способів, які було описано раніше Studier et al. [Studier, F.W., Rosenberg, A., Dunn, J. and Dubendorff, J., (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol., 185: 60-89]. Компетентні клітини E. coli штаму BL21 (DE3) трансформували плазмидою pET15-tcbA та висіювали на LB-агар, що містив 100мг/мл ампіциліну та 40мМ глюкози. Трансформовані клітини вирощували до густини кілька сотень ізольованих колоній на чашку. Після інкубування протягом ночі при 37 °C клітини вискрібали з чашок та суспендували в LB-бульйоні, що містив 100мг/мл ампіциліну. Типові об'єми культур були 200-500мл. У час 0 густини культур (OD₆₀₀) сягали 0,05 - 0,15 залежно від експерименту. Культури струшували при одній з трьох температур (22°C, 30°C або 37°C) до тих пір, доки не було отримано густину 0,15 - 0,5, після чого їх індукували 1мМ ізопропілтіо-бета-галактозидом (ІПТГ). Культури інкубували при обраній температурі протягом 4 - 5 годин та потім переносили на 4°C до обробки (12 - 72г).

Очищення та характеристика TcbA. експресованого в E. coli з плазмиди pET15-tcbA

Культури E. coli, що експресують пептиди TcbA, обробляли так. Клітини збирали центрифугуванням при 17000×g та середовище декантували й зберігали в окремому контейнері.

Середовище концентрували приблизно в 8 разів за допомогою фільтраційної системи M12 (Amicon, Beverly MA) та фільтру з відсіканням молекулярної маси 100кДа. Концентроване середовище наносили на аніонообмінну колонку та білки, що зв'язалися, елюювали 1,0М NaCl. Було виявлено, що пік елюції 1,0М NaCl спричинював смертність при застосуванні його проти личинок блішки довговусої, південної (SCR) (Таблиця 30). Фракцію 1,0М NaCl діалізували проти 10мМ буферу з фосфату натрію, pH7,0, концентрували та піддавали гел-фільтруванню на Sepharose CL-4B (Pharmacia, Piscataway, New Jersey). Зону профілю елюції CL-4B, що відповідає розрахованій молекулярній масі (приблизно 900кДа) нативного комплексу токсину W-14, збирали, концентрували та піддавали біотесту з личинками. Було виявлено, що зібрана фракція 900кДа мала інсектицидну активність (див. Таблицю 30 нижче), з симптоматологією, подібною до симптоматології, що спричинюється нативним комплексом токсину W-14. Цю фракцію піддавали дії Протеїнази К та термальній обробці. Активність в обох випадках зникала або зменшувалася, що доводить білкову природу носія активності. Крім того, ця активна фракція була імунологічно позитивною у відношенні до пептидів TcbA та TcbA_{ij} в імуноблотингу при тестуванні з моноклональними антитілами проти TcbA_{ij} (Таблиця 30).

Таблиця 30				
Результати імуноблотів та біотестів з SCR				
Фракція	Активність проти SCR		Імуноблот	Нативний розмір
	% смертності	% інгібування росту	Виявлені пептиди	(визначений на CL-4B)
TcbA в середовищі 1,0М іонообмін	+++	+++	TcbA	
TcbA в середовищі CL-4B +Протеїназа К	+++	+++	TcbA, TcbA _{iii}	~900кДа
TcbA в середовищі CL-4B +нагрівання	-	-	НТ	
TcbA суперу клітин CL-4B	-	+++	НТ	~900кДа
ПК= обробка Протеїназою К 2 години; Нагрівання= 100°C протягом 10 хвилини; НТ= не тестували. Система оцінки для смертності та інгібування росту в порівнянні з контрольними пробами: 5 - 24%="+", 25 - 49%="++", 5 - 100%="+++".				

Осад клітин повторно суспендували в 10мМ фосфатному буфері (з фосфатом натрію), рН7,0, та лізували, пропускаючи через клітинний розпилювач Bio-Neb™ (Glas-Col Inc., Terra Haute, IN). Осад обробляли ДНКазою для видалення ДНК та центрифугували при 17000 ×g для відділення клітинного осаду від супернатанту. Фракцію супернатанту декантували та фільтрували через фільтр 0,2 мікрон, щоб вилучити великі частки, та піддавали аніонообмінній хроматографії. Білки, що зв'язались, елюювали 1,0М NaCl, діалізували та концентрували за допомогою концентраторів Біотех™ (Millipore Corp, Bedford, MA) з відсіканням молекулярної маси 50000 Дальтон. Концентровану фракцію піддавали гель-фільтраційній хроматографії з застосуванням гранульованого матриксу CL-48. Дані біоаналізу для приготованого у такий спосіб матеріалу наведено в таблиці 30 та названо "TcbA суперу клітин".

В ще одному способі для обробки великої кількості матеріалу осаду клітин повторно суспендували в 10мМ натрій-фосфатному буфері, рН7,0, та обережно гомогенізували за допомогою млину для тканини (Kontes Glass Company (Vineland, NJ)) на 40мл. Залишки клітин осаджували центрифугуванням при 25000 ×g та клітинний супернатант декантували, пропускали крізь фільтр 0,2 мікрон та піддавали аніонообмінній хроматографії з застосуванням колонки 10/10 Pharmacia, упакованої гранулами Poros HQ 50. Білки, що зв'язались, елюювали градієнтом NaCl 0,0 - 1,0М. Фракції, що містили білок TcbA, об'єднували та концентрували за допомогою 50кДа-концентратору та піддавали гель-фільтраційній хроматографії з застосуванням гранульованого матриксу Pharmacia CL-4B. Фракції, що містили олігомер TcbA з молекулярною масою приблизно 900кДа, збирали та піддавали аніонообмінній хроматографії з застосуванням колонки Pharmacia Mono Q 10/10, зрівноваженої 20мМ Трис-буфером, рН7,3. Для елюції рекомбінантного білка TcbA використовували градієнт 0,0 - 1,0М NaCl. Рекомбінантний TcbA елюювався з колонки при концентрації солі приблизно 0,3 - 0,4М, при тій самій молярності, при якій нативний олігомер TcbA елюювався з колонки Mono Q 10/10. Було виявлено, що ця фракція рекомбінантного TcbA викликала смертність SCR в експериментах з біоаналізом, подібних до експериментів, що їх наведено в таблиці 30.

Формула винаходу

1. Полінуклеотид, функціонально зв'язаний з гетерологічним промотором, що кодує білок, який має активність токсину проти комах-шкідника, де нуклеотидна молекула, що кодує вказаний білок, зберігає здатність до гібридизації з комплементарною послідовністю нуклеїнової кислоти після гібридизації і промивання, причому вказану гібридизацію здійснюють при 60 °C в розчині, що містить 10 % (м/о) ПЕГ (поліетиленгліколю, М.м. приблизно 8000 Да), 0,6 x SSC, 7 % (м/о) SDS, 10 мМ натрійфосфатний буфер, 5 мМ EDTA і 100 мг/мл денатурованої ДНК із молочка лосося, де вказана послідовність нуклеїнової кислоти являє собою SEQ ID NO: 46.
2. Полінуклеотид за п. 1, який відрізняється тим, що промивання здійснюють в 0,25 x SSC і 0,2 % SDS при 60°C.
3. Полінуклеотид, що кодує білок, який має активність токсину проти комах-шкідника, де вказаний білок містить амінокислотну послідовність SEQ ID NO: 47, де вказаний полінуклеотид функціонально зв'язаний з гетерологічним промотором.
4. Полінуклеотид, що містить нуклеотидну послідовність, яка кодує білок, що має амінокислотну послідовність SEQ ID NO: 47, функціонально зв'язаний з гетерологічним промотором, причому вказаний білок має активність токсину проти комах-шкідника.
5. Трансгенна рослинна клітина, яка експресує полінуклеотид за п. 1.
6. Рекомбінантний білок, який має активність токсину проти комах-шкідника, де нуклеотидна послідовність, що кодує вказаний білок, зберігає здатність до гібридизації з послідовністю нуклеїнової кислоти при гібридизації і промиванні, причому вказану гібридизацію здійснюють при 60 °C в розчині, що містить 10 % (м/о) ПЕГ (поліетиленгліколю, М.м. приблизно 8000 Да), 0,6 x SSC, 7 % (м/о) SDS, 10 мМ натрійфосфатний буфер, 5 мМ EDTA і 100 мг/мл денатурованої ДНК із молочка лосося, де вказана послідовність нуклеїнової кислоти являє собою SEQ ID NO: 46.

7. Білок за п. 6, який відрізняється тим, що промивання здійснюють при 0,25 х SSC і 0,2 % SDS при 60 °C.

8. Рекombінантний білок, який має активність токсину проти комах-шкідника, що містить амінокислотну послідовність SEQ ID NO: 47.

5 9. Спосіб боротьби з комахою-шкідником, який передбачає пероральну доставку білка за п. 6 вказаному шкіднику.

10. Спосіб за п. 9, який відрізняється тим, що вказаний білок продукується і присутній у трансгенній рослині, яка є доступною для вказаного шкідника.

10

15

20

25

30

35

40

45

50

55

60

65

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

СПИСОК ПОСЛІДОВНОСТЕЙ

(i) GENERAL INFORMATION:

(ii) APPLICANT: Ensign, Jerald C
Bowen, David J
Petell, James
Facig, Raymond
Schoonover, Sue
French-Constant, Richard
Orr, Gregory L
Merlo, Donald J
Roberts, Jean L
Rocheleau, Thomas A
Blackburn, Michael B
Hey, Timothy D
Strickland, James A

(iii) НАЗВА ВИНАХОДУ: Insecticidal Protein Toxins From
Photorhabdus

(iii) ЧИСЛО ПОСЛІДОВНОСТЕЙ: 61

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Quarles & Brady
(B) STREET: 1 South Pinckney Street
(C) CITY: Madison
(D) STATE: WI
(E) COUNTRY: US
(F) ZIP: 53703

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/063,615
(B) FILING DATE: 18-MAY-1993

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/395,497
(B) FILING DATE: 28-FEB-1995

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/007,255
(B) FILING DATE: 06-NOV-1995

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/608,423
(B) FILING DATE: 28-FEB-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 09/705,484
(B) FILING DATE: 26-AUG-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seay, Nicholas J
(B) REGISTRATION NUMBER: 27386
(C) REFERENCE/DOCKET NUMBER: 960296.93804

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 608-251-5000
(B) TELEFAX: 608-251-9166

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 1:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 11 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 1:

Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 2:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 12 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 2:

Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Trp
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 3:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 19 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

U. 82485 C2

U A 8 2 4 8 5 C 2

(A) ДОВЖИНА: 14 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(v) ТИП ФРАГМЕНТУ: N-кінцевий

Ala Ser Pro Leu Ser Thr Ser Glu Leu Thr Ser Lys Leu Asn
1 5 10

(A) ДОВЖИНА: 9 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(v) ТИП ФРАГМЕНТУ: N-кінцевий

Ala Gly Asp Thr Ala Asn Ile Gly Asp
1 5

(A) ДОВЖИНА: 15 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(v) ТИП ФРАГМЕНТУ: N-кінцевий

Leu Gly Gly Ala Ala Thr Leu Leu Asp Leu Leu Leu Pro Gln Ile
 1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 7:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 11 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 7:

Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu
 1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 8:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 9 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 8:

Met Asn Leu Ala Ser Pro Leu Ile Ser
 1 5

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 9:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 16 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 9:

Met Ile Asn Leu Asp Ile Asn Glu Gln Asn Lys Ile Met Val Val Ser
1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 10:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 20 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 10:

Ala Ala Lys Asp Val Lys Phe Gly Ser Asp Ala Arg Val Lys Met Leu
1 5 10 15
Arg Gly Val Asn
20

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 11:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ: SEQ ID № 11:

- (A) ДОВЖИНА: 7515 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: CDS
- (B) МІСЦЕПОЛОЖЕННЯ: 1...7515

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 11:

ATG CAA AAC TCA TTA TCA AGC ACT ATC GAT ACT ATT TGT CAG AAA CTG	48
Met Gln Asn Ser Leu Ser Ser Thr Ile Asp Thr Ile Cys Gln Lys Leu	
1 5 10 15	
CAA TTA ACT TGT CCG GCG GAA ATT CCT TTG TAT CCC TTT GAT ACT TTC	96
Gln Leu Thr Cys Pro Ala Glu Ile Ala Leu Tyr Pro Phe Asp Thr Phe	
20 25 30	
CCG GAA AAA ACT CCG GGA ATG GTT AAT TCG CCG GAA GCA AAA CCG ATT	144
Arg Glu Lys Thr Arg Gly Met Val Asn Trp Gly Glu Ala Lys Arg Ile	
35 40 45	
TAT GAA ATT GCA CAA GCG GAA CAG GAT AGA AAC CTA CTT CAT GAA AAA	192
Tyr Glu Ile Ala Gln Ala Glu Gln Asp Arg Asn Leu Leu His Glu Lys	
50 55 60	
CCT ATT TTT GCC TAT GCT AAT CCG CTG CTG AAA AAC GCT GTT CCG TTG	240
Arg Ile Phe Ala Tyr Ala Asn Pro Leu Leu Lys Asn Ala Val Arg Leu	

U A 8 2 4 8 5 C 2

65	70	75	80	
GGT ACC CCG CAA ATC TTG GGT TTT ATA CAA GGT TAT AGT GAT CTC TTT	282			
Gly Thr Arg Gln Met Leu Gly Phe Ile Gln Gly Tyr Ser Asp Leu Phe				
85	90	95		
GGT AAT CGT GCT GAT AAC TAT GCG CCG CCG GCG TCG GTT GCA TCG ATC	336			
Gly Asn Arg Ala Asp Asn Tyr Ala Ala Pro Gly Ser Val Ala Ser Met				
100	105	110		
TTC TCA CCG GCG GCT TAT TTG ACC GAA TTG TAC CGT GAA GCC AAA AAC	384			
Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Lys Asn				
115	120	125		
TTG CAT GAC AGC AGC TCA ATT TAT TAC CTA GAT AAA CGT CCG CCG GAT	432			
Leu His Asp Ser Ser Ser Ile Tyr Tyr Leu Asp Lys Arg Arg Pro Asp				
130	135	140		
TTA GCA ACC TTA ATG CTC AGC CAG AAA AAT ATC GAT GAG GAA ATT TTA	480			
Leu Ala Ser Leu Met Ser Gln Lys Asn Met Asp Glu Glu Ile Ser				
145	150	155	160	
ACG CTC GCT CTC TCT AAT GAA TTG TCG CTT GCC GGG ATC GAA ACA AAA	528			
Thr Leu Ala Leu Ser Asn Glu Leu Cys Leu Ala Gly Ile Glu Thr Lys				
165	170	175		
ACA GGA AAA TCA CAA CAT GAA GTG ATG GAT ATG TTG TCA ACT TAT CGT	576			
Thr Gly Lys Ser Gln Asp Glu Val Met Asp Met Leu Ser Thr Tyr Arg				
180	185	190		
TTA AGT GGA GAG ACA CCG TAT CAT CAC GCT TAT GAA ACT GTT CGT GAA	624			
Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu				
195	200	205		
ATC GTT CAT GAA CCG GAT CCA CGA TTT CGT CAT TTG TCA CAG GCA CCC	672			
Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro				
210	215	220		
ATT GTT GCT GCT AAG CTC GAT CCG GTG ACT TTG TTG GGT ATT AGC TCG	720			
Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser				
225	230	235	240	
CAT ATT TCG CCA GAA CTC TAT AAC TTG CTC ATT GAG GAG ATC CCG GAA	768			
His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu				
245	250	255		
AAA CAT GAA GCC GCG CTT GAT ACC CTT TAT AAA ACA AAC TTT GGC GAT	816			
Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp				
260	265	270		
ATT ACT ACT GCT CAG TTA ATG TCC CCA AGT TAT CTC GCC CCG TAT TAT	864			
Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr				
275	280	285		
GCC CTC TCA CCG GAA GAT ATT GCC TAC GTG ACC ACT TCA TTA TCA CAT	912			
Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His				
290	295	300		
GTT GGA TAT AGC AGT GAT ATT CTC GTT ATT CCG TTG GTC GAT GGT GTG	960			
Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val				
305	310	315	320	
GGT AAG ATG GAA GTA GTT CGT GTT ACC CGA ACA CCA TCG GAT AAT TAT	1008			
Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr				
325	330	335		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

ACT AGT CAG ACC AAT TAT ATT GAG CTC TAT CCA CAG GGT GGC GAC AAT 1056
 Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn
 340 345 350
 TAT TTG ATC AAA TAC AAT CTA AGC AAT AGT TTT GGT TTG GAT GAT TTT 1104
 Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe
 355 360 365
 TAT CTG CAA TAT AAA GAT GGT TCC GCT GAT TGG ACT GAG ATT GCC CAT 1152
 Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His
 370 375 380
 AAT CCC TAT CCT GAT ATG GTC ATA AAT CAA AAG TAT GAA TCA CAG GCG 1200
 Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala
 385 390 395 400
 ACA ATC AAA CGT AGT GAC TCT GAC AAT ATA CTC AGT ATA GGC TTA CAA 1248
 Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln
 405 410 415
 AGA TGG CAT ACC GGT AGT TAT AAT TTT GCC GCC GCC AAT TTT AAA ATT 1296
 Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile
 420 425 430
 GAC CAA TAC TCC CCG AAA GCT TTC CTG CTT AAA ATC AAT AAG GCT ATT 1344
 Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile
 435 440 445
 CGG TTG CTC AAA GCT ACC GGC CTC TCT TTT GCT ACC TTG GAG CGT ATT 1392
 Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile
 450 455 460
 GTT GAT AGT GTT AAT AGC ACC AAA TCC ATC ACC GTT GAG GTA TTA AAC 1440
 Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn
 465 470 475 480
 AAG GTT TAT CGG GTA AAA TTC TAT ATT GAT CGT TAT GGC ATC AGT GAA 1488
 Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu
 485 490 495
 GAG ACA GCC GCT ATT TTG GCT AAT ATT AAT ATC TCT CAG CAA GCT GTT 1536
 Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val
 500 505 510
 GGC AAT CAG CTT AGC CAG TTT GAG CAA CTA TTT AAT CAC CCG CCG CTC 1584
 Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu
 515 520 525
 AAT GGT ATT CGC TAT GAA ATC AGT GAG GAC AAC TCC AAA CAT CTT CCT 1632
 Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro
 530 535 540
 AAT CCT GAT CTC AAC CTT AAA CCA GAC AGT ACC GGT GAT GAT CAA CCG 1680
 Asn Pro Asp Leu Asn Leu Lys Pro Asp Ser Thr Gly Asp Asp Gln Arg
 545 550 555 560
 AAG GCG GTT TTA AAA CGC GCG TTT CAG GTT AAC GCC AGT GAG TTG TAT 1728
 Lys Ala Val Leu Lys Arg Ala Phe Gln Val Asn Ala Ser Glu Leu Tyr
 565 570 575
 CAG ATC TTA TTG ATC ACT GAT CGT AAA GAA GAC GGT CTT ATC AAA AAT 1776
 Gln Met Leu Leu Ile Thr Asp Arg Lys Glu Asp Gly Val Ile Lys Asn
 580 585 590
 AAC TTA GAG AAT TTG TCT GAT CTG TAT TTG GTT AGT TTG CTC GCC CAG 1824
 Asn Leu Glu Asn Leu Ser Asp Leu Tyr Leu Val Ser Leu Leu Ala Gln

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

595	600	605	
ATT CAT AAC CTG ACT ATT GCT GAA TTG AAC ATT TTG TTG GTG ATT TGT 1872			
Ile His Asn Leu Thr Ile Ala Glu Leu Asn Ile Leu Leu Val Ile Cys			
610	615	620	
GGC TAT GGC GAC ACC AAC ATT TAT CAG ATT ACC GAC GAT AAT TTA GGC 1920			
Gly Tyr Gly Asp Thr Asn Ile Tyr Gln Ile Thr Asp Asp Asn Leu Ala			
625	630	635	640
AAA ATA CTG GAA ACA TTG TTG TGG ATC ACT CAA TGG TTG AAG ACC CAA 1958			
Lys Ile Val Glu Thr Leu Leu Trp Ile Thr Gln Trp Leu Lys Thr Gln			
645	650	655	
AAA TGG ACA GTT ACC GAC CTG TTT CTG ATG ACC ACG GCC ACT TAC AGC 2016			
Lys Trp Thr Val Thr Asp Leu Phe Leu Met Thr Thr Ala Thr Tyr Ser			
660	665	670	
ACC ACT TTA ACC CCA GAA ATT AGC AAT CTG ACG GCT ACG TTG TCT TCA 2064			
Thr Thr Leu Thr Pro Glu Ile Ser Asn Leu Thr Ala Thr Leu Ser Ser			
675	680	685	
ACT TTG CAT GGC AAA GAG AGT CTG ATT GGG GAA GAT CTG AAA AGA GCA 2112			
Thr Leu His Gly Lys Glu Ser Leu Ile Gly Glu Asp Leu Lys Arg Ala			
690	695	700	
ATG GCG CCT TGC TTC ACT TCG GCT TTG CAT TTG ACT TCT CAA GAA GTT 2150			
Met Ala Pro Cys Phe Thr Ser Ala Leu His Leu Thr Ser Gln Glu Val			
705	710	715	720
GGC TAT GAC CTG CTG TTG TGG ATA GAC CAG ATT CAA CCG GCA CAA ATA 2208			
Ala Tyr Asp Leu Leu Leu Trp Ile Asp Gln Ile Gln Pro Ala Gln Ile			
725	730	735	
ACT GTT GAT GCG TTT TCG GAA GAA CTG CAA ACA ACA CCA ACC AGC TTG 2256			
Thr Val Asp Gly Phe Trp Glu Glu Val Gln Thr Thr Pro Thr Ser Leu			
740	745	750	
AAG GTG ATT ACC TTT GCT CAG GTC CTG GCA CAA TTG AGC CTG ATC TAT 2304			
Lys Val Ile Thr Phe Ala Gln Val Leu Ala Gln Leu Ser Leu Ile Tyr			
755	760	765	
CGT CGT ATT GGG TTA AGT GAA ACG GAA CTG TCA CTG ATC GTG ACT CAA 2352			
Arg Arg Ile Gly Leu Ser Glu Thr Glu Leu Ser Leu Ile Val Thr Gln			
770	775	780	
TCT TCT CTG CTA GTG GCA GGC AAA AGC ATA CTG GAT CAC GGT CTG TTA 2400			
Ser Ser Leu Leu Val Ala Gly Lys Ser Ile Leu Asp His Gly Leu Leu			
785	790	795	800
ACC CTG ATG GCC TTG GAA GGT TTT CAT ACC TGG GTT AAT GGC TTG GCG 2448			
Thr Leu Met Ala Leu Glu Gly Phe His Thr Trp Val Asn Gly Leu Gly			
805	810	815	
CAA CAT GCC TCC TTG ATA TTG GCG GCG TTG AAA GAC GGA GCC TTG ACA 2496			
Gln His Ala Ser Leu Ile Leu Ala Ala Leu Lys Asp Gly Ala Leu Thr			
820	825	830	
GTT ACC GAT GTA GCA CAA GCT ATG AAT AAG GAG GAA TCT CTC CTA CAA 2544			
Val Thr Asp Val Ala Gln Ala Met Asn Lys Glu Glu Ser Leu Leu Gln			
835	840	845	
ATG GCA GCT AAT CAG GTG GAG AAG GAT CTA ACA AAA CTG ACC AGT TCG 2592			
Met Ala Ala Asn Gln Val Glu Lys Asp Leu Thr Lys Leu Thr Ser Trp			
850	855	860	

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

ACA CAG ATT GAC GCT ATT CTG CAA TCG TTA CAG ATG TCT TCG GCC TTG 2540
 Thr Gln Ile Asp Ala Ile Leu Gln Trp Leu Gln Met Ser Ser Ala Leu
 865 870 875 880

GCG GTT TCT CCA CTG GAT CTG CCA GCG ATG ATG GCC CTG AAA TAT GGG 2683
 Ala Val Ser Pro Leu Asp Leu Ala Gly Met Met Ala Leu Lys Tyr Gly
 885 890 895

ATA GAT CAT AAC TAT GCT GCC TCG CAA GCT GCG GCG GCT GCG CTG ATG 2736
 Ile Asp His Asn Tyr Ala Ala Trp Gln Ala Ala Ala Ala Ala Leu Met
 900 905 910

GCT GAT CAT GCT AAT CAG CCA CAG AAA AAA CTG GAT GAG ACC TTC AGT 2784
 Ala Asp His Ala Asn Gln Ala Gln Lys Lys Leu Asp Glu Thr Phe Ser
 915 920 925

AAG GCA TTA TGT AAC TAT TAT ATT AAT GCT GTT GTC GAT AGT GCT GCT 2832
 Lys Ala Leu Cys Asn Tyr Tyr Ile Asn Ala Val Val Asp Ser Ala Ala
 930 935 940

GGA GTA CGT GAT CGT AAC GGT TTA TAT ACC TAT TTG CTG ATT GAT AAT 2880
 Gly Val Arg Asp Arg Asn Gly Leu Tyr Thr Tyr Leu Leu Ile Asp Asn
 945 950 955 960

CAG GTT TCT GCC GAT GTG ATC ACT TCA CGT ATT GCA GAA GCT ATC GCC 2928
 Gln Val Ser Ala Asp Val Ile Thr Ser Arg Ile Ala Glu Ala Ile Ala
 965 970 975

GCT ATT CAA CTG TAC GTT AAC CGG GCT TTA AAC CGA GAT GAA GGT CAG 2976
 Gly Ile Gln Leu Tyr Val Asn Arg Ala Leu Asn Arg Asp Glu Gly Gln
 980 985 990

CTT GCA TCG GAC GTT AGT ACC CGT CAG TTC TTC ACT GAC TCG GAA CGT 3024
 Leu Ala Ser Asp Val Ser Thr Arg Gln Phe Phe Thr Asp Trp Glu Arg
 995 1000 1005

TAC AAT AAA CGT TAC AGT ACT TGG GCT GGT GTC TCT GAA CTG GTC TAT 3072
 Tyr Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Glu Leu Val Tyr
 1010 1015 1020

TAT CCA GAA AAC TAT GTT GAT CCC ACT CAG CGC ATT GGG CAA ACC AAA 3120
 Tyr Pro Glu Asn Tyr Val Asp Pro Thr Gln Arg Ile Gly Gln Thr Lys
 1025 1030 1035 1040

ATG ATG GAT GCG CTG TTG CAA TCC ATC AAC CAG AGC CAG CTA AAT GCC 3168
 Met Met Asp Ala Leu Leu Gln Ser Ile Asn Gln Ser Gln Leu Asn Ala
 1045 1050 1055

GAT ACC GTC GAA GAT GCT TTC AAA ACT TAT TTG ACC AGC TTT GAG CAG 3216
 Asp Thr Val Glu Asp Ala Phe Lys Thr Tyr Leu Thr Ser Phe Glu Gln
 1060 1065 1070

GTA GCA AAT CTG AAA GTA ATT AGT GCT TAC CAC GAT AAT GTG AAT GTG 3264
 Val Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Val Asn Val
 1075 1080 1085

GAT CAA GGA TTA ACT TAT TTT ATC GGT ATC GAC CAA GCA GCT CCG GGT 3312
 Asp Gln Gly Leu Thr Tyr Phe Ile Gly Ile Asp Gln Ala Ala Pro Gly
 1090 1095 1100

ACG TAT TAC TCG CGT AGT GTT GAT CAC AGC AAA TGT GAA AAT GGC AAG 3360
 Thr Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Cys Glu Asn Gly Lys
 1105 1110 1115 1120

TTT GCC GCT AAT GCT TCG GGT GAG TCG AAT AAA ATT ACC TGT GCT GTC 3408
 Phe Ala Ala Asn Ala Trp Gly Glu Trp Asn Lys Ile Thr Cys Ala Val

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

1125	1130	1135
AAT GCT TCG AAA AAT ATC ATC GGT CCG GTT GTT TAT ATG TCC CCG TTA 3456 Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu 1140 1145 1150		
TAT CTC CTA TCG CTC GAG CAG CAA TCA AAG AAA AGT GAT GAT GGT AAA 3504 Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys 1155 1160 1165		
ACC ACG ATT TAT CAA TAT AAC TTA AAA CTG GCT CAT ATT CGT TAC GAC 3552 Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp 1170 1175 1180		
GGT AGT TCG AAT ACA CCA TTT ACT TTT GAT CTG ACA GAA AAG GTA AAA 3600 Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys 1185 1190 1195 1200		
AAT TAC ACG TCG AGT ACT GAT GCT GCT GAA TCT TTA GGG TTC TAT TGT 3648 Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys 1205 1210 1215		
ACT GGT TAT CAA GCG GAA GAC ACT CTA TTA GTT ATG TTC TAT TCG ATG 3696 Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met 1220 1225 1230		
CAG AGT AGT TAT AGC TCC TAT ACC GAT AAT AAT GCG CCG GTC ACT GCG 3744 Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly 1235 1240 1245		
CTA TAT ATT TTC GCT GAT ATG TCA TCA GAC AAT ATG ACG AAT GCA CAA 3792 Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln 1250 1255 1260		
GCA ACT AAC TAT TCG AAT AAC AGT TAT CCG CAA TTT GAT ACT GTG ATG 3840 Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met 1265 1270 1275 1280		
GCA GAT CCG GAT AGC GAC AAT AAA AAA GTC ATA ACC AGA AGA GTT AAT 3888 Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn 1285 1290 1295		
AAC CGT TAT GCG GAG GAT TAT GAA ATT CCT TCC TCT GTG ACA AGT AAC 3936 Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn 1300 1305 1310		
AGT AAT TAT TCT TCG GCT GAT CAC AGT TTA ACC ATG CTT TAT GGT GGT 3984 Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly 1315 1320 1325		
AGT GTT CCT AAT ATT ACT TTT GAA TCG CCG GCA GAA GAT TTA AGG CTA 4032 Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu 1330 1335 1340		
TCT ACC AAT ATG GCA TTG AGT ATT ATT CAT AAT GGA TAT CCG GGA ACC 4080 Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr 1345 1350 1355 1360		
CGC CGT ATA CAA TGT AAT CTT ATG AAA CAA TAC GCT TCA TTA GGT GAT 4128 Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp 1365 1370 1375		
AAA TTT ATA ATT TAT GAT TCA TCA TTT GAT GAT GCA AAC CGT TTT AAT 4176 Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn 1380 1385 1390		

U A 8 2 4 8 5 C 2

СПИСОК ПОСЛІДОВНОСТЕЙ

(i) GENERAL INFORMATION:

(ii) APPLICANT: Ensign, Jerald C
Bowen, David J
Petell, James
Facig, Raymond
Schoonover, Sue
French-Constant, Richard
Orr, Gregory L
Merlo, Donald J
Roberts, Jean L
Rocheleau, Thomas A
Blackburn, Michael B
Hey, Timothy D
Strickland, James A

(iii) НАЗВА ВИНАХОДУ: Insecticidal Protein Toxins From
Photorhabdus

(iii) ЧИСЛО ПОСЛІДОВНОСТЕЙ: 61

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Quarles & Brady
(B) STREET: 1 South Pinckney Street
(C) CITY: Madison
(D) STATE: WI
(E) COUNTRY: US
(F) ZIP: 53703

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/063,615
(B) FILING DATE: 18-MAY-1993

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/395,497
(B) FILING DATE: 28-FEB-1995

(ix) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/007,255
(B) FILING DATE: 06-NOV-1995

(x) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/608,423
(B) FILING DATE: 28-FEB-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 09/705,484
(B) FILING DATE: 26-AUG-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seay, Nicholas J
(B) REGISTRATION NUMBER: 27386
(C) REFERENCE/DOCKET NUMBER: 960296.93804

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 608-251-5000
(B) TELEFAX: 608-251-9166

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 1:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 11 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 1:

Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 2:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 12 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 2:

Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Trp
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 3:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(A) ДОВЖИНА: 19 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

Leu Gly Gly Ala Ala Thr Leu Leu Asp Leu Leu Leu Pro Gln Ile
 1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 7:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 11 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 7:

Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu
 1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 8:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 9 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 8:

Met Asn Leu Ala Ser Pro Leu Ile Ser
 1 5

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 9:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 16 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 9:

Met Ile Asn Leu Asp Ile Asn Glu Gln Asn Lys Ile Met Val Val Ser
1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 10:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 20 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 10:

Ala Ala Lys Asp Val Lys Phe Gly Ser Asp Ala Arg Val Lys Met Leu
1 5 10 15
Arg Gly Val Asn
20

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 11:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ: SEQ ID № 11:

- (A) ДОВЖИНА: 7515 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: CDS
- (B) МІСЦЕПОЛОЖЕННЯ: 1...7515

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 11:

ATG CAA AAC TCA TTA TCA AGC ACT ATC GAT ACT ATT TGT CAG AAA CTG	48
Met Gln Asn Ser Leu Ser Ser Thr Ile Asp Thr Ile Cys Gln Lys Leu	
1 5 10 15	
CAA TTA ACT TGT CCG GCG GAA ATT CCT TTG TAT CCC TTT GAT ACT TTC	96
Gln Leu Thr Cys Pro Ala Glu Ile Ala Leu Tyr Pro Phe Asp Thr Phe	
20 25 30	
CCG GAA AAA ACT CCG GGA ATG GTT AAT TCG CCG GAA GCA AAA CCG ATT	144
Arg Glu Lys Thr Arg Gly Met Val Asn Trp Gly Glu Ala Lys Arg Ile	
35 40 45	
TAT GAA ATT GCA CAA GCG GAA CAG GAT AGA AAC CTA CTT CAT GAA AAA	192
Tyr Glu Ile Ala Gln Ala Glu Gln Asp Arg Asn Leu Leu His Glu Lys	
50 55 60	
CCT ATT TTT GCC TAT GCT AAT CCG CTG CTG AAA AAC GCT GTT CCG TTG	240
Arg Ile Phe Ala Tyr Ala Asn Pro Leu Leu Lys Asn Ala Val Arg Leu	

65	70	75	80	
GGT ACC CCG CAA ATC TTG GGT TTT ATA CAA GGT TAT AGT GAT CTC TTT	282			
Gly Thr Arg Gln Met Leu Gly Phe Ile Gln Gly Tyr Ser Asp Leu Phe				
85	90	95		
GGT AAT CGT GCT GAT AAC TAT GCG GCG CCG GCG TCG GTT GCA TCG ATC	336			
Gly Asn Arg Ala Asp Asn Tyr Ala Ala Pro Gly Ser Val Ala Ser Met				
100	105	110		
TTC TCA CCG GCG GCT TAT TTG ACC GAA TTG TAC CGT GAA GCC AAA AAC	384			
Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Lys Asn				
115	120	125		
TTG CAT GAC AGC AGC TCA ATT TAT TAC CTA GAT AAA CGT CCG CCG GAT	432			
Leu His Asp Ser Ser Ser Ile Tyr Tyr Leu Asp Lys Arg Arg Pro Asp				
130	135	140		
TTA GCA ACC TTA ATG CTC AGC CAG AAA AAT ATC GAT GAG GAA ATT TTA	480			
Leu Ala Ser Leu Met Ser Gln Lys Asn Met Asp Glu Glu Ile Ser				
145	150	155	160	
ACG CTC GCT CTC TCT AAT GAA TTG TCG CTT GCC GGG ATC GAA ACA AAA	528			
Thr Leu Ala Leu Ser Asn Glu Leu Cys Leu Ala Gly Ile Glu Thr Lys				
165	170	175		
ACA GGA AAA TCA CAA CAT GAA GTG ATG GAT ATG TTG TCA ACT TAT CGT	576			
Thr Gly Lys Ser Gln Asp Glu Val Met Asp Met Leu Ser Thr Tyr Arg				
180	185	190		
TTA AGT GGA GAG ACA CCT TAT CAT CAC GCT TAT GAA ACT GTT CGT GAA	624			
Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu				
195	200	205		
ATC GTT CAT GAA CGT GAT CCA CGA TTT CGT CAT TTG TCA CAG GCA CCC	672			
Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro				
210	215	220		
ATT GTT GCT GCT AAG CTC GAT CCT GTG ACT TTG TTG GGT ATT AGC TCC	720			
Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser				
225	230	235	240	
CAT ATT TCG CCA GAA CTC TAT AAC TTG CTC ATT GAG GAG ATC CCG GAA	768			
His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu				
245	250	255		
AAA CAT GAA GCC GCG CTT GAT ACC CTT TAT AAA ACA AAC TTT GGC GAT	816			
Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp				
260	265	270		
ATT ACT ACT GCT CAG TTA ATG TCC CCA AGT TAT CTC GCC CCG TAT TAT	864			
Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr				
275	280	285		
GCC CTC TCA CCG GAA GAT ATT GCC TAC GTG ACC ACT TCA TTA TCA CAT	912			
Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His				
290	295	300		
GTT GGA TAT AGC AGT GAT ATT CTC GTT ATT CCG TTG GTC GAT GGT GTG	960			
Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val				
305	310	315	320	
GGT AAG ATG GAA GTA GTT CGT GTT ACC CGA ACA CCA TCG GAT AAT TAT	1008			
Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr				
325	330	335		

U A 8 2 4 8 5 C 2

ACC AGT CAG ACC AAT TAT ATT GAG CTC TAT CCA CAG GGT GGC GAC AAT 1056
 Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn
 340 345 350
 TAT TTG ATC AAA TAC AAT CTA AGC AAT AGT TTT GGT TTG GAT GAT TTT 1104
 Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe
 355 360 365
 TAT CTG CAA TAT AAA GAT GGT TCC GCT GAT TGG ACT GAG ATT GCC CAT 1152
 Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His
 370 375 380
 AAT CCC TAT CCT GAT ATG GTC ATA AAT CAA AAG TAT GAA TCA CAG GCG 1200
 Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala
 385 390 395 400
 ACA ATC AAA CGT AGT GAC TCT GAC AAT ATA CTC AGT ATA GGC TTA CAA 1248
 Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln
 405 410 415
 AGA TGG CAT ACC GGT AGT TAT AAT TTT GCC GCC GCC AAT TTT AAA ATT 1296
 Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile
 420 425 430
 GAC CAA TAC TCC CCG AAA GCT TTC CTG CTT AAA ATC AAT AAG GCT ATT 1344
 Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile
 435 440 445
 CGG TTG CTC AAA GCT ACC GGC CTC TCT TTT GCT ACC TTG GAG CGT ATT 1392
 Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile
 450 455 460
 GTT GAT AGT GTT AAT AGC ACC AAA TCC ATC ACC GTT GAG GTA TTA AAC 1440
 Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn
 465 470 475 480
 AAG GTT TAT CGG GTA AAA TTC TAT ATT GAT CGT TAT GGC ATC AGT GAA 1488
 Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu
 485 490 495
 GAG ACA GCC GCT ATT TTG GCT AAT ATT AAT ATC TCT CAG CAA GCT GTT 1536
 Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val
 500 505 510
 GGC AAT CAG CTT AGC CAG TTT GAG CAA CTA TTT AAT CAC CCG CCG CTC 1584
 Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu
 515 520 525
 AAT GGT ATT CGC TAT GAA ATC AGT GAG GAC AAC TCC AAA CAT CTT CCT 1632
 Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro
 530 535 540
 AAT CCT GAT CTC AAC CTT AAA CCA GAC AGT ACC GGT GAT GAT CAA CCG 1680
 Asn Pro Asp Leu Asn Leu Lys Pro Asp Ser Thr Gly Asp Asp Gln Arg
 545 550 555 560
 AAG GCG GTT TTA AAA CGC GCG TTT CAG GTT AAC GCC AGT GAG TTG TAT 1728
 Lys Ala Val Leu Lys Arg Ala Phe Gln Val Asn Ala Ser Glu Leu Tyr
 565 570 575
 CAG ATC TTA TTG ATC ACT GAT CGT AAA GAA GAC GGT CTT ATC AAA AAT 1776
 Gln Met Leu Leu Ile Thr Asp Arg Lys Glu Asp Gly Val Ile Lys Asn
 580 585 590
 AAC TTA GAG AAT TTG TCT GAT CTG TAT TTG GTT AGT TTG CTC GCC CAG 1824
 Asn Leu Glu Asn Leu Ser Asp Leu Tyr Leu Val Ser Leu Leu Ala Gln

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

595	600	605	
ATT CAT AAC CTG ACT ATT GCT GAA TTG AAC ATT TTG TTG GTG ATT TGT 1872			
Ile His Asn Leu Thr Ile Ala Glu Leu Asn Ile Leu Leu Val Ile Cys			
610	615	620	
GGC TAT GGC GAC ACC AAC ATT TAT CAG ATT ACC GAC GAT AAT TTA GCC 1920			
Gly Tyr Gly Asp Thr Asn Ile Tyr Gln Ile Thr Asp Asp Asn Leu Ala			
625	630	635	640
AAA ATA CTG GAA ACA TTG TTG TGG ATC ACT CAA TGG TTG AAG ACC CAA 1958			
Lys Ile Val Glu Thr Leu Leu Trp Ile Thr Gln Trp Leu Lys Thr Gln			
645	650	655	
AAA TGG ACA GTT ACC GAC CTG TTT CTG ATG ACC ACG GCC ACT TAC AGC 2016			
Lys Trp Thr Val Thr Asp Leu Phe Leu Met Thr Thr Ala Thr Tyr Ser			
660	665	670	
ACC ACT TTA ACC CCA GAA ATT AGC AAT CTG ACG GCT ACG TTG TCT TCA 2064			
Thr Thr Leu Thr Pro Glu Ile Ser Asn Leu Thr Ala Thr Leu Ser Ser			
675	680	685	
ACT TTG CAT GGC AAA GAG AGT CTG ATT GGG GAA GAT CTG AAA AGA GCA 2112			
Thr Leu His Gly Lys Glu Ser Leu Ile Gly Glu Asp Leu Lys Arg Ala			
690	695	700	
ATG GCG CCT TGC TTC ACT TCG GCT TTG CAT TTG ACT TCT CAA GAA GTT 2150			
Met Ala Pro Cys Phe Thr Ser Ala Leu His Leu Thr Ser Gln Glu Val			
705	710	715	720
GGC TAT GAC CTG CTG TTG TGG ATA GAC CAG ATT CAA CCG GCA CAA ATA 2208			
Ala Tyr Asp Leu Leu Leu Trp Ile Asp Gln Ile Gln Pro Ala Gln Ile			
725	730	735	
ACT GTT GAT GCG TTT TCG GAA GAA CTG CAA ACA ACA CCA ACC AGC TTG 2256			
Thr Val Asp Gly Phe Trp Glu Glu Val Gln Thr Thr Pro Thr Ser Leu			
740	745	750	
AAG GTG ATT ACC TTT GCT CAG GTC CTG GCA CAA TTG AGC CTG ATC TAT 2304			
Lys Val Ile Thr Phe Ala Gln Val Leu Ala Gln Leu Ser Leu Ile Tyr			
755	760	765	
CGT CGT ATT GGG TTA AGT GAA ACG GAA CTG TCA CTG ATC GTG ACT CAA 2352			
Arg Arg Ile Gly Leu Ser Glu Thr Glu Leu Ser Leu Ile Val Thr Gln			
770	775	780	
TCT TCT CTG CTA GTG GCA GCG AAA AGC ATA CTG GAT CAC GGT CTG TTA 2400			
Ser Ser Leu Leu Val Ala Gly Lys Ser Ile Leu Asp His Gly Leu Leu			
785	790	795	800
ACC CTG ATG GCC TTG GAA GGT TTT CAT ACC TGG GTT AAT GGC TTG GCG 2448			
Thr Leu Met Ala Leu Glu Gly Phe His Thr Trp Val Asn Gly Leu Gly			
805	810	815	
CAA CAT GCC TCC TTG ATA TTG GCG GCG TTG AAA GAC GGA GCC TTG ACA 2496			
Gln His Ala Ser Leu Ile Leu Ala Ala Leu Lys Asp Gly Ala Leu Thr			
820	825	830	
GTT ACC GAT GTA GCA CAA GCT ATG AAT AAG GAG GAA TCT CTC CTA CAA 2544			
Val Thr Asp Val Ala Gln Ala Met Asn Lys Glu Glu Ser Leu Leu Gln			
835	840	845	
ATG GCA GCT AAT CAG GTG GAG AAG GAT CTA ACA AAA CTG ACC AGT TCG 2592			
Met Ala Ala Asn Gln Val Glu Lys Asp Leu Thr Lys Leu Thr Ser Trp			
850	855	860	

U A 8 2 4 8 5 C 2

ACA CAG ATT GAC GCT ATT CTG CAA TCG TTA CAG ATG TCT TCG GCC TTG 2640
 Thr Gln Ile Asp Ala Ile Leu Gln Trp Leu Gln Met Ser Ser Ala Leu
 865 870 875 880
 GCG GTT TCT CCA CTG GAT CTG CCA GCG ATG ATG GCC CTG AAA TAT GGG 2683
 Ala Val Ser Pro Leu Asp Leu Ala Gly Met Met Ala Leu Lys Tyr Gly
 885 890 895
 ATA GAT CAT AAC TAT GCT GCC TCG CAA GCT GCG GCG GCT GCG CTG ATG 2736
 Ile Asp His Asn Tyr Ala Ala Trp Gln Ala Ala Ala Ala Ala Leu Met
 900 905 910
 GCT GAT CAT GCT AAT CAG CCA CAG AAA AAA CTG GAT GAG ACC TTC AGT 2784
 Ala Asp His Ala Asn Gln Ala Gln Lys Lys Leu Asp Glu Thr Phe Ser
 915 920 925
 AAG GCA TTA TGT AAC TAT TAT ATT AAT GCT GTT GTC GAT AGT GCT GCT 2832
 Lys Ala Leu Cys Asn Tyr Tyr Ile Asn Ala Val Val Asp Ser Ala Ala
 930 935 940
 CGA GTA CGT GAT CGT AAC GGT TTA TAT ACC TAT TTG CTG ATT GAT AAT 2880
 Gly Val Arg Asp Arg Asn Gly Leu Tyr Thr Tyr Leu Leu Ile Asp Asn
 945 950 955 960
 CAG GTT TCT GCC GAT GTC ATC ACT TCA CGT ATT GCA GAA GCT ATC GCC 2928
 Gln Val Ser Ala Asp Val Ile Thr Ser Arg Ile Ala Glu Ala Ile Ala
 965 970 975
 GGT ATT CAA CTG TAC GTT AAC CGG GCT TTA AAC CGA GAT GAA GGT CAG 2976
 Gly Ile Gln Leu Tyr Val Asn Arg Ala Leu Asn Arg Asp Glu Gly Gln
 980 985 990
 CTT GCA TCG GAC GTT AGT ACC CGT CAG TTC TTC ACT GAC TCG GAA CGT 3024
 Leu Ala Ser Asp Val Ser Thr Arg Gln Phe Phe Thr Asp Trp Glu Arg
 995 1000 1005
 TAC AAT AAA CGT TAC AGT ACT TGG GCT GGT GTC TCT GAA CTG GTC TAT 3072
 Tyr Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Glu Leu Val Tyr
 1010 1015 1020
 TAT CCA GAA AAC TAT GTT GAT CCC ACT CAG CGC ATT GGG CAA ACC AAA 3120
 Tyr Pro Glu Asn Tyr Val Asp Pro Thr Gln Arg Ile Gly Gln Thr Lys
 1025 1030 1035 1040
 ATG ATG GAT GCG CTG TTG CAA TCC ATC AAC CAG AGC CAG CTA AAT GCC 3168
 Met Met Asp Ala Leu Leu Gln Ser Ile Asn Gln Ser Gln Leu Asn Ala
 1045 1050 1055
 GAT ACC GTC GAA GAT GCT TTC AAA ACT TAT TTG ACC AGC TTT GAG CAG 3216
 Asp Thr Val Glu Asp Ala Phe Lys Thr Tyr Leu Thr Ser Phe Glu Gln
 1060 1065 1070
 GTA GCA AAT CTG AAA GTA ATT AGT GCT TAC CAC GAT AAT GTG AAT GTG 3264
 Val Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Val Asn Val
 1075 1080 1085
 GAT CAA GGA TTA ACT TAT TTT ATC GGT ATC GAC CAA GCA GCT CCG GGT 3312
 Asp Gln Gly Leu Thr Tyr Phe Ile Gly Ile Asp Gln Ala Ala Pro Gly
 1090 1095 1100
 ACG TAT TAC TCG CGT AGT GTT GAT CAC AGC AAA TGT GAA AAT GGC AAG 3360
 Thr Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Cys Glu Asn Gly Lys
 1105 1110 1115 1120
 TTT GCC GCT AAT GCT TCG GGT GAG TCG AAT AAA ATT ACC TGT GCT GTC 3408
 Phe Ala Ala Asn Ala Trp Gly Glu Trp Asn Lys Ile Thr Cys Ala Val

U A 8 2 4 8 5 C 2

1125	1130	1135
AAT GCT TCG AAA AAT ATC ATC GGT CCG GTT GTT TAT ATG TCC CCG TTA 3456 Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu 1140 1145 1150		
TAT CTC CTA TCG CTC GAG CAG CAA TCA AAG AAA AGT GAT GAT GGT AAA 3504 Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys 1155 1160 1165		
ACC ACG ATT TAT CAA TAT AAC TTA AAA CTG GCT CAT ATT CGT TAC GAC 3552 Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp 1170 1175 1180		
GGT AGT TCG AAT ACA CCA TTT ACT TTT GAT CTG ACA GAA AAG GTA AAA 3600 Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys 1185 1190 1195 1200		
AAT TAC ACG TCG AGT ACT GAT GCT GCT GAA TCT TTA GGG TTC TAT TGT 3648 Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys 1205 1210 1215		
ACT GGT TAT CAA GCG GAA GAC ACT CTA TTA GTT ATG TTC TAT TCG ATG 3696 Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met 1220 1225 1230		
CAG AGT AGT TAT AGC TCC TAT ACC GAT AAT AAT GCG CCG GTC ACT GCG 3744 Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly 1235 1240 1245		
CTA TAT ATT TTC GCT GAT ATG TCA TCA GAC AAT ATG ACG AAT GCA CAA 3792 Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln 1250 1255 1260		
GCA ACT AAC TAT TCG AAT AAC AGT TAT CCG CAA TTT GAT ACT GTG ATG 3840 Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met 1265 1270 1275 1280		
GCA GAT CCG GAT AGC GAC AAT AAA AAA GTC ATA ACC AGA AGA GTT AAT 3888 Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn 1285 1290 1295		
AAC CGT TAT GCG GAG GAT TAT GAA ATT CCT TCC TCT GTG ACA AGT AAC 3936 Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn 1300 1305 1310		
AGT AAT TAT TCT TCG GCT GAT CAC AGT TTA ACC ATG CTT TAT GGT GGT 3984 Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly 1315 1320 1325		
AGT GTT CCT AAT ATT ACT TTT GAA TCG GCG GCA GAA GAT TTA AGG CTA 4032 Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu 1330 1335 1340		
TCT ACC AAT ATG GCA TTG AGT ATT ATT CAT AAT GGA TAT GCG GGA ACC 4080 Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr 1345 1350 1355 1360		
GGC CGT ATA CAA TGT AAT CTT ATG AAA CAA TAC GCT TCA TTA GGT GAT 4128 Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp 1365 1370 1375		
AAA TTT ATA ATT TAT GAT TCA TCA TTT GAT GAT GCA AAC CGT TTT AAT 4176 Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn 1380 1385 1390		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

CTG GTG CGA TTC TTT AAA TTC CGA AAA GAG GAG AAC TCA GAT GAT AGT 4311
Leu Val Pro Leu Phe Lys Phe Gly Lys Asp Glu Asn Ser Asp Asp Ser
1395 1400 1405

ATT TGT ATA TAT AAT GAA AAC GGT TCT TCT GAA GAT AAG AAG TCG TAT 4312
Ile Cys Ile Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr
1410 1415 1420

TTT TCT TCG AAA GAT GAC AAT AAA ACA GCG GAT TAT AAT GGT CGA ACT 4320
Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
1425 1430 1435 1440

GAA TGT ATA GAT GGT CGA ACC AGT AAC AAA GAT TTT TAT TAT AAT CTC 4350
Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
1445 1450 1455

CAG GAG ATT GAA GTA ATT AGT GTT ACT GGT GCG TAT TCG TCG AGT TAT 4410
Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
1460 1465 1470

AAA ATA TCG AAC CCG ATT AAT ATC AAT ACG GCG ATT GAT AGT GGT AAA 4460
Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
1475 1480 1485

GTA AAA GTC ACC GTA AAA GCG GGT GGT GAC GAT CAA ATC TTT ACT GCT 4510
Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
1490 1495 1500

GAT AAT AGT ACC TAT GTT CCT CAG CAA CCG GCA CCC AGT TTT GAG GAG 4560
Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
1505 1510 1515 1520

ATG ATT TAT CAG TTC AAT AAC CTG ACA ATA GAT TGT AAG AAT TTA AAT 4600
Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
1525 1530 1535

TTC ATC GAC AAT CAG GCA CAT ATT GAG ATT GAT TTC ACC GCT ACG GCA 4650
Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
1540 1545 1550

CAA GAT GCG CGA TTC TTG GGT GCA GAA ACT TTT ATT ATC CCG GTA ACT 4700
Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
1555 1560 1565

AAA AAA GTT CTC GGT ACT GAG AAC GTG ATT GCG TTA TAT AGC GAA AAT 4752
Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn
1570 1575 1580

AAC GGT GTT CAA TAT ATG CAA ATT GCG GCA TAT CGT ACC CGT TTG AAT 4800
Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
1585 1590 1595 1600

ACG TTA TTC GCT CAA CAG TTG GTT ACC CGT GGT AAT CGT GCG ATT GAT 4840
Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
1605 1610 1615

GCA GTG CTC AGT ATG GAA ACT CAG AAT ATT CAG GAA CCG CAA TTA CGA 4890
Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
1620 1625 1630

GCG GCG ACA TAT GTG CAG GTT GTG TTG GAT AAA TAT GAT GAG TCT ATT 4940
Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
1635 1640 1645

CAT GCG ACT AAT AAA ACC TTT GCT ATT GAA TAT GTT GAT ATA TTT AAA 4992
His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys

U A 8 2 4 8 5 C 2

1650	1655	1660
GAG AAC GAT AGT TTT GTG ATT TAT CAA GGA GAA CTT AGC GAA ACA AGT 5040 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser 1655 1670 1675 1680		
CAA ACT GTT GTG AAA GTT TTC TTA TCC TAT TTT ATA GAG GCG ACT GGA 5088 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly 1625 1690 1695		
AAT AAG AAC CAC TTA TGG GTA CGT GCT AAA TAC CAA AAG GAA ACG ACT 5136 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr 1700 1705 1710		
GAT AAG ATC TTG TTC GAC CGT ACT GAT GAG AAA GAT CCG CAC GGT TGG 5184 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp 1715 1720 1725		
TTT CTC ACC GAC GAT CAC AAG ACC TTT AGT GGT CTC TCT TCC GCA CAG 5232 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln 1730 1735 1740		
GCA TTA AAG AAC GAC AGT GAA CCG ATG GAT TTC TCT GCG GCC AAT GCT 5280 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala 1745 1750 1755 1760		
CTC TAT TTC TGG GAA CTG TTC TAT TAC ACG CCG ATG ATG ATG GCT CAT 5328 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His 1765 1770 1775		
GCT TTG TTG CAG GAA CAG AAT TTT GAT GCG GCG AAC CAT TGG TTC GGT 5376 Arg Leu Leu Gln Glu Gln Asn Phe Asp Ala Ala Asn His Trp Phe Arg 1780 1785 1790		
TAT GTC TGG AGT CCA TCC GGT TAT ATC GTT GAT GGT AAA ATT GCT ATC 5424 Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile 1795 1800 1805		
TAC CAC TGG AAC CTG CGA CCG CTG GAA GAA GAC ACC AGT TGG AAT GCA 5472 Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala 1810 1815 1820		
CAA CAA CTG GAC TCC ACC GAT CCA GAT GCT GTA GCC CAA GAT GAT CCG 5520 Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro 1825 1830 1835 1840		
ATG CAC TAC AAG GTG GCT ACC TTT ATG GCG ACG TTG GAT CTG CTA ATG 5568 Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met 1845 1850 1855		
GCC CGT GGT GAT GCT GCT TAC CCG CAG TTA GAG CGT GAT ACG TTG GCT 5616 Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala 1860 1865 1870		
GAA GCT AAA ATG TGG TAT ACA CAG GCG CTT AAT CTG TTG GGT GAT CAG 5664 Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu 1875 1880 1885		
CCA CAA GTG ATG CTG AGT ACG ACT TGG GCT AAT CCA ACA TTG GGT AAT 5712 Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn 1890 1895 1900		
GCT GCT TCA AAA ACC ACA CAG CAG GTT GGT CAG CAA GTG CTT ACC CAG 5760 Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln 1905 1910 1915 1920		

TTG GGT TTC AAT ACC AGG GTA AAA ACC GCG TTG CTA GGA AGA GCG AAT 5408
 Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn
 1925 1930 1935

TCC CTC ACC GGT TTA TTC CTG CCG CAG GAA AAT ACC AAG CTC AAA GCG 5356
 Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly
 1940 1945 1950

TAC TGG CCG ACA CTG CCG CAG CGT ATG TTT AAT TTA CGT CAT AAT CTC 5904
 Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu
 1955 1960 1965

TCG ATT GAC GCG CAG CCG CTC TCC TTG CCG CTG TAT GCT AAA CCG GCT 5352
 Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala
 1970 1975 1980

GAT CCA AAA GGT TTA CTC AGT GCG GCG GTT TCA GCT TCT CAA CCG GGA 6000
 Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly
 1985 1990 1995 2000

GCG GAC TTG CCG AAG CCG CCG CTC ACT ATT CAC CCG TTC CCT CAA ATG 6048
 Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met
 2005 2010 2015

CTA GAA GCG GCA CCG GCG TTG GTT AAC CAG CTT ATA CAG TTC GGT AGT 6096
 Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser
 2020 2025 2030

TCA CTA TTG CCG TAC AGT GAG CGT CAG GAT GCG GAA GGT ATG AGT CAA 6144
 Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln
 2035 2040 2045

CTA CTG CAA ACC CAA GCG AGC GAG TTA ATA CTG ACC AGT ATT CGT ATG 6192
 Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met
 2050 2055 2060

CAG GAT AAC CAA TTG GCA GAG CTG GAT TCG GAA AAA ACC GCG TTG CAA 6240
 Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln
 2065 2070 2075 2080

CTC TCT TTA GCT GGA GTG CAA CAA CCG TTT GAC AGC TAT AGC CAA CTG 6288
 Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu
 2085 2090 2095

TAT GAG GAG AAC ATC AAC GCA GGT GAG CAG CGA CCG CTC GCG TTA CCG 6336
 Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg
 2100 2105 2110

TCA GAA TCT GCT ATT GAG TCT CAG GGA GCG CAG ATT TCC GGT ATG GCA 6384
 Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala
 2115 2120 2125

GCG GCG GGT GTT GAT ATG GCA CCA AAT ATC TTC GCG CTC GCT GAT GCG 6432
 Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly
 2130 2135 2140

GCG ATG CAT TAT GGT GCT ATT GCC TAT GCC ATC GCT GAC GGT ATT GAG 6480
 Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu
 2145 2150 2155 2160

TTG AGT GCT TCT GCG AAG ATG GTT GAT GCG CAG AAA GTT GCT CAG TCG 6528
 Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln Ser
 2165 2170 2175

GAA ATA TAT CCG GGT CCG GGT CAA GAA TGG AAA ATT CAG CGT GAC AAC 6576
 Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

2180 GCA CAA GCG GAG ATT AAC GAG TTA AAC GCG CAA CTG GAA TCA CTG TCT 6624 Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser 2195 2200 2205 ATT GCG CGT GAA GCG GCT GAA ATG CAA AAA GAG TAC CTG AAA ACC GAG 6672 Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln 2210 2215 2220 CAA GCT CAG GCG CAG GCA CAA CTT ACT TTC TTA AGA AGC AAA TTC AGT 6720 Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser 2225 2230 2235 2240 AAT CAA GCG TTA TAT AGT TCG TTA CGA GCG CGT TTG TCA GGT ATT TAT 6768 Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr 2245 2250 2255 TTC CAG TTC TAT GAC TTG GCC GTA TCA CGT TGC CTG ATG GCA GAG CAA 6816 Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln 2260 2265 2270 TCC TAT CAA TCG GAA GCT AAT GAT AAT TCC ATT AGC TTT CTC AAA CCG 6864 Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro 2275 2280 2285 GGT GCA TCG CAA GGA ACT TAC GCC GCG TTA TTG TGT GGA GAA GCT TTG 6912 Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala Leu 2290 2295 2300 ATA CAA AAT CTG GCA CAA ATG GAA GAG GCA TAT CTG AAA TGG GAA TCT 6960 Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser 2305 2310 2315 2320 CGC GCT TTC GAA GTA GAA CGC ACG GTT TCA TTG GCA GTG GTT TAT GAT 7008 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp 2325 2330 2335 TCA CTG GAA GGT AAT GAT CGT TTT AAT TTA GCG GAA CAA ATA CCT GCA 7056 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala 2340 2345 2350 TTA TTG CAT AAG GCG GAG GGA ACA GCA GGA ACT AAA GAA AAT GCG TTA 7104 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu 2355 2360 2365 TCA TTG GCT AAT GCT ATC CTG TCA GCT TCG GTC AAA TTG TCC GAC TTG 7152 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu 2370 2375 2380 AAA CTG GGA ACG GAT TAT CCA GAC AGT ATC GTT GGT AGC AAC AAG GTT 7200 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val 2385 2390 2395 2400 CGT CGT ATT AAG CAA ATC AGT GTT TCG CTA CCT GCA TTG GTT GCG CCT 7248 Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro 2405 2410 2415 TAT CAG GAT GTT CAG GCT ATG CTC ACC TAT GGT GCG AGT ACT CAA TTG 7296 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu 2420 2425 2430 CGC AAA GGT TGT TCA GCG TTG GCT GTC TCT CAT GGT ACC AAT GAT AGT 7344 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser 2435 2440 2445	2135	2190
--	-------------	-------------

GGT CAG TTC CAG TTG GAT TTC AAT CAC GGC AAA TAC CTG CCA TTT GAA 7392
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460

GGT ATT GCT CTT GAT GAT CAG GGT ACA CTG AAT CTT CAA TTT CCG AAT 7440
 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480

GCT ACC GAC AAG CAG AAA GCA ATA TTG CAA ACT ATG ACC GAT ATT ATT 7488
 Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495

TTG CAT ATT CGT TAT ACC ATC CGT TAA 7515
 Leu His Ile Arg Tyr Thr Ile Arg
 2500 2505

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 12:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 2505 амінокислот
 (B) ТИП: амінокислота
 (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 12:

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

U A 8 2 4 8 5 C 2

Leu Ser Gly Glu Thr Pro Tyr His His Ala Tyr Glu Thr Val Arg Glu	195	200	205
Ile Val His Glu Arg Asp Pro Gly Phe Arg His Leu Ser Gln Ala Pro	210	215	220
Ile Val Ala Ala Lys Leu Asp Pro Val Thr Leu Leu Gly Ile Ser Ser	225	230	235 240
His Ile Ser Pro Glu Leu Tyr Asn Leu Leu Ile Glu Glu Ile Pro Glu	245	250	255
Lys Asp Glu Ala Ala Leu Asp Thr Leu Tyr Lys Thr Asn Phe Gly Asp	260	265	270
Ile Thr Thr Ala Gln Leu Met Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr	275	280	285
Gly Val Ser Pro Glu Asp Ile Ala Tyr Val Thr Thr Ser Leu Ser His	290	295	300
Val Gly Tyr Ser Ser Asp Ile Leu Val Ile Pro Leu Val Asp Gly Val	305	310	315 320
Gly Lys Met Glu Val Val Arg Val Thr Arg Thr Pro Ser Asp Asn Tyr	325	330	335
Thr Ser Gln Thr Asn Tyr Ile Glu Leu Tyr Pro Gln Gly Gly Asp Asn	340	345	350
Tyr Leu Ile Lys Tyr Asn Leu Ser Asn Ser Phe Gly Leu Asp Asp Phe	355	360	365
Tyr Leu Gln Tyr Lys Asp Gly Ser Ala Asp Trp Thr Glu Ile Ala His	370	375	380
Asn Pro Tyr Pro Asp Met Val Ile Asn Gln Lys Tyr Glu Ser Gln Ala	385	390	395 400
Thr Ile Lys Arg Ser Asp Ser Asp Asn Ile Leu Ser Ile Gly Leu Gln	405	410	415
Arg Trp His Ser Gly Ser Tyr Asn Phe Ala Ala Ala Asn Phe Lys Ile	420	425	430
Asp Gln Tyr Ser Pro Lys Ala Phe Leu Leu Lys Met Asn Lys Ala Ile	435	440	445
Arg Leu Leu Lys Ala Thr Gly Leu Ser Phe Ala Thr Leu Glu Arg Ile	450	455	460
Val Asp Ser Val Asn Ser Thr Lys Ser Ile Thr Val Glu Val Leu Asn	465	470	475 480
Lys Val Tyr Arg Val Lys Phe Tyr Ile Asp Arg Tyr Gly Ile Ser Glu	485	490	495
Glu Thr Ala Ala Ile Leu Ala Asn Ile Asn Ile Ser Gln Gln Ala Val	500	505	510
Gly Asn Gln Leu Ser Gln Phe Glu Gln Leu Phe Asn His Pro Pro Leu	515	520	525
Asn Gly Ile Arg Tyr Glu Ile Ser Glu Asp Asn Ser Lys His Leu Pro	530	535	540

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Asn	Pro	Asp	Leu	Asn	Leu	Lys	Pro	Asp	Ser	Thr	Gly	Asp	Asp	Gln	Arg	545	550	555	560
Lys	Ala	Val	Leu	Lys	Arg	Ala	Phe	Gln	Val	Asn	Ala	Ser	Glu	Leu	Tyr	565	570	575	
Gln	Met	Leu	Leu	Ile	Thr	Asp	Arg	Lys	Glu	Asp	Gly	Val	Ile	Lys	Asn	580	585	590	
Asn	Leu	Glu	Asn	Leu	Ser	Asp	Leu	Tyr	Leu	Val	Ser	Leu	Leu	Ala	Gln	595	600	605	
Ile	His	Asn	Leu	Thr	Ile	Ala	Glu	Leu	Asn	Ile	Leu	Leu	Val	Ile	Cys	610	615	620	
Gly	Tyr	Gly	Asp	Thr	Asn	Ile	Tyr	Gln	Ile	Thr	Asp	Asp	Asn	Leu	Ala	625	630	635	640
Lys	Ile	Val	Glu	Thr	Leu	Leu	Trp	Ile	Thr	Gln	Trp	Leu	Lys	Thr	Gln	645	650	655	
Lys	Trp	Thr	Val	Thr	Asp	Leu	Phe	Leu	Met	Thr	Thr	Ala	Thr	Tyr	Ser	660	665	670	
Thr	Thr	Leu	Thr	Pro	Glu	Ile	Ser	Asn	Leu	Thr	Ala	Thr	Leu	Ser	Ser	675	680	685	
Thr	Leu	His	Gly	Lys	Glu	Ser	Leu	Ile	Gly	Glu	Asp	Leu	Lys	Arg	Ala	690	695	700	
Met	Ala	Pro	Cys	Phe	Thr	Ser	Ala	Leu	His	Leu	Thr	Ser	Gln	Glu	Val	705	710	715	720
Ala	Tyr	Asp	Leu	Leu	Leu	Trp	Ile	Asp	Gln	Ile	Gln	Pro	Ala	Gln	Ile	725	730	735	
Thr	Val	Asp	Gly	Phe	Trp	Glu	Glu	Val	Gln	Thr	Thr	Pro	Thr	Ser	Leu	740	745	750	
Lys	Val	Ile	Thr	Phe	Ala	Gln	Val	Leu	Ala	Gln	Leu	Ser	Leu	Ile	Tyr	755	760	765	
Arg	Arg	Ile	Gly	Leu	Ser	Glu	Thr	Glu	Leu	Ser	Leu	Ile	Val	Thr	Gln	770	775	780	
Ser	Ser	Leu	Leu	Val	Ala	Gly	Lys	Ser	Ile	Leu	Asp	His	Gly	Leu	Leu	785	790	795	800
Thr	Leu	Met	Ala	Leu	Glu	Gly	Phe	His	Thr	Trp	Val	Asn	Gly	Leu	Gly	805	810	815	
Gln	His	Ala	Ser	Leu	Ile	Leu	Ala	Ala	Leu	Lys	Asp	Gly	Ala	Leu	Thr	820	825	830	
Val	Thr	Asp	Val	Ala	Gln	Ala	Met	Asn	Lys	Glu	Glu	Ser	Leu	Leu	Gln	835	840	845	
Met	Ala	Ala	Asn	Gln	Val	Glu	Lys	Asp	Leu	Thr	Lys	Leu	Thr	Ser	Trp	850	855	860	
Thr	Gln	Ile	Asp	Ala	Ile	Leu	Gln	Trp	Leu	Gln	Met	Ser	Ser	Ala	Leu	865	870	875	880
Ala	Val	Ser	Pro	Leu	Asp	Leu	Ala	Gly	Met	Met	Ala	Leu	Lys	Tyr	Gly	885	890	895	

U A 8 2 4 8 5 C 2

Ile Asp His Asn Tyr Ala Ala Trp Gln Ala Ala Ala Ala Ala Leu Met
 900 905 910
 Ala Asp His Ala Asn Gln Ala Gln Lys Lys Leu Asp Glu Thr Phe Ser
 915 920 925
 Lys Ala Leu Cys Asn Tyr Tyr Ile Asn Ala Val Val Asp Ser Ala Ala
 930 935 940
 Gly Val Arg Asp Arg Asn Gly Leu Tyr Thr Tyr Leu Leu Ile Asp Asn
 945 950 955 960
 Gln Val Ser Ala Asp Val Ile Thr Ser Arg Ile Ala Glu Ala Ile Ala
 965 970 975
 Gly Ile Gln Leu Tyr Val Asn Arg Ala Leu Asn Arg Asp Glu Gly Gln
 980 985 990
 Leu Ala Ser Asp Val Ser Thr Arg Gln Phe Phe Thr Asp Trp Glu Arg
 995 1000 1005
 Tyr Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Glu Leu Val Tyr
 1010 1015 1020
 Tyr Pro Glu Asn Tyr Val Asp Pro Thr Gln Arg Ile Gly Gln Thr Lys
 1025 1030 1035 1040
 Met Met Asp Ala Leu Leu Gln Ser Ile Asn Gln Ser Gln Leu Asn Ala
 1045 1050 1055
 Asp Thr Val Glu Asp Ala Phe Lys Thr Tyr Leu Thr Ser Phe Glu Gln
 1060 1065 1070
 Val Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Val Asn Val
 1075 1080 1085
 Asp Gln Gly Leu Thr Tyr Phe Ile Gly Ile Asp Gln Ala Ala Pro Gly
 1090 1095 1100
 Thr Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Cys Glu Asn Gly Lys
 1105 1110 1115 1120
 Phe Ala Ala Asn Ala Trp Gly Glu Trp Asn Lys Ile Thr Cys Ala Val
 1125 1130 1135
 Asn Pro Trp Lys Asn Ile Ile Arg Pro Val Val Tyr Met Ser Arg Leu
 1140 1145 1150
 Tyr Leu Leu Trp Leu Glu Gln Gln Ser Lys Lys Ser Asp Asp Gly Lys
 1155 1160 1165
 Thr Thr Ile Tyr Gln Tyr Asn Leu Lys Leu Ala His Ile Arg Tyr Asp
 1170 1175 1180
 Gly Ser Trp Asn Thr Pro Phe Thr Phe Asp Val Thr Glu Lys Val Lys
 1185 1190 1195 1200
 Asn Tyr Thr Ser Ser Thr Asp Ala Ala Glu Ser Leu Gly Leu Tyr Cys
 1205 1210 1215
 Thr Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Ser Met
 1220 1225 1230
 Gln Ser Ser Tyr Ser Ser Tyr Thr Asp Asn Asn Ala Pro Val Thr Gly
 1235 1240 1245

U A 8 2 4 8 5 C 2

Leu Tyr Ile Phe Ala Asp Met Ser Ser Asp Asn Met Thr Asn Ala Gln
1250 1255 1260

Ala Thr Asn Tyr Trp Asn Asn Ser Tyr Pro Gln Phe Asp Thr Val Met
1265 1270 1275 1280

Ala Asp Pro Asp Ser Asp Asn Lys Lys Val Ile Thr Arg Arg Val Asn
1285 1290 1295

Asn Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Thr Ser Asn
1300 1305 1310

Ser Asn Tyr Ser Trp Gly Asp His Ser Leu Thr Met Leu Tyr Gly Gly
1315 1320 1325

Ser Val Pro Asn Ile Thr Phe Glu Ser Ala Ala Glu Asp Leu Arg Leu
1330 1335 1340

Ser Thr Asn Met Ala Leu Ser Ile Ile His Asn Gly Tyr Ala Gly Thr
1345 1350 1355 1360

Arg Arg Ile Gln Cys Asn Leu Met Lys Gln Tyr Ala Ser Leu Gly Asp
1365 1370 1375

Lys Phe Ile Ile Tyr Asp Ser Ser Phe Asp Asp Ala Asn Arg Phe Asn
1380 1385 1390

Leu Val Pro Leu Phe Lys Phe Gly Lys Asp Glu Asn Ser Asp Asp Ser
1395 1400 1405

Ile Cys Ile Tyr Asn Glu Asn Pro Ser Ser Glu Asp Lys Lys Trp Tyr
1410 1415 1420

Phe Ser Ser Lys Asp Asp Asn Lys Thr Ala Asp Tyr Asn Gly Gly Thr
1425 1430 1435 1440

Gln Cys Ile Asp Ala Gly Thr Ser Asn Lys Asp Phe Tyr Tyr Asn Leu
1445 1450 1455

Gln Glu Ile Glu Val Ile Ser Val Thr Gly Gly Tyr Trp Ser Ser Tyr
1460 1465 1470

Lys Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
1475 1480 1485

Val Lys Val Thr Val Lys Ala Gly Gly Asp Asp Gln Ile Phe Thr Ala
1490 1495 1500

Asp Asn Ser Thr Tyr Val Pro Gln Gln Pro Ala Pro Ser Phe Glu Glu
1505 1510 1515 1520

Met Ile Tyr Gln Phe Asn Asn Leu Thr Ile Asp Cys Lys Asn Leu Asn
1525 1530 1535

Phe Ile Asp Asn Gln Ala His Ile Glu Ile Asp Phe Thr Ala Thr Ala
1540 1545 1550

Gln Asp Gly Arg Phe Leu Gly Ala Glu Thr Phe Ile Ile Pro Val Thr
1555 1560 1565

Lys Lys Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn
1570 1575 1580

Asn Gly Val Gln Tyr Met Gln Ile Gly Ala Tyr Arg Thr Arg Leu Asn
1585 1590 1595 1600

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Thr Leu Phe Ala Gln Gln Leu Val Ser Arg Ala Asn Arg Gly Ile Asp
 1605 1610 1615
 Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly
 1620 1625 1630
 Ala Gly Thr Tyr Val Gln Leu Val Leu Asp Lys Tyr Asp Glu Ser Ile
 1635 1640 1645
 His Gly Thr Asn Lys Ser Phe Ala Ile Glu Tyr Val Asp Ile Phe Lys
 1650 1655 1660
 Glu Asn Asp Ser Phe Val Ile Tyr Gln Gly Glu Leu Ser Glu Thr Ser
 1665 1670 1675 1680
 Gln Thr Val Val Lys Val Phe Leu Ser Tyr Phe Ile Glu Ala Thr Gly
 1685 1690 1695
 Asn Lys Asn His Leu Trp Val Arg Ala Lys Tyr Gln Lys Glu Thr Thr
 1700 1705 1710
 Asp Lys Ile Leu Phe Asp Arg Thr Asp Glu Lys Asp Pro His Gly Trp
 1715 1720 1725
 Phe Leu Ser Asp Asp His Lys Thr Phe Ser Gly Leu Ser Ser Ala Gln
 1730 1735 1740
 Ala Leu Lys Asn Asp Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ala
 1745 1750 1755 1760
 Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Met Met Ala His
 1765 1770 1775
 Arg Leu Leu Gln Glu Gln Asn Phe Asp Ala Ala Asn His Trp Phe Arg
 1780 1785 1790
 Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Asp Gly Lys Ile Ala Ile
 1795 1800 1805
 Tyr His Trp Asn Val Arg Pro Leu Glu Glu Asp Thr Ser Trp Asn Ala
 1810 1815 1820
 Gln Gln Leu Asp Ser Thr Asp Pro Asp Ala Val Ala Gln Asp Asp Pro
 1825 1830 1835 1840
 Met His Tyr Lys Val Ala Thr Phe Met Ala Thr Leu Asp Leu Leu Met
 1845 1850 1855
 Ala Arg Gly Asp Ala Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Ala
 1860 1865 1870
 Glu Ala Lys Met Trp Tyr Thr Gln Ala Leu Asn Leu Leu Gly Asp Glu
 1875 1880 1885
 Pro Gln Val Met Leu Ser Thr Thr Trp Ala Asn Pro Thr Leu Gly Asn
 1890 1895 1900
 Ala Ala Ser Lys Thr Thr Gln Gln Val Arg Gln Gln Val Leu Thr Gln
 1905 1910 1915 1920
 Leu Arg Leu Asn Ser Arg Val Lys Thr Pro Leu Leu Gly Thr Ala Asn
 1925 1930 1935
 Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn Ser Lys Leu Lys Gly
 1940 1945 1950

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn Leu Arg His Asn Leu
1955 1960 1965

5 Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu Tyr Ala Lys Pro Ala
1970 1975 1980

Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser Ala Ser Gln Gly Gly
1985 1990 1995 2000

1 Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His Arg Phe Pro Gln Met
2005 2010 2015

Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu Ile Gln Phe Gly Ser
2020 2025 2030

Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala Glu Ala Met Ser Gln
2035 2040 2045

Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu Thr Ser Ile Arg Met
2050 2055 2060

Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu Lys Thr Ala Leu Gln
2065 2070 2075 2080

Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp Ser Tyr Ser Gln Leu
2085 2090 2095

Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg Ala Leu Ala Leu Arg
2100 2105 2110

Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln Ile Ser Arg Met Ala
2115 2120 2125

Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe Gly Leu Ala Asp Gly
2130 2135 2140

Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile Ala Asp Gly Ile Glu
2145 2150 2155 2160

Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu Lys Val Ala Gln Ser
2165 2170 2175

Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys Ile Gln Arg Asp Asn
2180 2185 2190

Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln Leu Glu Ser Leu Ser
2195 2200 2205

Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu Tyr Leu Lys Thr Gln
2210 2215 2220

Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu Arg Ser Lys Phe Ser
2225 2230 2235 2240

Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg Leu Ser Gly Ile Tyr
2245 2250 2255

Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys Leu Met Ala Glu Gln
2260 2265 2270

Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile Ser Phe Val Lys Pro
2275 2280 2285

Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu Cys Gly Glu Ala Leu
2290 2295 2300

U A 8 2 4 8 5 C 2

Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr Leu Lys Trp Glu Ser
 2305 2310 2315 2320
 Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu Ala Val Val Tyr Asp
 2325 2330 2335
 Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala Glu Gln Ile Pro Ala
 2340 2345 2350
 Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr Lys Glu Asn Gly Leu
 2355 2360 2365
 Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val Lys Leu Ser Asp Leu
 2370 2375 2380
 Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val Gly Ser Asn Lys Val
 2385 2390 2395 2400
 Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro Ala Leu Val Gly Pro
 2405 2410 2415
 Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly Gly Ser Thr Gln Leu
 2420 2425 2430
 Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser
 2435 2440 2445
 Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Tyr Leu Pro Phe Glu
 2450 2455 2460
 Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn Leu Gln Phe Pro Asn
 2465 2470 2475 2480
 Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr Met Ser Asp Ile Ile
 2485 2490 2495
 Leu His Ile Arg Tyr Thr Ile Arg
 2500 2505

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 13:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 12 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 13:

Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Xaa Ala
 1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 14:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 12 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 14:

Gln Asn Ser Gln Thr Phe Ser Val Gly Glu Leu
5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 15:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 9 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 15:

Ala Gln Asp Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 16:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 5 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 16:

Met Gln Asn Ser Leu
1 5

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 17:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 10 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 17:

Ala Phe Asn Ile Asp Asp Val Ser Leu Phe
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 18:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 16 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 18:

Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 19:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 21 амінокислота
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 19:

Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile Gly Ser
1 5 10 15

Leu Gln Leu Phe Ile
20

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 20:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 12 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 20:

Met Tyr Tyr Ile Gln Ala Gln Gln Leu Gly Pro
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 21:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 26 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ:
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 21:

Gly Ile Asp Ala Val Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro
1 5 10 15
Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
20 25

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 22:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 15 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 22:

Ile Ser Asn Pro Ile Asn Ile Asn Thr Gly Ile Asp Ser Ala Lys
1 5 10 15

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 23:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: амінокислота
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 23:

Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 24:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 22 амінокислоти
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: пептид

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 24:

Val Leu Gly Thr Glu Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly

1 5 10 15
Val Gln Tyr Met Gln Ile
20

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 25:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 6005 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: RBS
- (B) МІСЦЕПОЛОЖЕННЯ: 1...9

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: CDS
- (B) МІСЦЕПОЛОЖЕННЯ: 16...3585
- (D) ІНША ІНФОРМАЦІЯ: /продукт="P8"

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 25:

AAGAAGGAAT TCATT ATG TCT GAA TCT TTA TTT ACA CAA ACG TTG AAA GAA	51
Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu	
1 5 10	
CCC CCC CGT GAT GCA TTG GTT GCT CAT TAT ATT GCT ACT CAG GTG CCC	99
Ala Arg Arg Asp Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro	
15 20 25	
GCA GAT TTA AAA GAG AGT ATC CAG ACC GCG GAT GAT CTG TAC GAA TAT	147
Ala Asp Leu Lys Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr	
30 35 40	
CTG TTG CTG GAT ACC AAA ATT AGC GAT CTG GTT ACT ACT TCA CCG CTG	195
Leu Leu Leu Asp Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu	
45 50 55 60	
TCC GAA GCG ATT GCG AGT CTG CAA TTG TTT ATT CAT CGT GCG ATA GAG	243
Ser Glu Ala Ile Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu	
65 70 75	
GCC TAT GAC GGC ACG CTG GCA GAC TCA GCA AAA CCC TAT TTT GCC GAT	291
Gly Tyr Asp Gly Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp	
80 85 90	
GAA CAG TTT TTA TAT AAC TCG GAT AGT TTT AAC CAC CGT TAT AGC ACT	339
Glu Gln Phe Leu Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr	
95 100 105	
TGG GCT GGC AAC GAA CCG TTG AAA TTC TAT GCC GCG GAT TAT ATT GAT	387
Trp Ala Gly Lys Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp	
110 115 120	
CCA ACA TTG CGA TTG AAT AAG ACC CAG ATA TTT ACC GCA TTT GAA CAA	435
Pro Thr Leu Arg Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln	
125 130 135 140	

GGT ATT TCT CAA GGG AAA TTA AAA AGT GAA TTA CTC GAA TCT AAA TTA 433
Gly Ile Ser Gln Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu
145 150 155

CCT GAT TAT CTA ATT AGT TAT GAC ACT TTA GCC ACC GTT GAT TAT ATT 531
Arg Asp Tyr Leu Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile
160 165 170

ACT GCC TGC CAA GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT 579
Thr Ala Cys Gln Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg
175 180 185

ACA CAG AAT GCA CCC TAT GCA TTT TAT TCG CGA AAA TTA ACT TTA GTC 627
Thr Gln Asn Ala Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val
190 195 200

ACT GAT GGC GGT AAG TTG AAA CCA GAT CAA TCG TCA CAG TCG CGA GCA 675
Thr Asp Gly Gly Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala
205 210 215 220

ATT AAT GCC GGG ATT AGT CAG GCA TAT TCA GCG CAT GTC GAG CCT TTC 723
Ile Asn Ala Gly Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe
225 230 235

TCG GAA AAT AAC AAG CTG CAC ATC CGT TGG TTT ACT ATC TCG AAA GAA 771
Trp Glu Asn Asn Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu
240 245 250

GAT AAA ATA GAT TTT GTT TAT AAA AAC ATC TCG CTC ATG AGT AGC GAT 819
Asp Lys Ile Asp Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp
255 260 265

TAT AGC TCG GCA TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC 867
Tyr Ser Trp Ala Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp
270 275 280

TAC AAT AGA GTT GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT 915
Tyr Asn Arg Val Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala
285 290 295 300

TCA CAA TAT GGT TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGC ACT 963
Ser Gln Tyr Gly Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr
305 310 315

GTA CTT ATT TTT CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG 1011
Val Leu Ile Phe Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val
320 325 330

ACG TTA TGT TAT GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA 1059
Thr Leu Cys Tyr Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr
335 340 345

GGA AGT GCA AAT TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA GCC 1107
Gly Ser Ala Asn Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg
350 355 360

ATG TGT CAT GGA CAA AGT TAC AAT GAT AAT AAC TAC TCG AAT TTT ACA 1155
Met Cys His Gly Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr
365 370 375 380

CTC TCT ATT AAT ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA 1203
Leu Ser Ile Asn Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser
385 390 395

GAT GGA AAA CAA TTT ACA CCA CCT TCT GGT TCT GCC ATT GAT TTA CAC 1251
Asp Gly Lys Gln Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His

U A 8 2 4 8 5 C 2

400	425	410	
CTC CTT AAT TAT GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT 1299			
Leu Pro Asn Tyr Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp			
415	420	425	
TCA CTA CTT AAT TAT GAC GTT CAG GCG CAG TTT GCG GGA TCT AAT CCG 1347			
Ser Leu Leu Asn Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro			
430	435	440	
GTT GAT AAT TTC AGT GGT CCC TAT GGT ATT TAT CTA TGG GAA ATC TTC 1395			
Val Asp Asn Phe Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe			
445	450	455	460
TTC CAT ATT CCG TTC CTT GTT ACG GTC CGT ATG CAA ACC GAA CAA CGT 1443			
Phe His Ile Pro Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg			
465	470	475	
TAC GAA GAC GCG GAC ACT TGG TAC AAA TAT ATT TTC CCG AGC GCC GGT 1491			
Tyr Glu Asp Ala Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly			
480	485	490	
TAT CCG GAT GCT AAT GGC CAG CTC ATT ATG GAT GCG AGT AAA CCA CGT 1539			
Tyr Arg Asp Ala Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg			
495	500	505	
TAT TGG AAT GTG ATG CCA TTG CAA CTC GAT ACC GCA TGG GAT ACC ACA 1587			
Tyr Trp Asn Val Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr			
510	515	520	
CAG CCC GCC ACC ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG 1635			
Gln Pro Ala Thr Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met			
525	530	535	540
CAT TAC AAG CTG GCG ATA TTC CTG CAT ACC CTT GAT CTA TTG ATT GCC 1683			
His Tyr Lys Leu Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala			
545	550	555	
CGA GCG GAC ACC GCT TAC CGT CAA CTT GAA GCG GAT ACT CTA GTC GAA 1731			
Arg Gly Asp Ser Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu			
560	565	570	
GCC AAA ATG TAC TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CGC CCT 1779			
Ala Lys Met Tyr Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro			
575	580	585	
GAT ATC CAT ACC ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA 1827			
Asp Ile His Thr Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu			
590	595	600	
GCT GCG GCT ATT GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG 1875			
Ala Gly Ala Ile Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met			
605	610	615	620
ACG TTC GCT GCC TGG CTA AGC GCA GCG GAT ACC GCA AAT ATT GGC GAC 1923			
Thr Phe Ala Ala Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp			
625	630	635	
GGT GAT TTC TTG CCA CCG TAC AAC GAT GTA CTA CTC GGT TAC TGG GAT 1971			
Gly Asp Phe Leu Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp			
640	645	650	
AAA CTT GAG TTA CCG CTA TAC AAC CTG CCG CAC AAT CTG AGT CTG GAT 2019			
Lys Leu Glu Leu Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp			
655	660	665	

U A 8 2 4 8 5 C 2

GGT CAA CCG CTA AAT CTG CCA CTG TAT GCC ACG CCG GTA GAC CCG AAA 2557
 Gly Gln Pro Leu Asn Leu Pro Leu Tyr Ala Thr Pro Val Asp Pro Lys
 670 675 680

ACC CTG CAA CCG CAG CAA GCC GGA GGG GAC GGT ACA GCC AGT AGT CCG 2115
 Thr Leu Gln Arg Gln Gln Ala Gly Gly Asp Gly Thr Gly Ser Ser Pro
 685 690 695 700

GGT GGT GGT CAA GCC AGT GTT CAG GGC TGG CGC TAT CCG TTA TTG GTA 2153
 Ala Gly Gly Gln Gly Ser Val Gln Gly Trp Arg Tyr Pro Leu Leu Val
 705 710 715

GAA CCG GCC CCG TCT GCC CTG AGT TTG TTG ACT CAG TTC GGC AAC AGC 2211
 Glu Arg Ala Arg Ser Ala Val Ser Leu Leu Thr Gln Phe Gly Asn Ser
 720 725 730

TTA CAA ACA ACG TTA GAA CAT CAG GAT AAT GAA AAA ATG ACG ATA CTG 2259
 Leu Gln Thr Thr Leu Glu His Gln Asp Asn Glu Lys Met Thr Ile Leu
 735 740 745

TTG CAG ACT CAA CAG GAA GCC ATC CTG AAA CAT CAG CAC GAT ATA CAA 2307
 Leu Gln Thr Gln Gln Glu Ala Ile Leu Lys His Gln His Asp Ile Gln
 750 755 760

CAA AAT AAT CTA AAA GGA TTA CAA CAC AGC CTG ACC GCA TTA CAG GCT 2355
 Gln Asn Asn Leu Lys Gly Leu Gln His Ser Leu Thr Ala Leu Gln Ala
 765 770 775 780

AGC CGT GAT GCC GAC ACA TTG CCG CAA AAA CAT TAC AGC GAC CTG ATT 2403
 Ser Arg Asp Gly Asp Thr Leu Arg Gln Lys His Tyr Ser Asp Leu Ile
 785 790 795

AAC GGT GGT CTA TCT CCG GCA GAA ATC GCC GGT CTG ACA CTA CCG AGC 2451
 Asn Gly Gly Leu Ser Ala Ala Glu Ile Ala Gly Leu Thr Leu Arg Ser
 800 805 810

ACC GCC ATG ATT ACC AAT GCC GTT GCA ACG GGA TTG CTG ATT GCC GCG 2499
 Thr Ala Met Ile Thr Asn Gly Val Ala Thr Gly Leu Leu Ile Ala Gly
 815 820 825

GGA ATC GCC AAC CCG GTA CCT AAC GTC TTC GGG CTG GCT AAC GGT GGA 2547
 Gly Ile Ala Asn Ala Val Pro Asn Val Phe Gly Leu Ala Asn Gly Gly
 830 835 840

TGG GAA TGG GGA GCG CCA TTA ATT GGC TCC GGG CAA GCA ACC CAA GTT 2595
 Ser Glu Trp Gly Ala Pro Leu Ile Gly Ser Gly Gln Ala Thr Gln Val
 845 850 855 860

GGC GCC GCG ATC CAG GAT CAG AGC GCG GGC ATT TCA GAA GTG ACA GCA 2643
 Gly Ala Gly Ile Gln Asp Gln Ser Ala Gly Ile Ser Glu Val Thr Ala
 865 870 875

GGC TAT CAG CGT CGT CAG GAA GAA TGG GCA TTG CAA CCG GAT ATT GCT 2691
 Gly Tyr Gln Arg Arg Gln Glu Glu Trp Ala Leu Gln Arg Asp Ile Ala
 880 885 890

GAT AAC GAA ATA ACC CAA CTG GAT GCC CAG ATA CAA AGC CTG CAA GAG 2739
 Asp Asn Glu Ile Thr Gln Leu Asp Ala Gln Ile Gln Ser Leu Gln Glu
 895 900 905

CAA ATC ACC ATG GCA CAA AAA CAG ATC ACG CTC TCT GAA ACC GAA CAA 2787
 Gln Ile Thr Met Ala Gln Lys Gln Ile Thr Leu Ser Glu Thr Glu Gln
 910 915 920

GCG AAT GCC CAA GCG ATT TAT GAC CTG CAA ACC ACT CGT TTT ACC GCG 2835
 Ala Asn Ala Gln Ala Ile Tyr Asp Leu Gln Thr Thr Arg Phe Thr Gly

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

925	930	935	940
CAG GCA CTG TAT AAC TCG ATG GCC GGT CGT CTC TCC GCG CTC TAT TAC 2333			
Gln Ala Leu Tyr Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr	945	950	955
CAA ATG TAT GAT TCC ACT CTG CCA ATC TGT CTC CAG CCA AAA GCC GCA 2331			
Gln Met Tyr Asp Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala	960	965	970
TTA GTA CAG GAA TTA GCC GAG AAA GAG AGC GAC AGT CTT TTC CAG GTT 2379			
Leu Val Gln Glu Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val	975	980	985
CCG GTG TCG AAT GAT CTG TCG CAA GCG CTG TTA GCA GCA GAA GGT TTA 3027			
Pro Val Trp Asn Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Le.	990	995	1000
AGT TCA GAG CTA CAG AAA CTG GAT GCC ATC TCG CTT GCA CGT GGT GGT 3075			
Ser Ser Glu Leu Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly	1005	1010	1015
ATT GGG CTA GAA GCC ATC CGC ACC GTG TCG CTG GAT ACC CTG TTT GCC 3123			
Ile Gly Leu Glu Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly	1025	1030	1035
ACA GGG ACG TTA AGT GAA AAT ATC AAT AAA GTG CTT AAC GCG GAA ACG 3171			
Thr Gly Thr Leu Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr	1040	1045	1050
GTA TCT CCA TCC GGT GGC GTC ACT CTG GCG CTG ACA GCG GAT ATC TTC 3219			
Val Ser Pro Ser Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe	1055	1060	1065
CAA GCA ACA CTG GAT TTG AGT CAG CTA GGT TTG GAT AAC TCT TAC AAC 3267			
Gln Ala Thr Leu Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn	1070	1075	1080
TTG GGT AAC GAG AAG AAA CGT CGT ATT AAA CGT ATC GCC GTC ACC CTG 3315			
Leu Gly Asn Glu Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu	1085	1090	1095
CCA ACA CTT CTG GCG CCA TAT CAA GAT CTT GAA GCC ACA CTG GTA ATG 3363			
Pro Thr Leu Leu Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met	1105	1110	1115
GGT GCG GAA ATC GCC GCC TTA TCA CAC GGT GTG AAT GAC CGA GCC CGG 3411			
Gly Ala Glu Ile Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg	1120	1125	1130
TTT GTT ACC GAC TTT AAC GAC AGC CGT TTT CTG CCT TTT GAA GGT CGA 3459			
Phe Val Thr Asp Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg	1135	1140	1145
GAT GCA ACA ACC GCG ACA CTG GAG CTC AAT ATT TTC CAT GCG GGT AAA 3507			
Asp Ala Thr Thr Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys	1150	1155	1160
GAG GGA ACG CAA CAC GAG TTG CTC GCG AAT CTG AGT GAC ATC ATT GTG 3555			
Glu Gly Thr Gln His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val	1165	1170	1175
CAT CTG AAT TAC ATC ATT CGA GAC GCG TAA ATTTCTTTTC TTGTGCGATT 3605			
His Leu Asn Tyr Ile Ile Arg Asp Ala	1185	1190	

U
A
8
2
4
8
5
C
2

ACAGGTCCCT ATCAGGGGCG TGTATTAAAG GAGTACTTTA TGCAGGATTC ACCAGAAGTA 3535
 TCGATTACAA CGGTGTCACT TCCCAAAGGT GCGGCTGCTA TCAATGGCAT GGGAGAAGCA 3725
 CTGAATGCTG CCGGCCCCGA TGGAAATGGCC TCCCTATCTC TGCCATTACC CCTTTGGACC 3785
 GGCAGAGGGA CGGCTCCTGG ATTATCGCTG ATTTACAGCA ACAGTGCAGG TAATGGGCTT 3845
 TTGGGCATCG GCTGGCAATG CGGTGTTATG TCCATTAGCC GACGGACCCA ACATGGCATT 3905
 CGACAATACG GTAATGACGA CAGGTTCCCTA TCCCCACAAG GCGAGGTCAT GAATATCGCC 3965
 CTGAATGACC AAGGGCAACC TGATATCCGT CAAGACGTTA AAACGGTCCA AGCGGTTACC 4025
 TTGCCAATTT CCTATACCGT GACCCGCTAT CAAGCCCCCG AGATCCTGGA TTTCAGTA 4085
 ATCGAATACT GGCAACCTGC CTCCGGTCAA GAAGGACCGG CTTTCTGGCT GATATCGACA 4145
 CCGGACGGGC ATCTACACAT CTTAGGGAAA ACCCGCCAGG CTGTCTGGC AAATCCGCAA 4205
 AATGACCAAC AAATCGCCCA GTGGTTGCTG GAAGAACTG TGACCCCAAG CGGTGAACAT 4265
 GTCAGCTATC AATATCGAGC CGAAGATGAA GCGCATTTGT ACGACAATGA AAAAACCGCT 4325
 CATCCCAATG TTACCGCACA GCGCTATCTG GTACAGGTGA ACTACAGGCA ACATCAAACC 4385
 ACAAGCCAGC CTGTTCTGAC TGGATAACCG ACCTCCCGCA CCGAAGAGT GGCTGTTTCA 4445
 TCTGGTCTTT GACCACGGTG AGCGCGTACC TCACTTCATA CCGTGCCAAC ATGGGATGCA 4505
 GGTACAGCGC AATGGTCTGT ACGCCCGGAT ATCTTCTCTC GGTATGAATA TGGTTTTGAA 4565
 GTGCGTACTC GCGGCTTATG TCAACAAGTG CTGATGTTTC ACCGCACCGC GCTCATGGCC 4625
 GGAGAAGCCA GTACCAATGA CGCCCCGAA CTGGTTGGAC GCTTAATACT GGAATATGAC 4685
 AAAAACGCCA GCGTCACCAC GTTGATTACC ATCCGTCAAT TAAGCCATGA ATCGGACGGG 4745
 AGGCCAGTCA CCCAGCCACC ACTAGAATA GCCTGGCAAC GGTTCGATCT GGAGAAAATC 4805
 CCGACATGGC AAGGCTTTGA CGCACTAGAT AATTTTAACT CGCAGCAACG TTATCAACTG 4865
 GTTGATCTGC GCGGAGAAGG GTTGCCAGGT ATGCTGTATC AAGATCGAGG CGCTTGGTGG 4925
 TATAAAGCTC CGCAACGTC AAGAGACGGA GACAGCAATG CCGTCACTTA CGACAAAATC 4985
 GCGCCACTGC CTACCCTACC CAATTTCAG GATAATGCCT CATTCATGGA TATCAACGGA 5045
 GACGGCCAAC TGGATTGGGT TGTTACCGCC TCCCGTATTC GCGGATACCA TAGTCAGCAA 5105
 CCGATGGAA AGTGGACGCA CTTTACGCCA ATCAATGCCT TGCCCGTGA ATATTTTCAT 5165
 CCAAGCATCC AGTTCCCTGA CCTTACCGGG GCAGGCTTAT CTGATTTAGT GTTGATCGGG 5225
 CCGAAAAGCG TCGCTCTATA TGCCAACCAG CGAAACGGCT GCGGTAAAGG AGAAGATGTC 5285
 CCCCAAATCA CAGGTATCAC CCGTCCCTGC ACAGGGACCG ATGCCCCGAA ACTGGTGGCT 5345
 TTCAGTGATA TGCTCGGTTT CCGTCAACAA CATCTGGTGG AAATCAAGGG TAATCGGTC 5405
 ACCTGTTGGC CGAATCTAGG GCATGGCCGT TTCGGTCAAC CACTAACTCT GTCAGGATTT 5465
 AGCCAGCCCC AAAATAGCTT CAATCCCGAA CGGCTGTTTC TGGCGGATAT CGACGGCTCC 5525
 GGCACCACCG ACCTTATCTA TCGCAATCC GGCTCTTTGC TCATTTATCT CAACCAAAGT 5585

U
A
8
2
4
8
5
C
2

CGTAATCAGT TTGATCCCCC GTTGACATTA GCSTTCCCAG AAGGCCTACA ATTTGACAAC 5445
 ACTTCCCAAC TTCAAGTGGC CCAATATTCAG GCATTAGGGA TAGCCAGCTT GATTCTGACT 5795
 GTGCCACATA TCGCCCCACA TCACTGGCGT TOTGACCTGT CACTGACCAA ACCCTGGTTG 5755
 TTGAATGTAA TGAACAATAA CCGGGGGCGCA CATCACACGC TACATTATCG TAGTTCCGGC 5825
 CAATTCTGGT TGGATGAAA ATTACAGCTC ACCAAAGCAG GCAAATCTCC GGCTTCTTAT 5885
 CTGCCCTTTC CAATGCATTT GCTATGGTAT ACCGAAATTC AGGATGAAAT CAGCGCCAAC 5945
 CGGCTCACCA CTGAAGTCAA CTACAGCCAC GGCCTCTGGG ATGCTAAAGA GC CGAATTC 6005

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 26:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 1190 амінокислот

(B) ТИП: амінокислота

(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 1:

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

U A 8 2 4 8 5 C 2

Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly
 210 215 220
 Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn
 225 230 235 240
 Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp
 245 250 255
 Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala
 260 265 270
 Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val
 275 280 285
 Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly
 290 295 300
 Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe
 305 310 315 320
 Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr
 325 330 335
 Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn
 340 345 350
 Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly
 355 360 365
 Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn
 370 375 380
 Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln
 385 390 395 400
 Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr
 405 410 415
 Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn
 420 425 430
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe
 435 440 445
 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro
 450 455 460
 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala
 465 470 475 480
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu
 530 535 540
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Gln Ala Lys Met Tyr
565 570 575

Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
580 585 590

Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
595 600 605

Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
610 615 620

Trp Leu Ser Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu
625 630 635 640

Pro Pro Tyr Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu
645 650 655

Arg Leu Tyr Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu
660 665 670

Asn Leu Pro Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg
675 680 685

Gln Gln Ala Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln
690 695 700

Gly Ser Val Gln Gly Trp Arg Tyr Pro Leu Leu Val Glu Arg Ala Arg
705 710 715 720

Ser Ala Val Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr
725 730 735

Leu Glu His Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln
740 745 750

Gln Glu Ala Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu
755 760 765

Lys Gly Leu Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly
770 775 780

Asp Thr Leu Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu
785 790 795 800

Ser Ala Ala Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile
805 810 815

Thr Asn Gly Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn
820 825 830

Ala Val Pro Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly
835 840 845

Ala Pro Leu Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile
850 855 860

Gln Asp Gln Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg
865 870 875 880

Arg Gln Glu Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile
885 890 895

Thr Gln Leu Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met
900 905 910

U A 8 2 4 8 5 C 2

Ala Gln Lys Gln Ile Thr Leu Ser Glu Thr Gln Gln Ala Asn Ala Gln
915 920 925

Ala Ile Tyr Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr
930 935 940

Asn Trp Met Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp
945 950 955 960

Ser Thr Leu Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu
965 970 975

Leu Gly Glu Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn
980 985 990

Asp Leu Trp Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu
995 1000 1005

Gln Lys Leu Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu
1010 1015 1020

Ala Ile Arg Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu
1025 1030 1035 1040

Ser Glu Asn Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser
1045 1050 1055

Gly Gly Val Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu
1060 1065 1070

Asp Leu Ser Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu
1075 1080 1085

Lys Lys Arg Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu
1090 1095 1100

Gly Pro Tyr Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile
1105 1110 1115 1120

Ala Ala Leu Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp
1125 1130 1135

Phe Asn Asp Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr
1140 1145 1150

Gly Thr Leu Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln
1155 1160 1165

His Glu Leu Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr
1170 1175 1180

Ile Ile Arg Asp Ala *
1185 1190

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 27:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1881 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

(A) ІМ'Я/КЛЮЧ: CDS
 (B) МІСЦЕПОЛОЖЕННЯ: 1...1881
 (D) ІНША ІНФОРМАЦІЯ: /продукт="P8"

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 27:

ATG TCT GAA TCT TTA TTT ACA CAA ACG TTG AAA GAA GCG CGC GGT GAT	48
Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp	
1 5 10 15	
GCA TTG GTT GCT CAT TAT ATT GCT ACT CAG CTG CCC GCA GAT TTA AAA	96
Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys	
20 25 30	
GAG AGT ATC CAG ACC GCG GAT GAT CTG TAC GAA TAT CTG TTG CTG GAT	144
Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Leu Asp	
35 40 45	
ACC AAA ATT AGC GAT CTG GTT ACT ACT TCA CCG CTG TCC GAA GCG ATT	192
Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile	
50 55 60	
GGC AGT CTG CAA TTG TTT ATT CAT CGT GCG ATA GAG GGC TAT GAC GGC	240
Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly	
65 70 75 80	
ACG CTG GCA GAC TCA GCA AAA CCC TAT TTT GCC GAT GAA CAG TTT TTA	288
Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp Glu Gln Phe Leu	
85 90 95	
TAT AAC TGG GAT AGT TTT AAC CAC CGT TAT ACC ACT TGG GCT GGC AAG	336
Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr Trp Ala Gly Lys	
100 105 110	
GAA CCG TTG AAA TTC TAT GCC GGG GAT TAT ATT GAT CCA ACA TTG CGA	384
Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp Pro Thr Leu Arg	
115 120 125	
TTG AAT AAG ACC GAG ATA TTT ACC GCA TTT GAA CAA GGT ATT TCT CAA	432
Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln Gly Ile Ser Gln	
130 135 140	
GGG AAA TTA AAA AGT GAA TTA GTC GAA TCT AAA TTA CGT GAT TAT CTA	480
Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu Arg Asp Tyr Leu	
145 150 155 160	
ATT AGT TAT GAC ACT TTA GCC ACC CTT GAT TAT ATT ACT GCC TGC CAA	528
Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile Thr Ala Cys Gln	
165 170 175	
GGC AAA GAT AAT AAA ACC ATC TTC TTT ATT GGC CGT ACA CAG AAT GCA	576
Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala	
180 185 190	
CCC TAT GCA TTT TAT TGG CGA AAA TTA ACT TTA GTC ACT GAT GGC GGT	624
Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly	
195 200 205	
AAG TTG AAA CCA GAT CAA TGG TCA GAG TGG CGA GCA ATT AAT GCC GGG	672
Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly	
210 215 220	

U A 8 2 4 8 5 C 2

ATT AGT GAG GCA TAT TCA GGG CAT CTC GAG CTT TTC TCG GAA AAT AAC	711
Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn	
225 230 235 240	
AAG CTC CAC ATC COT TCG TTT ACT ATC TCG AAA GAA GAT AAA ATA GAT	758
Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp	
245 250 255	
TTT GTT TAT AAA AAC ATC TCG CTC ATC AGT AGC GAT TAT AGC TCG GCA	816
Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala	
260 265 270	
TCA AAG AAA AAA ATC TTG GAA CTT TCT TTT ACT GAC TAC AAT AGA GTT	854
Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val	
275 280 285	
GGA GCA ACA GGA TCA TCA AGC CCG ACT GAA GTA GCT TCA CAA TAT GGT	912
Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly	
290 295 300	
TCT GAT GCT CAG ATG AAT ATT TCT GAT GAT GGG ACT GTA CTT ATT TTT	960
Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe	
305 310 315 320	
CAG AAT GCC GGC GGA GCT ACT CCC AGT ACT GGA GTG ACG TTA TGT TAT	1008
Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr	
325 330 335	
GAC TCT GGC AAC GTG ATT AAG AAC CTA TCT AGT ACA GGA AGT GCA AAT	1056
Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn	
340 345 350	
TTA TCG TCA AAG GAT TAT GCC ACA ACT AAA TTA CGC ATG TGT CAT GGA	1104
Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly	
355 360 365	
CAA AGT TAC AAT GAT AAT AAC TAC TGC AAT TTT ACA CTC TCT ATT AAT	1152
Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn	
370 375 380	
ACA ATA GAA TTC ACC TCC TAC GGC ACA TTC TCA TCA GAT GGA AAA CAA	1200
Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln	
385 390 395 400	
TTT ACA CCA CCT TCT CGT TCT GCC ATT GAT TTA CAC CTC CCT AAT TAT	1248
Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr	
405 410 415	
GTA GAT CTC AAC GCG CTA TTA GAT ATT AGC CTC GAT TCA CTA CTT AAT	1296
Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn	
420 425 430	
TAT GAC GTT CAG GGC CAG TTT GGC GGA TCT AAT CCG GTT GAT AAT TTC	1344
Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe	
435 440 445	
AGT GGT CCC TAT GGT ATT TAT CTA TCG GAA ATC TTC TTC CAT ATT CCG	1392
Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro	
450 455 460	
TTC CTT GTT ACC GTC CGT ATG CAA ACC GAA CAA CGT TAC GAA GAC GCG	1440
Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala	
465 470 475 480	
GAC ACT TGG TAC AAA TAT ATT TTC CCG AGC GCC GGT TAT CGC GAT GCT	1488

U A 8 2 4 8 5 C 2

Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 AAT GGC CAG CTC ATT ATG GAT GGC AGT AAA CCA CGT TAT TGG AAT GTG 1535
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 ATG CCA TTG CAA CTG GAT ACC GCA TGG GAT ACC ACA CAG CCC GCC ACC 1534
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 ACT GAT CCA GAT GTG ATC GCT ATG GCG GAC CCG ATG CAT TAC AAG CTC 1612
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu
 530 535 540
 CCG ATA TTC CTG CAT ACC CTT GAT CTA TTC ATT GCC CGA GGC GAC AGC 1680
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560
 GCT TAC CGT CAA CTT GAA GCG GAT ACT CTA GTC GAA GCC AAA ATG TAC 1728
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr
 565 570 575
 TAC ATT CAG GCA CAA CAG CTA CTG GGA CCG CCG CCT GAT ATC CAT ACC 1776
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
 580 585 590
 ACC AAT ACT TGG CCA AAT CCC ACC TTG AGT AAA GAA GCT GCC GCT ATT 1824
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
 595 600 605
 GCC ACA CCG ACA TTC CTC AGT TCA CCG GAG GTG ATG ACC TTC GCT GCC 1872
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
 610 615 620
 TGG CTA AGC 1881
 Trp Leu Ser
 625

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 28:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 627 амінокислот
- (B) ТИП: амінокислота
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 28:

Met Ser Glu Ser Leu Phe Thr Gln Thr Leu Lys Glu Ala Arg Arg Asp
 1 5 10 15
 Ala Leu Val Ala His Tyr Ile Ala Thr Gln Val Pro Ala Asp Leu Lys
 20 25 30
 Glu Ser Ile Gln Thr Ala Asp Asp Leu Tyr Glu Tyr Leu Leu Leu Asp
 35 40 45
 Thr Lys Ile Ser Asp Leu Val Thr Thr Ser Pro Leu Ser Glu Ala Ile
 50 55 60
 Gly Ser Leu Gln Leu Phe Ile His Arg Ala Ile Glu Gly Tyr Asp Gly

U A 8 2 4 8 5 C 2

65	70	75	80
Thr Leu Ala Asp Ser Ala Lys Pro Tyr Phe Ala Asp Glu Gln Phe Leu	85	90	95
Tyr Asn Trp Asp Ser Phe Asn His Arg Tyr Ser Thr Trp Ala Gly Lys	100	105	110
Glu Arg Leu Lys Phe Tyr Ala Gly Asp Tyr Ile Asp Pro Thr Leu Arg	115	120	125
Leu Asn Lys Thr Glu Ile Phe Thr Ala Phe Glu Gln Gly Ile Ser Gln	130	135	140
Gly Lys Leu Lys Ser Glu Leu Val Glu Ser Lys Leu Arg Asp Tyr Leu	145	150	155
Ile Ser Tyr Asp Thr Leu Ala Thr Leu Asp Tyr Ile Thr Ala Cys Gln	165	170	175
Gly Lys Asp Asn Lys Thr Ile Phe Phe Ile Gly Arg Thr Gln Asn Ala	180	185	190
Pro Tyr Ala Phe Tyr Trp Arg Lys Leu Thr Leu Val Thr Asp Gly Gly	195	200	205
Lys Leu Lys Pro Asp Gln Trp Ser Glu Trp Arg Ala Ile Asn Ala Gly	210	215	220
Ile Ser Glu Ala Tyr Ser Gly His Val Glu Pro Phe Trp Glu Asn Asn	225	230	235
Lys Leu His Ile Arg Trp Phe Thr Ile Ser Lys Glu Asp Lys Ile Asp	245	250	255
Phe Val Tyr Lys Asn Ile Trp Val Met Ser Ser Asp Tyr Ser Trp Ala	260	265	270
Ser Lys Lys Lys Ile Leu Glu Leu Ser Phe Thr Asp Tyr Asn Arg Val	275	280	285
Gly Ala Thr Gly Ser Ser Ser Pro Thr Glu Val Ala Ser Gln Tyr Gly	290	295	300
Ser Asp Ala Gln Met Asn Ile Ser Asp Asp Gly Thr Val Leu Ile Phe	305	310	315
Gln Asn Ala Gly Gly Ala Thr Pro Ser Thr Gly Val Thr Leu Cys Tyr	325	330	335
Asp Ser Gly Asn Val Ile Lys Asn Leu Ser Ser Thr Gly Ser Ala Asn	340	345	350
Leu Ser Ser Lys Asp Tyr Ala Thr Thr Lys Leu Arg Met Cys His Gly	355	360	365
Gln Ser Tyr Asn Asp Asn Asn Tyr Cys Asn Phe Thr Leu Ser Ile Asn	370	375	380
Thr Ile Glu Phe Thr Ser Tyr Gly Thr Phe Ser Ser Asp Gly Lys Gln	385	390	395
Phe Thr Pro Pro Ser Gly Ser Ala Ile Asp Leu His Leu Pro Asn Tyr	405	410	415
Val Asp Leu Asn Ala Leu Leu Asp Ile Ser Leu Asp Ser Leu Leu Asn			

U A 8 2 4 8 5 C 2

420 425 430
 Tyr Asp Val Gln Gly Gln Phe Gly Gly Ser Asn Pro Val Asp Asn Phe
 435 440 445
 Ser Gly Pro Tyr Gly Ile Tyr Leu Trp Glu Ile Phe Phe His Ile Pro
 450 455 460
 Phe Leu Val Thr Val Arg Met Gln Thr Glu Gln Arg Tyr Glu Asp Ala
 465 470 475 480
 Asp Thr Trp Tyr Lys Tyr Ile Phe Arg Ser Ala Gly Tyr Arg Asp Ala
 485 490 495
 Asn Gly Gln Leu Ile Met Asp Gly Ser Lys Pro Arg Tyr Trp Asn Val
 500 505 510
 Met Pro Leu Gln Leu Asp Thr Ala Trp Asp Thr Thr Gln Pro Ala Thr
 515 520 525
 Thr Asp Pro Asp Val Ile Ala Met Ala Asp Pro Met His Tyr Lys Leu
 530 535 540
 Ala Ile Phe Leu His Thr Leu Asp Leu Leu Ile Ala Arg Gly Asp Ser
 545 550 555 560
 Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Val Glu Ala Lys Met Tyr
 565 570 575
 Tyr Ile Gln Ala Gln Gln Leu Leu Gly Pro Arg Pro Asp Ile His Thr
 580 585 590
 Thr Asn Thr Trp Pro Asn Pro Thr Leu Ser Lys Glu Ala Gly Ala Ile
 595 600 605
 Ala Thr Pro Thr Phe Leu Ser Ser Pro Glu Val Met Thr Phe Ala Ala
 610 615 620
 Trp Leu Ser
 625

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 29:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1689 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: CDS
- (B) МІСЦЕПОЛОЖЕННЯ: 1...1689
- (C) ІНША ІНФОРМАЦІЯ: /продукт="S8"

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 29:

GCA GGC GAT ACC GCA AAT ATT GGC GAC GGT GAT TTC TTG CCA CCG TAC 43
 Ala Gly Asp Thr Ala Asn Ile Gly Asp Gly Asp Phe Leu Pro Pro Tyr
 1 5 10 15

AAC GAT GTA CTA CTC GGT TAC TGG GAT AAA CTT GAG TTA CGC CTA TAC	36
Asn Asp Val Leu Leu Gly Tyr Trp Asp Lys Leu Glu Leu Arg Leu Tyr	
20 25 30	
AAC CTC CGC CAC AAT CTG AGT CTG GAT GGT CAA CGC CTA AAT CTG CCA	144
Asn Leu Arg His Asn Leu Ser Leu Asp Gly Gln Pro Leu Asn Leu Pro	
35 40 45	
CTG TAT GCC ACC CCG GTA GAC CCG AAA ACC CTG CAA CGC CAG CAA GCC	192
Leu Tyr Ala Thr Pro Val Asp Pro Lys Thr Leu Gln Arg Gln Gln Ala	
50 55 60	
GGA GGG GAC GGT ACA GGC AGT AGT CCG GCT GGT GGT CAA GGC AGT GTT	240
Gly Gly Asp Gly Thr Gly Ser Ser Pro Ala Gly Gly Gln Gly Ser Val	
65 70 75 80	
CAG GGC TGG CGC TAT CCG TTA TTG GTA GAA CGC GCC CGC TCT GCC GTG	288
Gln Gly Trp Arg Pro Leu Leu Val Glu Arg Ala Arg Ser Ala Val	
85 90 95	
AGT TTG TTG ACT CAG TTC GGC AAC AGC TTA CAA ACA ACG TTA GAA CAT	336
Ser Leu Leu Thr Gln Phe Gly Asn Ser Leu Gln Thr Thr Leu Glu His	
100 105 110	
CAG GAT AAT GAA AAA ATG ACG ATA CTG TTG CAG ACT CAA CAG GAA GCC	384
Gln Asp Asn Glu Lys Met Thr Ile Leu Leu Gln Thr Gln Gln Glu Ala	
115 120 125	
ATC CTG AAA CAT CAG CAC GAT ATA CAA CAA AAT AAT CTA AAA GGA TTA	432
Ile Leu Lys His Gln His Asp Ile Gln Gln Asn Asn Leu Lys Gly Leu	
130 135 140	
CAA CAC AGC CTG ACC GCA TTA CAG GCT AGC CGT GAT GGC GAC ACA TTG	480
Gln His Ser Leu Thr Ala Leu Gln Ala Ser Arg Asp Gly Asp Thr Leu	
145 150 155 160	
CGG CAA AAA CAT TAC AGC GAC CTG ATT AAC GGT GGT CTA TCT GCG GCA	528
Arg Gln Lys His Tyr Ser Asp Leu Ile Asn Gly Gly Leu Ser Ala Ala	
165 170 175	
GAA ATC GCC GGT CTG ACA CTA CGC AGC ACC GCC ATG ATT ACC AAT GGC	576
Glu Ile Ala Gly Leu Thr Leu Arg Ser Thr Ala Met Ile Thr Asn Gly	
180 185 190	
GTT GCA ACG GGA TTG CTG ATT GCC GGC GGA ATC GCC AAC GCG GTA CCT	624
Val Ala Thr Gly Leu Leu Ile Ala Gly Gly Ile Ala Asn Ala Val Pro	
195 200 205	
AAC GTC TTC GGG CTG GCT AAC GGT GGA TCG GAA TCG GGA GCG CCA TTA	672
Asn Val Phe Gly Leu Ala Asn Gly Gly Ser Glu Trp Gly Ala Pro Leu	
210 215 220	
ATT GGC TCC GGG CAA GCA ACC CAA GTT GGC GCC GGC ATC CAG GAT CAG	720
Ile Gly Ser Gly Gln Ala Thr Gln Val Gly Ala Gly Ile Gln Asp Gln	
225 230 235 240	
AGC GCG GGC ATT TCA GAA GTG ACA CCA GGC TAT CAG CGT CGT CAG GAA	768
Ser Ala Gly Ile Ser Glu Val Thr Ala Gly Tyr Gln Arg Arg Gln Glu	
245 250 255	
GAA TGG GCA TTG CAA CCG GAT ATT GCT GAT AAC GAA ATA ACC CAA CTG	816
Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile Thr Gln Leu	
260 265 270	
GAT GCC CAG ATA CAA AGC CTG CAA GAG CAA ATC ACG ATG GCA CAA AAA	864

Asp	Ala	Gln	Ile	Gln	Ser	Leu	Gln	Glu	Gln	Ile	Thr	Met	Ala	Gln	Lys	
	275						280					285				
CAG	ATC	ACG	CTC	TCT	GAA	ACC	GAA	CAA	GCG	AAT	GCC	CAA	GCG	ATT	TAT	312
Gln	Ile	Thr	Leu	Ser	Glu	Thr	Glu	Gln	Ala	Asn	Ala	Gln	Ala	Ile	Tyr	
	290					295					300					
CAC	CTG	CAA	ACC	ACT	CGT	TTT	ACC	GCG	CAG	GCA	CTG	TAT	AAC	TOG	ATG	360
Asp	Leu	Gln	Thr	Thr	Arg	Phe	Thr	Gly	Gln	Ala	Leu	Tyr	Asn	Trp	Met	
305					310				315					320		
GCC	GGT	CGT	CTC	TTC	GCG	CTC	TAT	TAC	CAA	ATG	TAT	GAT	TCC	ACT	CTG	1008
Ala	Gly	Arg	Leu	Ser	Ala	Leu	Tyr	Tyr	Gln	Met	Tyr	Asp	Ser	Thr	Leu	
				325					330					335		
CCA	ATC	TGT	CTC	CAG	CCA	AAA	GCC	GCA	TTA	GTA	CAG	GAA	TTA	GCC	GAG	1056
Pro	Ile	Cys	Leu	Gln	Pro	Lys	Ala	Ala	Leu	Val	Gln	Glu	Leu	Gly	Glu	
			340					345					350			
AAA	GAG	AGC	GAC	AGT	CTT	TTC	CAG	GTT	CCG	GTG	TOG	AAT	GAT	CTG	TOG	1104
Lys	Glu	Ser	Asp	Ser	Leu	Phe	Gln	Val	Pro	Val	Trp	Asn	Asp	Leu	Trp	
	355					360						365				
CAA	GCG	CTG	TTA	GCA	GGA	GAA	GGT	TTA	AGT	TCA	GAG	CTA	CAG	AAA	CTG	1152
Gln	Gly	Leu	Leu	Ala	Gly	Glu	Gly	Leu	Ser	Ser	Glu	Leu	Gln	Lys	Leu	
	370					375					380					
GAT	GCC	ATC	TGC	CTT	GCA	CGT	GGT	GGT	ATT	GCG	CTA	GAA	GCC	ATC	GCG	1200
Asp	Ala	Ile	Trp	Leu	Ala	Arg	Gly	Gly	Ile	Gly	Leu	Glu	Ala	Ile	Arg	
385					390				395					400		
ACC	CTG	TCC	CTG	GAT	ACC	CTG	TTT	GCG	ACA	GCG	ACG	TTA	AGT	GAA	AAT	1248
Thr	Val	Ser	Leu	Asp	Thr	Leu	Phe	Gly	Thr	Gly	Thr	Leu	Ser	Glu	Asn	
				405				410						415		
ATC	AAT	AAA	GTG	CTT	AAC	GCG	GAA	ACG	GTA	TCT	CCA	TCC	GGT	GCG	GTG	1296
Ile	Asn	Lys	Val	Leu	Asn	Gly	Glu	Thr	Val	Ser	Pro	Ser	Gly	Gly	Val	
			420				425						430			
ACT	CTG	GCG	CTG	ACA	GCG	GAT	ATC	TTC	CAA	GCA	ACA	CTG	GAT	TTG	AGT	1344
Thr	Leu	Ala	Leu	Thr	Gly	Asp	Ile	Phe	Gln	Ala	Thr	Leu	Asp	Leu	Ser	
			435				440						445			
CAG	CTA	GGT	TTG	GAT	AAC	TCT	TAC	AAC	TTG	GGT	AAC	GAG	AAG	AAA	CGT	1392
Gln	Leu	Gly	Leu	Asp	Asn	Ser	Tyr	Asn	Leu	Gly	Asn	Glu	Lys	Lys	Arg	
	450					455					460					
CGT	ATT	AAA	CGT	ATC	GCC	CTC	ACC	CTG	CCA	ACA	CTT	CTG	GGG	CCA	TAT	1440
Arg	Ile	Lys	Arg	Ile	Ala	Val	Thr	Leu	Pro	Thr	Leu	Leu	Gly	Pro	Tyr	
465					470				475					480		
CAA	GAT	CTT	GAA	GCC	ACA	CTG	GTA	ATG	GGT	GCG	GAA	ATC	GCC	GCC	TTA	1488
Gln	Asp	Leu	Glu	Ala	Thr	Leu	Val	Met	Gly	Ala	Glu	Ile	Ala	Ala	Leu	
				485					490					495		
TCA	CAC	GGT	GTG	AAT	CAC	GGA	GCG	CGG	TTT	GTT	ACC	GAC	TTT	AAC	GAC	1536
Ser	His	Gly	Val	Asn	Asp	Gly	Gly	Arg	Phe	Val	Thr	Asp	Phe	Asn	Asp	
			500					505					510			
AGC	CGT	TTT	CTG	CCT	TTT	GAA	GGT	CGA	GAT	GCA	ACA	ACC	GCG	ACA	CTG	1584
Ser	Arg	Phe	Leu	Pro	Phe	Glu	Gly	Arg	Asp	Ala	Thr	Thr	Gly	Thr	Leu	
		515					520					525				
GAG	CTC	AAT	ATT	TTC	CAT	GCG	GGT	AAA	GAG	GGA	ACG	CAA	CAC	GAG	TTC	1632
Glu	Leu	Asn	Ile	Phe	His	Ala	Gly	Lys	Glu	Gly	Thr	Gln	His	Glu	Leu	
	530						535					540				

CTC GCG AAT CTC ACT GAC ATC ATT CTC CAT CTC AAT TAC ATC ATT CGA 1530
 Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg
 545 550 555 560

GAC GCG TAA
 Asp Ala

1533

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 30:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 563 амінокислоти

(B) ТИП: амінокислота

(C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 30:

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

225	230	235	240
Ser Ala Gly Ile	Ser Glu Val Thr Ala Gly Tyr Gln Arg Arg Gln Glu		
	245	250	255
Glu Trp Ala Leu Gln Arg Asp Ile Ala Asp Asn Glu Ile Thr Gln Leu			
	260	265	270
Asp Ala Gln Ile Gln Ser Leu Gln Glu Gln Ile Thr Met Ala Gln Lys			
	275	280	285
Gln Ile Thr Leu Ser Glu Thr Glu Gln Ala Asn Ala Gln Ala Ile Tyr			
	290	295	300
Asp Leu Gln Thr Thr Arg Phe Thr Gly Gln Ala Leu Tyr Asn Trp Met			
	305	310	315
Ala Gly Arg Leu Ser Ala Leu Tyr Tyr Gln Met Tyr Asp Ser Thr Leu			
	325	330	335
Pro Ile Cys Leu Gln Pro Lys Ala Ala Leu Val Gln Glu Leu Gly Glu			
	340	345	350
Lys Glu Ser Asp Ser Leu Phe Gln Val Pro Val Trp Asn Asp Leu Trp			
	355	360	365
Gln Gly Leu Leu Ala Gly Glu Gly Leu Ser Ser Glu Leu Gln Lys Leu			
	370	375	380
Asp Ala Ile Trp Leu Ala Arg Gly Gly Ile Gly Leu Glu Ala Ile Arg			
	385	390	395
Thr Val Ser Leu Asp Thr Leu Phe Gly Thr Gly Thr Leu Ser Glu Asn			
	405	410	415
Ile Asn Lys Val Leu Asn Gly Glu Thr Val Ser Pro Ser Gly Gly Val			
	420	425	430
Thr Leu Ala Leu Thr Gly Asp Ile Phe Gln Ala Thr Leu Asp Leu Ser			
	435	440	445
Gln Leu Gly Leu Asp Asn Ser Tyr Asn Leu Gly Asn Glu Lys Lys Arg			
	450	455	460
Arg Ile Lys Arg Ile Ala Val Thr Leu Pro Thr Leu Leu Gly Pro Tyr			
	465	470	475
Gln Asp Leu Glu Ala Thr Leu Val Met Gly Ala Glu Ile Ala Ala Leu			
	485	490	495
Ser His Gly Val Asn Asp Gly Gly Arg Phe Val Thr Asp Phe Asn Asp			
	500	505	510
Ser Arg Phe Leu Pro Phe Glu Gly Arg Asp Ala Thr Thr Gly Thr Leu			
	515	520	525
Glu Leu Asn Ile Phe His Ala Gly Lys Glu Gly Thr Gln His Glu Leu			
	530	535	540
Val Ala Asn Leu Ser Asp Ile Ile Val His Leu Asn Tyr Ile Ile Arg			
	545	550	555
Asp Ala			560

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 31:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 4458 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(11) ТИП МОЛЕКУЛИ: ДНК (геномна)

(ix) ОЗНАКА:

- (A) ІМ'Я/КЛЮЧ: CDS
- (B) МІСЦЕПОЛОЖЕННЯ: 1...4458

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 25:

ATG CAG GAT TCA CCA GAA GTA TCG ATT ACA ACG CTG TCA CTT CCC AAA 48
Met Gln Asp Ser Pro Glu Val Ser Ile Thr Thr Leu Ser Leu Pro Lys
1 5 10 15

CGT GGC GGT GCT ATC AAT GGC ATG GGA GAA GCA CTG AAT GCT GCG GGC 96
Gly Gly Gly Ala Ile Asn Gly Met Gly Glu Ala Leu Asn Ala Ala Gly
20 25 30

CCT GAT GGA ATG GCC TCC CTA TCT CTG CCA TTA CCC CTT TCG ACC GGC 144
Pro Asp Gly Met Ala Ser Leu Ser Leu Pro Leu Pro Leu Ser Thr Gly
35 40 45

AGA GGG ACG GCT CCT GGA TTA TCG CTG ATT TAC AGC AAC AGT GCA GGT 192
Arg Gly Thr Ala Pro Gly Leu Ser Leu Ile Tyr Ser Asn Ser Ala Gly
50 55 60

AAT GGG CCT TTC GGC ATC GGC TGG CAA TGC GGT GTT ATC TCC ATT AGC 240
Asn Gly Pro Phe Gly Ile Gly Trp Gln Cys Gly Val Met Ser Ile Ser
65 70 75 80

CGA CGC ACC CAA CAT GGC ATT CCA CAA TAC GGT AAT GAC GAC ACG TTC 288
Arg Arg Thr Gln His Gly Ile Pro Gln Tyr Gly Asn Asp Asp Thr Phe
85 90 95

CTA TCC CCA CAA GGC GAG GTC ATG AAT ATC GCC CTG AAT GAC CAA GGC 336
Leu Ser Pro Gln Gly Glu Val Met Asn Ile Ala Leu Asn Asp Gln Gly
100 105 110

CAA CCT GAT ATC CGT CAA GAC GTT AAA ACG CTG CAA GGC GTT ACC TTG 384
Gln Pro Asp Ile Arg Gln Asp Val Lys Thr Leu Gln Gly Val Thr Leu
115 120 125

CCA ATT TCC TAT ACC GTG ACC CGC TAT CAA GCC CGC CAG ATC CTG GAT 432
Pro Ile Ser Tyr Thr Val Thr Arg Tyr Gln Ala Arg Gln Ile Leu Asp
130 135 140

TTC AGT AAA ATC GAA TAC TGG CAA CCT GCC TCC GGT CAA GAA GGA GGC 480
Phe Ser Lys Ile Glu Tyr Trp Gln Pro Ala Ser Gly Gln Glu Gly Arg
145 150 155 160

GCT TTC TGG CTG ATA TCG ACA CCG GAC GGG CAT CTA CAC ATC TTA GGC 528
Ala Phe Trp Leu Ile Ser Thr Pro Asp Gly His Leu His Ile Leu Gly
165 170 175

AAA ACC GCG CAG GCT TGT CTG GCA AAT CCG CAA AAT GAC CAA CAA ATC 576
Lys Thr Ala Gln Ala Cys Leu Ala Asn Pro Gln Asn Asp Gln Gln Ile

130	135	190	
GGC CAG TGG TTG CTG GAA GAA ACT CTG AGC CCA GCC GGT GAA CAT GTC 624			
Ala Gln Trp Leu Leu Glu Glu Thr Val Thr Pro Ala Gly Glu His Val			
195	200	205	
AGC TAT CAA TAT CGA GGC GAA GAT GAA GCC CAT TGT GAC GAC AAT GAA 672			
Ser Tyr Gln Tyr Arg Ala Glu Asp Glu Ala His Cys Asp Asp Asn Glu			
210	215	220	
AAA ACC GCT CAT CCC AAT GTT ACC GCA CAG CGC TAT CTG GTA CAG GTG 720			
Lys Thr Ala His Pro Asn Val Thr Ala Gln Arg Tyr Leu Val Gln Val			
225	230	235	240
AAC TAC GGC AAC ATC AAA CCA CAA GCC AGC CTG TTC GTA CTG GAT AAC 758			
Asn Tyr Gly Asn Ile Lys Pro Gln Ala Ser Leu Phe Val Leu Asp Asn			
245	250	255	
GCA CCT CCC GCA CCG GAA GAG TGG CTG TTT CAT CTG GTC TTT GAC CAC 816			
Ala Pro Pro Ala Pro Glu Glu Trp Leu Phe His Leu Val Phe Asp His			
260	265	270	
GGT GAG CGC GAT ACC TCA GTT CAT ACC CTG CCA ACA TGG GAT GCA GGT 864			
Gly Glu Arg Asp Thr Ser Leu His Thr Val Pro Thr Trp Asp Ala Gly			
275	280	285	
ACA GCG CAA TGG TCT GTA CGC CCG GAT ATC TTC TCT CGC TAT GAA TAT 912			
Thr Ala Gln Trp Ser Val Arg Pro Asp Ile Phe Ser Arg Tyr Glu Tyr			
290	295	300	
GGT TTT GAA GTG CGT ACT CGC CCG TTA TGT CAA CAA GTG CTG ATG TTT 960			
Gly Phe Glu Val Arg Thr Arg Arg Leu Cys Gln Gln Val Leu Met Phe			
305	310	315	320
CAC CGC ACC GCG CTC ATG GCC GGA GAA GCC AGT ACC AAT GAC GCC CCG 1008			
His Arg Thr Ala Leu Met Ala Gly Glu Ala Ser Thr Asn Asp Ala Pro			
325	330	335	
GAA CTG GTT GGA CGC TTA ATA CTG GAA TAT GAC AAA AAC GCC AGC GTC 1056			
Glu Leu Val Gly Arg Leu Ile Leu Glu Tyr Asp Lys Asn Ala Ser Val			
340	345	350	
ACC ACC TTG ATT ACC ATC CGT CAA TTA AGC CAT GAA TCG GAC GGG AGG 1104			
Thr Thr Leu Ile Thr Ile Arg Gln Leu Ser His Glu Ser Asp Gly Arg			
355	360	365	
CCA GTC ACC CAG CCA CCA CTA GAA CTA GCC TGG CAA CGG TTT GAT CTG 1152			
Pro Val Thr Gln Pro Pro Leu Glu Leu Ala Trp Gln Arg Phe Asp Leu			
370	375	380	
GAG AAA ATC CCG ACA TGG CAA CGC TTT GAC GCA CTA GAT AAT TTT AAC 1200			
Glu Lys Ile Pro Thr Trp Gln Arg Phe Asp Ala Leu Asp Asn Phe Asn			
385	390	395	400
TGG CAG CAA CGT TAT CAA CTG GTT GAT CTG CGG GGA GAA GCG TTG CCA 1248			
Ser Gln Gln Arg Tyr Gln Leu Val Asp Leu Arg Gly Glu Gly Leu Pro			
405	410	415	
GGT ATG CTG TAT CAA GAT CGA GGC GGT TGG TGG TAT AAA GCT CCG CAA 1296			
Gly Met Leu Tyr Gln Asp Arg Gly Ala Trp Trp Tyr Lys Ala Pro Gln			
420	425	430	
CGT CAG GAA GAC GGA GAC AGC AAT GCC GTC ACT TAC GAC AAA ATC GCC 1344			
Arg Gln Glu Asp Gly Asp Ser Asn Ala Val Thr Tyr Asp Lys Ile Ala			
435	440	445	

U A 8 2 4 8 5 C 2

TTA	TTG	CGT	ACC	CTA	CCC	AAT	TTG	CAG	GAT	AAT	CCC	TCA	TTG	ATG	GAT	1391
Pro	Leu	Pro	Thr	Leu	Pro	Asn	Leu	Gln	Asp	Asn	Ala	Ser	Leu	Met	Asp	
450						455					460					
ATC	AAC	GGA	GAC	GGC	CAA	CTG	GAT	TGG	GTT	GTT	ACC	GGC	TCC	GGT	ATT	1446
Ile	Asn	Gly	Asp	Gly	Gln	Leu	Asp	Trp	Val	Val	Thr	Ala	Ser	Gly	Ile	
465					470				475						480	
GGC	GGA	TAC	CAT	AGT	CAG	CAA	CCC	GAT	GGA	AAG	TGG	ACC	CAC	TTT	AGC	1488
Arg	Gly	Tyr	His	Ser	Gln	Gln	Pro	Asp	Gly	Lys	Trp	Thr	His	Phe	Thr	
				485					490					495		
CCA	ATC	AAT	CCC	TTG	CCC	GTG	GAA	TAT	TTT	CAT	CCA	AGC	ATC	CAG	TTC	1536
Pro	Ile	Asn	Ala	Leu	Pro	Val	Glu	Tyr	Phe	His	Pro	Ser	Ile	Gln	Phe	
			500					505					510			
GCT	GAC	CTT	ACC	GGG	GCA	GGC	TTA	TCT	GAT	TTA	GTG	TTG	ATC	GGG	CCG	1584
Ala	Asp	Leu	Thr	Gly	Ala	Gly	Leu	Ser	Asp	Leu	Val	Leu	Ile	Gly	Pro	
		515					520					525				
AAA	AGC	GTG	CGT	CTA	TAT	CCC	AAC	CAG	CGA	AAC	GGC	TGG	CGT	AAA	GGA	1632
Lys	Ser	Val	Arg	Leu	Tyr	Ala	Asn	Gln	Arg	Asn	Gly	Trp	Arg	Lys	Gly	
		530					535					540				
GAA	GAT	GTG	CCC	CAA	TCC	ACA	GGT	ATC	ACC	CTG	CCT	GTG	ACA	GGG	ACC	1680
Glu	Asp	Val	Pro	Gln	Ser	Thr	Gly	Ile	Thr	Leu	Pro	Val	Thr	Gly	Thr	
545					550					555					560	
GAT	CCC	CCC	AAA	CTG	GTG	GCT	TTC	AGT	GAT	ATG	CTC	GGT	TCC	GGT	CAA	1728
Asp	Ala	Arg	Lys	Leu	Val	Ala	Phe	Ser	Asp	Met	Leu	Gly	Ser	Gly	Gln	
				565					570					575		
CAA	CAT	CTG	GTG	GAA	ATC	AAG	GGT	AAT	CGC	GTG	ACC	TGT	TGG	CCG	AAT	1776
Gln	His	Leu	Val	Glu	Ile	Lys	Gly	Asn	Arg	Val	Thr	Cys	Trp	Pro	Asn	
			580					585					590			
CTA	GGG	CAT	GGC	CGT	TTC	GGT	CAA	CCA	CTA	ACT	CTG	TCA	GGA	TTT	AGC	1824
Leu	Gly	His	Gly	Arg	Phe	Gly	Gln	Pro	Leu	Thr	Leu	Ser	Gly	Phe	Ser	
		595					600					605				
CAG	CCC	GAA	AAT	AGC	TTC	AAT	CCC	GAA	CGG	CTG	TTT	CTG	GGG	GAT	ATC	1872
Gln	Pro	Glu	Asn	Ser	Phe	Asn	Pro	Glu	Arg	Leu	Phe	Leu	Ala	Asp	Ile	
		610				615					620					
GAC	GGC	TCC	GGC	ACC	ACC	GAC	CTT	ATC	TAT	CGC	CAA	TCC	GGC	TCT	TTG	1920
Asp	Gly	Ser	Gly	Thr	Thr	Asp	Leu	Ile	Tyr	Ala	Gln	Ser	Gly	Ser	Leu	
		625			630					635					640	
CTC	ATT	TAT	CTC	AAC	CAA	AGT	GGT	AAT	CAG	TTT	GAT	GGC	CCG	TTG	ACA	1968
Leu	Ile	Tyr	Leu	Asn	Gln	Ser	Gly	Asn	Gln	Phe	Asp	Ala	Pro	Leu	Thr	
				645					650					655		
TTA	GGG	TTG	CCA	GAA	GGC	GTA	CAA	TTT	GAC	AAC	ACT	TGC	CAA	CTT	CAA	2016
Leu	Ala	Leu	Pro	Glu	Gly	Val	Gln	Phe	Asp	Asn	Thr	Cys	Gln	Leu	Gln	
			660					665					670			
CTC	GGC	GAT	ATT	CAG	GGA	TTA	GGG	ATA	GCC	AGC	TTG	ATT	CTG	ACT	CTG	2064
Val	Ala	Asp	Ile	Gln	Gly	Leu	Gly	Ile	Ala	Ser	Leu	Ile	Leu	Thr	Val	
		675					680					685				
CCA	CAT	ATC	GGC	CCA	CAT	CAC	TGG	CGT	TGT	GAC	CTG	TCA	CTG	ACC	AAA	2112
Pro	His	Ile	Ala	Pro	His	His	Trp	Arg	Cys	Asp	Leu	Ser	Leu	Thr	Lys	
		690				695					700					
CCC	TGG	TTG	TTG	AAT	GTA	ATG	AAC	AAT	AAC	CGG	GGC	GCA	CAT	CAC	ACC	2160
Pro	Trp	Leu	Leu	Asn	Val	Met	Asn	Asn	Asn	Arg	Gly	Ala	His	His	Thr	

U A 8 2 4 8 5 C 2

705	710	715	720
CTA CAT TAT CGT AGT TCC GCG CAA TTC TCG TTG GAT GAA AAA TTA CAG 2308 Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln 725 730 735			
CTC ACC AAA GCA GGC AAA TCT CCG GCT TGT TAT CTC CCG TTT CCA ATC 2356 Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met 740 745 750			
CAT TTG CTA TGG TAT ACC GAA ATT CAG GAT GAA ATC AGC GGC AAC CCG 2304 His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg 755 760 765			
CTC ACC AGT GAA GTC AAC TAC AGC CAC GGC CTC TGG GAT GGT AAA GAG 2352 Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu 770 775 780			
CGG GAA TTC AGA CGA TTT GCG TGC ATC AAA CAG ACA GAT ACC ACA ACG 2400 Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr 785 790 795 800			
TTT TCT CAC GGC ACC GGC CCC GAA CAG GCG GCA CCG TCG CTC AGT ATT 2448 Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile 805 810 815			
AGC TGG TTT CCC ACC GGC ATG GAT GAA GTA GAC AGC CAA TTA GCT ACG 2496 Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr 820 825 830			
GAA TAT TGG CAG GCA GAC ACC CAA GCT TAT AGC GGA TTT GAA ACC CGT 2544 Glu Tyr Trp Gln Ala Asp Thr Gln Ala Tyr Ser Gly Phe Glu Thr Arg 835 840 845			
TAT ACC GTC TGG GAT CAC ACC AAC CAG ACA GAC CAA GCA TTT ACC CCC 2592 Tyr Thr Val Trp Asp His Thr Asn Gln Thr Asp Gln Ala Phe Thr Pro 850 855 860			
AAT GAG ACA CAA CGT AAC TGG CTC ACC CGA CCG CTT AAA GGC CAA CTG 2640 Asn Glu Thr Gln Arg Asn Trp Leu Thr Arg Ala Leu Lys Gly Gln Leu 865 870 875 880			
CTA CCG ACT CAG CTC TAC GGT CTC GAC GGA ACA GAT AAG CAA ACA GTC 2688 Leu Arg Thr Glu Leu Tyr Gly Leu Asp Gly Thr Asp Lys Gln Thr Val 885 890 895			
CCT TAT ACC GTC AGT GAA TCG CCG TAT CAG GTA CCG TCT ATT CCC GTA 2736 Pro Tyr Thr Val Ser Glu Ser Arg Tyr Gln Val Arg Ser Ile Pro Val 900 905 910			
AAT AAA GAA ACT GAA TTA TCT GCC TGG GTG ACT GCT ATT GAA AAT CCG 2784 Asn Lys Glu Thr Glu Leu Ser Ala Trp Val Thr Ala Ile Glu Asn Arg 915 920 925			
AGC TAC CAC TAT GAA CGT ATC ATC ACT GAC CCA CAG TTC AGC CAG AGT 2832 Ser Tyr His Tyr Glu Arg Ile Ile Thr Asp Pro Gln Phe Ser Gln Ser 930 935 940			
ATC AAG TTG CAA CAC GAT ATC TTT GGT CAA TCA CTC CAA AGT GTC GAT 2880 Ile Lys Leu Gln His Asp Ile Phe Gly Gln Ser Leu Gln Ser Val Asp 945 950 955 960			
ATT CCC TGG CCG CCC CCG GAA AAA CCA GCA CTG AAT CCC TAC CCG CCT 2928 Ile Ala Trp Pro Arg Arg Glu Lys Pro Ala Val Asn Pro Tyr Pro Pro 965 970 975			

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

ATT CTG CCG GAA ACC CTA TTT GAG ACC AGT TAT GAT GAT CAA CAA TAA 1075
 Thr Leu Pro Glu Thr Leu Phe Asp Ser Ser Tyr Asp Asp Gln Gln Gln
 980 985 990

 CTA TTA CGT CTG GTG AGA CAA AAA AAT ACC TGG CAT CAC CTG ACT GAT 1024
 Leu Leu Arg Leu Val Arg Gln Lys Asn Ser Trp His His Leu Thr Asp
 995 1000 1005

 GCG GAA AAC TGG CGA TTA GGT TTA CCG AAT GCA CAA CCG GGT GAT GTT 1072
 Gly Glu Asn Trp Arg Leu Gly Leu Pro Asn Ala Gln Arg Arg Asp Val
 1010 1015 1020

 TAT ACT TAT GAC CCG ACC AAA ATT GCA ACC GAA CCG ATT TCC GTT GAA 1120
 Tyr Thr Tyr Asp Arg Ser Lys Ile Pro Thr Glu Gly Ile Ser Leu Glu
 1025 1030 1035 1040

 ATC TTG CTG AAA GAT GAT GCG CTG CTA GCA GAT GAA AAA GCG GCC GTT 1168
 Ile Leu Leu Lys Asp Asp Gly Leu Leu Ala Asp Glu Lys Ala Ala Val
 1045 1050 1055

 TAT CTG GGA CAA CAA CAG ACC TTT TAC ACC GCC GGT CAA GCG GAA GTC 1216
 Tyr Leu Gly Gln Gln Gln Thr Phe Tyr Thr Ala Gly Gln Ala Glu Val
 1060 1065 1070

 ACT CTA GAA AAA CCG ACC TTA CAA GCA CTG GTC CCG TTC CAA GAA ACC 1264
 Thr Leu Glu Lys Pro Thr Leu Gln Ala Leu Val Ala Phe Gln Glu Thr
 1075 1080 1085

 GCC ATG ATG GAC GAT ACC TCA TTA CAG CCG TAT GAA GCC GTG ATT GAA 1112
 Ala Met Met Asp Asp Thr Ser Leu Gln Ala Tyr Glu Gly Val Ile Glu
 1090 1095 1100

 CAG CAA GAG TTG AAT ACC GCG CTG ACA CAG GCC GGT TAT CAG CAA GTC 1160
 Glu Gln Glu Leu Asn Thr Ala Leu Thr Gln Ala Gly Tyr Gln Gln Val
 1105 1110 1115 1120

 CCG CCG TTG TTT AAT ACC AGA TCA GAA ACC CCG GTA TGG CCG GCA CCG 1408
 Ala Arg Leu Phe Asn Thr Arg Ser Glu Ser Pro Val Trp Ala Ala Arg
 1125 1130 1135

 CAA GGT TAT ACC GAT TAC GGT GAC GCC GCA CAG TTC TGG CCG CCT CAG 1456
 Gln Gly Tyr Thr Asp Tyr Gly Asp Ala Ala Gln Phe Trp Arg Pro Gln
 1140 1145 1150

 GCT CAG CGT AAC TCG TTG CTG ACA CCG AAA ACC ACA CTG ACC TGG GAT 1504
 Ala Gln Arg Asn Ser Leu Leu Thr Gly Lys Thr Thr Leu Thr Trp Asp
 1155 1160 1165

 ACC CAT CAT TGT GTA ATA ATA CAG ACT CAA GAT GCC GCT GGA TTA ACC 1552
 Thr His His Cys Val Ile Ile Gln Thr Gln Asp Ala Ala Gly Leu Thr
 1170 1175 1180

 ACC CAA GCC CAT TAC GAT TAT CGT TTC CTT ACA CCG GTA CAA CTG ACA 1600
 Thr Gln Ala His Tyr Asp Tyr Arg Phe Leu Thr Pro Val Gln Leu Thr
 1185 1190 1195 1200

 GAT ATT AAT GAT AAT CAA CAT ATT GTG ACT CTG GAC GCC CTA GGT CCG 1648
 Asp Ile Asn Asp Asn Gln His Ile Val Thr Leu Asp Ala Leu Gly Arg
 1205 1210 1215

 GTA ACC ACC AGC CCG TTC TGG GCG ACA GAG GCA GGA CAA GCC GCA GCG 1696
 Val Thr Thr Ser Arg Phe Trp Gly Thr Glu Ala Gly Gln Ala Ala Gly
 1220 1225 1230

 TAT TCC AAC CAG CCC TTC ACA CCA CCG GAC TCC GTA GAT AAA GCG CTG 1744
 Tyr Ser Asn Gln Pro Phe Thr Pro Pro Asp Ser Val Asp Lys Ala Leu

U A 8 2 4 8 5 C 2

1235	1240	1245
GCA TTA ACC GGC GCA CTC CTT GTT GCC CAA TGT TTA CTC TAT GCG GTT 1250	Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val 1255	
GAT ACC TGG ATG CCG TCG TTA TGT TTG TGT CAG CTT TGT CAG TCA CAA 1265	Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln 1270	1275 1280
GAA GAG GCA GAA GCG CTA TCG GCG CAA CTG CGT GCC GCT CAT ATG ATT 1285	Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile 1290	1295
ACC GAA GAT GCG AAA CTG TGT GCG TTA AGC GCG AAA CGA GGA ACA AGC 1300	Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser 1305	1310
CAT CAG AAC CTG ACG ATT CAA CTT ATT TCG CTA TTG GCA AGT ATT GCG 1315	His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro 1320	1325
CGT TTA CCG CCA CAT GTA CTG GGG ATC ACC ACT GAT CCG TAT GAT ACC 1330	Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser 1335	1340
GAT CCG CAA CAG CAG CAC CAA CAG ACG GTG ACC TTT AGT GAC GGT TTT 1345	Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe 1350	1355 1360
GGC CGG TTA CTC CAG AGT TCA GCT CGT CAT GAG TCA GGT GAT GCC TGG 1365	Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp 1370	1375
CAA CGT AAA GAG GAT GGC GCG CTG GTC CTG GAT GCA AAT GCG GTT CTC 1380	Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu 1385	1390
GTC AGT GCC CCT ACA GAC ACC CGA TGG GCC GTT TCC GGT CGC ACA GAA 1395	Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu 1400	1405
TAT GAC GAC AAA GGC CAA CCT GTG CGT ACT TAT CAA CCC TAT TTT CTA 1410	Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu 1415	1420
AAT GAC TGG CGT TAC GTT AGT GAT GAC AGC GCA CGA GAT GAC CTG TTT 1425	Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe 1430	1435 1440
GCC GAT ACC CAC CTT TAT GAT CCA TTG CGA CCG GAA TAC AAA GTC ATC 1445	Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile 1450	1455
ACT GCT AAG AAA TAT TTG CGA GAA AAG CTG TAC ACC CCG TGG TTT ATT 1460	Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile 1465	1470
GTC AGT GAG GAT GAA AAC GAT ACA GCA TCA AGA ACC CCA TAG 1475	Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro 1480	4458 1485

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 32:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 1486 амінокислот

(B) ТИП: амінокислота

(C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 32:

Met	Gln	Asp	Ser	Pro	Glu	Val	Ser	Ile	Thr	Thr	Leu	Ser	Leu	Pro	Lys
1				5					10					15	
Gly	Gly	Gly	Ala	Ile	Asn	Gly	Met	Gly	Glu	Ala	Leu	Asn	Ala	Ala	Gly
			20				25						30		
Pro	Asp	Gly	Met	Ala	Ser	Leu	Ser	Leu	Pro	Leu	Pro	Leu	Ser	Thr	Gly
		35					40					45			
Arg	Gly	Thr	Ala	Pro	Gly	Leu	Ser	Leu	Ile	Tyr	Ser	Asn	Ser	Ala	Gly
	50					55					60				
Asn	Gly	Pro	Phe	Gly	Ile	Gly	Trp	Gln	Cys	Gly	Val	Met	Ser	Ile	Ser
65				70					75						80
Arg	Arg	Thr	Gln	His	Gly	Ile	Pro	Gln	Tyr	Gly	Asn	Asp	Asp	Thr	Phe
			85						90					95	
Leu	Ser	Pro	Gln	Gly	Glu	Val	Met	Asn	Ile	Ala	Leu	Asn	Asp	Gln	Gly
			100					105					110		
Gln	Pro	Asp	Ile	Arg	Gln	Asp	Val	Lys	Thr	Leu	Gln	Gly	Val	Thr	Leu
		115					120					125			
Pro	Ile	Ser	Tyr	Thr	Val	Thr	Arg	Tyr	Gln	Ala	Arg	Gln	Ile	Leu	Asp
	130						135				140				
Phe	Ser	Lys	Ile	Glu	Tyr	Trp	Gln	Pro	Ala	Ser	Gly	Gln	Glu	Gly	Arg
145				150					155						160
Ala	Phe	Trp	Leu	Ile	Ser	Thr	Pro	Asp	Gly	His	Leu	His	Ile	Leu	Gly
			165					170					175		
Lys	Thr	Ala	Gln	Ala	Cys	Leu	Ala	Asn	Pro	Gln	Asn	Asp	Gln	Gln	Ile
			180					185					190		
Ala	Gln	Trp	Leu	Leu	Glu	Glu	Thr	Val	Thr	Pro	Ala	Gly	Glu	His	Val
		195					200					205			
Ser	Tyr	Gln	Tyr	Arg	Ala	Glu	Asp	Glu	Ala	His	Cys	Asp	Asp	Asn	Glu
	210					215					220				
Lys	Thr	Ala	His	Pro	Asn	Val	Thr	Ala	Gln	Arg	Tyr	Leu	Val	Gln	Val
225					230					235				240	
Asn	Tyr	Gly	Asn	Ile	Lys	Pro	Gln	Ala	Ser	Leu	Phe	Val	Leu	Asp	Asn
			245						250					255	
Ala	Pro	Pro	Ala	Pro	Glu	Glu	Trp	Leu	Phe	His	Leu	Val	Phe	Asp	His
			260				265						270		
Gly	Glu	Arg	Asp	Thr	Ser	Leu	His	Thr	Val	Pro	Thr	Trp	Asp	Ala	Gly
	275						280					285			
Thr	Ala	Gln	Trp	Ser	Val	Arg	Pro	Asp	Ile	Phe	Ser	Arg	Tyr	Glu	Tyr

U A 8 2 4 8 5 C 2

290	295	300
Gly Phe Glu Val Arg Thr Arg Arg Leu Cys Gln Gln Val Leu Met Phe 305 310 315 320		
His Arg Thr Ala Leu Met Ala Gly Glu Ala Ser Thr Asn Asp Ala Pro 325 330 335		
Glu Leu Val Gly Arg Leu Ile Leu Glu Tyr Asp Lys Asn Ala Ser Val 340 345 350		
Thr Thr Leu Ile Thr Ile Arg Gln Leu Ser His Glu Ser Asp Gly Arg 355 360 365		
Pro Val Thr Gln Pro Pro Leu Glu Leu Ala Trp Gln Arg Phe Asp Leu 370 375 380		
Glu Lys Ile Pro Thr Trp Gln Arg Phe Asp Ala Leu Asp Asn Phe Asn 385 390 395 400		
Ser Gln Gln Arg Tyr Gln Leu Val Asp Leu Arg Gly Glu Gly Leu Pro 405 410 415		
Gly Met Leu Tyr Gln Asp Arg Gly Ala Trp Trp Tyr Lys Ala Pro Gln 420 425 430		
Arg Gln Glu Asp Gly Asp Ser Asn Ala Val Thr Tyr Asp Lys Ile Ala 435 440 445		
Pro Leu Pro Thr Leu Pro Asn Leu Gln Asp Asn Ala Ser Leu Met Asp 450 455 460		
Ile Asn Gly Asp Gly Gln Leu Asp Trp Val Val Thr Ala Ser Gly Ile 465 470 475 480		
Arg Gly Tyr His Ser Gln Gln Pro Asp Gly Lys Trp Thr His Phe Thr 485 490 495		
Pro Ile Asn Ala Leu Pro Val Glu Tyr Phe His Pro Ser Ile Gln Phe 500 505 510		
Ala Asp Leu Thr Gly Ala Gly Leu Ser Asp Leu Val Leu Ile Gly Pro 515 520 525		
Lys Ser Val Arg Leu Tyr Ala Asn Gln Arg Asn Gly Trp Arg Lys Gly 530 535 540		
Glu Asp Val Pro Gln Ser Thr Gly Ile Thr Leu Pro Val Thr Gly Thr 545 550 555 560		
Asp Ala Arg Lys Leu Val Ala Phe Ser Asp Met Leu Gly Ser Gly Gln 565 570 575		
Gln His Leu Val Glu Ile Lys Gly Asn Arg Val Thr Cys Trp Pro Asn 580 585 590		
Leu Gly His Gly Arg Phe Gly Gln Pro Leu Thr Leu Ser Gly Phe Ser 595 600 605		
Gln Pro Glu Asn Ser Phe Asn Pro Glu Arg Leu Phe Leu Ala Asp Ile 610 615 620		
Asp Gly Ser Gly Thr Thr Asp Leu Ile Tyr Ala Gln Ser Gly Ser Leu 625 630 635 640		
Leu Ile Tyr Leu Asn Gln Ser Gly Asn Gln Phe Asp Ala Pro Leu Thr		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

645	650	655
Leu Ala Leu Pro Glu Gly Val Gln Phe Asp Asn Thr Cys Gln Leu Gln 660 665 670		
Val Ala Asp Ile Gln Gly Leu Gly Ile Ala Ser Leu Ile Leu Thr Val 675 680 685		
Pro His Ile Ala Pro His His Trp Arg Cys Asp Leu Ser Leu Thr Lys 690 695 700		
Pro Trp Leu Leu Asn Val Met Asn Asn Asn Arg Gly Ala His His Thr 705 710 715 720		
Leu His Tyr Arg Ser Ser Ala Gln Phe Trp Leu Asp Glu Lys Leu Gln 725 730 735		
Leu Thr Lys Ala Gly Lys Ser Pro Ala Cys Tyr Leu Pro Phe Pro Met 740 745 750		
His Leu Leu Trp Tyr Thr Glu Ile Gln Asp Glu Ile Ser Gly Asn Arg 755 760 765		
Leu Thr Ser Glu Val Asn Tyr Ser His Gly Val Trp Asp Gly Lys Glu 770 775 780		
Arg Glu Phe Arg Gly Phe Gly Cys Ile Lys Gln Thr Asp Thr Thr Thr 785 790 795 800		
Phe Ser His Gly Thr Ala Pro Glu Gln Ala Ala Pro Ser Leu Ser Ile 805 810 815		
Ser Trp Phe Ala Thr Gly Met Asp Glu Val Asp Ser Gln Leu Ala Thr 820 825 830		
Glu Tyr Trp Gln Ala Asp Thr Gln Ala Tyr Ser Gly Phe Glu Thr Arg 835 840 845		
Tyr Thr Val Trp Asp His Thr Asn Gln Thr Asp Gln Ala Phe Thr Pro 850 855 860		
Asn Glu Thr Gln Arg Asn Trp Leu Thr Arg Ala Leu Lys Gly Gln Leu 865 870 875 880		
Leu Arg Thr Glu Leu Tyr Gly Leu Asp Gly Thr Asp Lys Gln Thr Val 885 890 895		
Pro Tyr Thr Val Ser Glu Ser Arg Tyr Gln Val Arg Ser Ile Pro Val 900 905 910		
Asn Lys Glu Thr Glu Leu Ser Ala Trp Val Thr Ala Ile Glu Asn Arg 915 920 925		
Ser Tyr His Tyr Glu Arg Ile Ile Thr Asp Pro Gln Phe Ser Gln Ser 930 935 940		
Ile Lys Leu Gln His Asp Ile Phe Gly Gln Ser Leu Gln Ser Val Asp 945 950 955 960		
Ile Ala Trp Pro Arg Arg Glu Lys Pro Ala Val Asn Pro Tyr Pro Pro 965 970 975		
Thr Leu Pro Glu Thr Leu Phe Asp Ser Ser Tyr Asp Asp Gln Gln Gln 980 985 990		
Leu Leu Arg Leu Val Arg Gln Lys Asn Ser Trp His His Leu Thr Asp		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

935	1000	1005
Gly Glu Asn Trp Arg Leu Gly Leu Pro Asn Ala Gln Arg Arg Asp Val 1010	1015	1020
Tyr Thr Tyr Asp Arg Ser Lys Ile Pro Thr Glu Gly Ile Ser Leu Glu 1025	1030	1035 1040
Ile Leu Leu Lys Asp Asp Gly Leu Leu Ala Asp Glu Lys Ala Ala Val 1045	1050	1055
Tyr Leu Gly Gln Gln Gln Thr Phe Tyr Thr Ala Gly Gln Ala Glu Val 1060	1065	1070
Thr Leu Glu Lys Pro Thr Leu Gln Ala Leu Val Ala Phe Gln Glu Thr 1075	1080	1085
Ala Met Met Asp Asp Thr Ser Leu Gln Ala Tyr Glu Gly Val Ile Glu 1090	1095	1100
Glu Gln Glu Leu Asn Thr Ala Leu Thr Gln Ala Gly Tyr Gln Gln Val 1105	1110	1115 1120
Ala Arg Leu Phe Asn Thr Arg Ser Glu Ser Pro Val Trp Ala Ala Arg 1125	1130	1135
Gln Gly Tyr Thr Asp Tyr Gly Asp Ala Ala Gln Phe Trp Arg Pro Gln 1140	1145	1150
Ala Gln Arg Asn Ser Leu Leu Thr Gly Lys Thr Thr Leu Thr Trp Asp 1155	1160	1165
Thr His His Cys Val Ile Ile Gln Thr Gln Asp Ala Ala Gly Leu Thr 1170	1175	1180
Thr Gln Ala His Tyr Asp Tyr Arg Phe Leu Thr Pro Val Gln Leu Thr 1185	1190	1195 1200
Asp Ile Asn Asp Asn Gln His Ile Val Thr Leu Asp Ala Leu Gly Arg 1205	1210	1215
Val Thr Thr Ser Arg Phe Trp Gly Thr Glu Ala Gly Gln Ala Ala Gly 1220	1225	1230
Tyr Ser Asn Gln Pro Phe Thr Pro Pro Asp Ser Val Asp Lys Ala Leu 1235	1240	1245
Ala Leu Thr Gly Ala Leu Pro Val Ala Gln Cys Leu Val Tyr Ala Val 1250	1255	1260
Asp Ser Trp Met Pro Ser Leu Ser Leu Ser Gln Leu Ser Gln Ser Gln 1265	1270	1275 1280
Glu Glu Ala Glu Ala Leu Trp Ala Gln Leu Arg Ala Ala His Met Ile 1285	1290	1295
Thr Glu Asp Gly Lys Val Cys Ala Leu Ser Gly Lys Arg Gly Thr Ser 1300	1305	1310
His Gln Asn Leu Thr Ile Gln Leu Ile Ser Leu Leu Ala Ser Ile Pro 1315	1320	1325
Arg Leu Pro Pro His Val Leu Gly Ile Thr Thr Asp Arg Tyr Asp Ser 1330	1335	1340
Asp Pro Gln Gln Gln His Gln Gln Thr Val Ser Phe Ser Asp Gly Phe		

U A 8 2 4 8 5 C 2

1345 1350 1355 1360
 Gly Arg Leu Leu Gln Ser Ser Ala Arg His Glu Ser Gly Asp Ala Trp
 1365 1370 1375
 Gln Arg Lys Glu Asp Gly Gly Leu Val Val Asp Ala Asn Gly Val Leu
 1380 1385 1390
 Val Ser Ala Pro Thr Asp Thr Arg Trp Ala Val Ser Gly Arg Thr Glu
 1395 1400 1405
 Tyr Asp Asp Lys Gly Gln Pro Val Arg Thr Tyr Gln Pro Tyr Phe Leu
 1410 1415 1420
 Asn Asp Trp Arg Tyr Val Ser Asp Asp Ser Ala Arg Asp Asp Leu Phe
 1425 1430 1435 1440
 Ala Asp Thr His Leu Tyr Asp Pro Leu Gly Arg Glu Tyr Lys Val Ile
 1445 1450 1455
 Thr Ala Lys Lys Tyr Leu Arg Glu Lys Leu Tyr Thr Pro Trp Phe Ile
 1460 1465 1470
 Val Ser Glu Asp Glu Asn Asp Thr Ala Ser Arg Thr Pro
 1475 1480 1485

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 33:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 3288 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 25:

ATG	GTG	ACT	GTT	ATG	CAA	AAT	AAA	ATA	TCA	TTT	TTA	TCA	GGT	ACA	TCC	48
Met	Val	Thr	Val	Met	Gln	Asn	Lys	Ile	Ser	Phe	Leu	Ser	Gly	Thr	Ser	
1				5					10					15		
CAA	CAG	CCC	CTG	CTT	GAC	GCC	GGT	TAT	CAA	AAC	GTA	TTT	GAT	ATC	GCA	96
Glu	Gln	Pro	Leu	Leu	Asp	Ala	Gly	Tyr	Gln	Asn	Val	Phe	Asp	Ile	Ala	
			20					25					30			
TCA	ATC	AGC	CGG	GCT	ACT	TTC	GTT	CAA	TCC	GTT	CCC	ACC	CTG	CCC	GTT	144
Ser	Ile	Ser	Arg	Ala	Thr	Phe	Val	Gln	Ser	Val	Pro	Thr	Leu	Pro	Val	
		35				40					45					
AAA	CAG	GCT	CAT	ACC	GTC	TAT	CGT	CAG	GCG	CGG	CAA	CGT	GCG	GAA	AAT	192
Lys	Glu	Ala	His	Thr	Val	Tyr	Arg	Gln	Ala	Arg	Gln	Arg	Ala	Glu	Asn	
	50					55					60					
CTG	AAA	TCC	CTC	TAC	CGA	GCC	TGG	CAA	TTG	CGT	CAG	GAG	CCG	GTT	ATT	240
Leu	Lys	Ser	Leu	Tyr	Arg	Ala	Trp	Gln	Leu	Arg	Gln	Glu	Pro	Val	Ile	
65					70				75					80		
AAA	GCG	CTG	GCT	AAA	CTT	AAC	CTA	CAA/TCC	AAC	GTT	TCT	GTG	CTT	CAA		238
Lys	Gly	Leu	Ala	Lys	Leu	Asn	Leu/Gln	Ser	Asn	Val	Ser	Val	Leu	Gln		
			85					90					95			
GAT	GCT	TTG	GTA	GAG	AAT	ATT	GCC	GGT	GAT	GGG	GAT	TTC	ACC	GAT	TTA	336
Asp	Ala	Leu	Val	Glu	Asn	Ile	Gly	Gly	Asp	Gly	Asp	Phe	Ser	Asp	Leu	

100						105						110						
ATG	AAC	CGT	GGC	AGT	CAA	TAT	GCT	GAC	GCT	GGC	TCT	ATT	CAA	TCC	CTA	384		
Met	Asn	Arg	Ala	Ser	Gln	Tyr	Ala	Asp	Ala	Ala	Ser	Ile	Gln	Ser	Leu			
		115					120					125						
TTT	TCA	CCG	GGC	CGT	TAT	GCT	TCC	GCA	CTC	TAC	AGA	GTT	GCT	AAA	GAT	432		
Phe	Ser	Pro	Gly	Arg	Tyr	Ala	Ser	Ala	Leu	Tyr	Arg	Val	Ala	Lys	Asp			
		130				135					140							
CTG	CAT	AAA	TCA	GAT	TCC	AGT	TTG	CAT	ATT	GAT	AAT	CGC	CGC	GCT	GAT	480		
Leu	His	Lys	Ser	Asp	Ser	Ser	Leu	His	Ile	Asp	Asn	Arg	Arg	Ala	Asp			
		145				150				155				160				
CTG	AAG	GAT	CTG	ATA	TTA	AGC	GAA	ACG	ACG	ATG	AAT	AAA	GAG	GTC	ACT	528		
Leu	Lys	Asp	Leu	Ile	Leu	Ser	Glu	Thr	Thr	Met	Asn	Lys	Glu	Val	Thr			
				165					170					175				
TCC	CTT	GAT	ATC	TTG	TTG	GAT	GTG	CTA	CAA	AAA	GGC	GGT	AAA	GAT	ATT	576		
Ser	Leu	Asp	Ile	Leu	Leu	Asp	Val	Leu	Gln	Lys	Gly	Gly	Lys	Asp	Ile			
			180					185					190					
ACT	GAG	CTG	TCC	GGC	GCA	TTC	TTC	CCA	ATG	ACG	TTA	CCT	TAT	GAC	GAT	624		
Thr	Glu	Leu	Ser	Gly	Ala	Phe	Phe	Pro	Met	Thr	Leu	Pro	Tyr	Asp	Asp			
		195					200					205						
CAT	CTG	TCC	CAA	ATC	GAT	TCC	GCT	TTA	TCG	GCA	CAA	GGC	AGA	ACG	CTG	672		
His	Leu	Ser	Gln	Ile	Asp	Ser	Ala	Leu	Ser	Ala	Gln	Ala	Arg	Thr	Leu			
		210				215					220							
AAC	GGT	CTG	TGG	AAT	ACT	TTG	ACA	GAT	ACC	ACG	GCA	CAA	GGC	GTT	TCA	720		
Asn	Gly	Val	Trp	Asn	Thr	Leu	Thr	Asp	Thr	Thr	Ala	Gln	Ala	Val	Ser			
		225				230				235				240				
GAA	CAA	ACC	AGT	AAT	ACG	AAT	ACA	CCG	AAA	CTG	TTC	GCT	GGC	CAA	GAT	768		
Glu	Gln	Thr	Ser	Asn	Thr	Asn	Thr	Arg	Lys	Leu	Phe	Ala	Ala	Gln	Asp			
				245					250					255				
GGT	AAT	CAA	GAT	ACA	TTT	TTT	TCC	GGA	AAC	ACT	TTT	TAT	TTC	AAA	GGC	816		
Gly	Asn	Gln	Asp	Thr	Phe	Phe	Ser	Gly	Asn	Thr	Phe	Tyr	Phe	Lys	Ala			
			260				265						270					
GTG	GGA	TTC	AGC	GGG	CAA	CCT	ATG	GTT	TAC	CTG	TCA	CAG	TAC	ACC	AGC	864		
Val	Gly	Phe	Ser	Gly	Gln	Pro	Met	Val	Tyr	Leu	Ser	Gln	Tyr	Thr	Ser			
		275					280					285						
GGG	AAC	GGC	ATT	GTC	GGC	GCA	CAA	TTG	ATT	GCA	GGT	AAT	CCA	GAC	CAA	912		
Gly	Asn	Gly	Ile	Val	Gly	Ala	Gln	Leu	Ile	Ala	Gly	Asn	Pro	Asp	Gln			
		290				295					300							
GGC	GGC	GGC	GCA	ATA	GTC	GCA	CCG	TTG	AAA	CTC	ACT	TGG	TCA	ATC	GCA	960		
Ala	Ala	Ala	Ala	Ile	Val	Ala	Pro	Leu	Lys	Leu	Thr	Trp	Ser	Met	Ala			
		305				310				315				320				
AAA	CAG	TGT	TAC	TAC	CTC	GTC	GCT	CCC	GAT	GGT	ACA	ACG	ATG	GGA	GAC	1008		
Lys	Gln	Cys	Tyr	Tyr	Leu	Val	Ala	Pro	Asp	Gly	Thr	Thr	Met	Gly	Asp			
				325					330					335				
GGT	AAT	GTT	CTG	ACC	GGC	TGT	TTC	TTA	AGA	GGC	AAC	AGC	CCA	ACT	AAC	1056		
Gly	Asn	Val	Leu	Thr	Gly	Cys	Phe	Leu	Arg	Gly	Asn	Ser	Pro	Thr	Asn			
			340					345					350					
CCG	GAT	AAA	GAC	GGT	ATT	TTT	GCT	CAG	GTA	GCC	AAC	AAA	TCA	GGC	AGT	1104		
Pro	Asp	Lys	Asp	Gly	Ile	Phe	Ala	Gln	Val	Ala	Asn	Lys	Ser	Gly	Ser			
		355					360					365						

ACT CAG CCT TTG CCA AGC TTC CAT CTG CCG GTC ACA CTG GAA CAC AGC	1152
Thr Gln Pro Leu Pro Ser Phe His Leu Pro Val Thr Leu Glu His Ser	
370 375 380	
GAG AAT AAA GAT CAG TAC TAT CTG AAA ACA GAG CAG GGT TAT ATC ACG	1200
Glu Asn Lys Asp Gln Tyr Tyr Leu Lys Thr Glu Gln Gly Tyr Ile Thr	
385 390 395 400	
GTA GAT AGT TCC GGA CAG TCA AAT TCG AAA AAC GCG CTG GTT ATC AAT	1248
Val Asp Ser Ser Gly Gln Ser Asn Trp Lys Asn Ala Leu Val Ile Asn	
405 410 415	
GGG ACA AAA GAC AAG GGG CTG TTA TTA ACC TTT TGC AGC GAT AGC TCA	1296
Gly Thr Lys Asp Lys Gly Leu Leu Leu Thr Phe Cys Ser Asp Ser Ser	
420 425 430	
GGC ACT CCG ACA AAC CCT GAT GAT GTG ATT CCT CCC GCT ATC AAT GAT	1344
Gly Thr Pro Thr Asn Pro Asp Asp Val Ile Pro Pro Ala Ile Asn Asp	
435 440 445	
ATT CCA TCG CCG CCA GCC CGC GAA ACA CTG TCA CTG ACG CCG GTC AGT	1392
Ile Pro Ser Pro Pro Ala Arg Glu Thr Leu Ser Leu Thr Pro Val Ser	
450 455 460	
TAT CAA TTG ATG ACC AAT CCG GCA CCG ACA GAA GAT GAT ATT ACC AAC	1440
Tyr Gln Leu Met Thr Asn Pro Ala Pro Thr Glu Asp Asp Ile Thr Asn	
465 470 475 480	
CAT TAT GGT TTT AAC GGC GCT AGC TTA CCG GCT TCT CCA TTG TCA ACC	1488
His Tyr Gly Phe Asn Gly Ala Ser Leu Arg Ala Ser Pro Leu Ser Thr	
485 490 495	
AGC GAG TTG ACC AGC AAA CTG AAT TCT ATC GAT ACT TTC TGT GAG AAG	1536
Ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr Phe Cys Glu Lys	
500 505 510	
ACC CCG TTA AGC TTC AAT CAG TTA ATG GAT TTG ACC GCT CAG CAA TCT	1584
Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr Ala Gln Gln Ser	
515 520 525	
TAC AGT CAA AGC AGC ATT GAT GCG AAA GCA GCC AGC CGC TAT GTT CGT	1632
Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser Arg Tyr Val Arg	
530 535 540	
TTT GGG GAA ACC ACC CCA ACC CGC GTC AAT GTC TAC GGT GCC GCT TAT	1680
Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr Gly Ala Ala Tyr	
545 550 555 560	
CTG AAC AGC ACA CTG GCA GAC GCG GCT GAT GGT CAA TAT CTG TGG ATT	1728
Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln Tyr Leu Trp Ile	
565 570 575	
CAG ACT GAT GGC AAG AGC CTA AAT TTC ACT GAC GAT ACG GTA GTC GCC	1776
Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp Thr Val Val Ala	
580 585 590	
TTA GCC GGT CGC GCT GAA AAG CTG GTA CGT TTA TCA TCC CAG ACC GGG	1824
Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser Ser Gln Thr Gly	
595 600 605	
CTA TCA TTT GAA GAA TTG GAC TGG CTG ATT GCC AAT GCC AGT CGT AGT	1872
Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn Ala Ser Arg Ser	
610 615 620	
GTG CCG GAC CAC CAC GAC AAA ATT GTG CTG GAT AAG CCG GTC CTT GAA	1920
Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys Pro Val Leu Glu	

525	530	535	540	
GCA CTG GCA GAG TAT GTC AGC CTA AAA CAG CGC TAT GCG GTT GAT GCG	Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala	1953		
645	650		655	
AAT ACC TTT GCG ACC TTC ATT AGT GCA GTA AAT CCT TAT ACG CCA GAT	Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp	2015		
660	665		670	
CAG ACA CCC AGT TTC TAT GAA ACC GGT TTC CGC TCT GCG GAC GGT AAT	Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn	2064		
675	680		685	
CAT CTC ATT GCG CTA GGT ACA CAG GTG AAA TAT GCA GAA AAT GAG CAG	His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln	2112		
690	695		700	
GAT GAG TTA GCG GCC ATA TGC TGC AAA GCA TTG GGT GTC ACC AGT GAT	Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly Val Thr Ser Asp	2160		
705	710		715	720
GAA CTG CTC CGT ATT GGT CGC TAT TGC TTC GGT AAT GCA GGC AGT TTT	Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn Ala Gly Ser Phe	2208		
725	730		735	
ACC TTG GAT GAA TAT ACC GCC AGT CAG TTG TAT CGC TTC GCG GCC ATT	Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg Phe Gly Ala Ile	2256		
740	745		750	
CCC CGT TTG TTT GCG CTG ACA TTT GCC CAA GCC GAA ATT TTA TGG CGT	Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu Ile Leu Trp Arg	2304		
755	760		765	
CTG ATG GAA GGC GGA AAA GAT ATC TTA TTG CAA CAG TTA GGT CAG GCA	Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln Leu Gly Gln Ala	2352		
770	775		780	
AAA TCC CTG CAA CCA CTC GCT ATT TTA CGC CGT ACC GAG CAG CTG CTG	Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr Glu Gln Val Leu	2400		
785	790		795	800
GAT TCG ATG TCG TCC GTA AAT CTA AGT CTG ACT TAT CTG CAA GCG ATG	Asp Trp Met Ser Ser Val Asn Leu Ser Leu Thr Tyr Leu Gln Gly Met	2448		
805	810		815	
GTA AGT ACG CAA TGG AGC GGT ACC GCC ACC GCT CAG ATG TTC AAT TTC	Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu Met Phe Asn Phe	2496		
820	825		830	
TTG GAA AAC GTT TGT GAC AGC GTG AAT AGT CAA GCT GCC ACT AAA GAA	Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala Ala Thr Lys Glu	2544		
835	840		845	
ACA ATG GAT TCG GCG TTA CAG CAG AAA GTG CTG CCG GCG CTA AGC GCC	Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg Ala Leu Ser Ala	2592		
850	855		860	
GGT TTC GCG ATT AAG AGC AAT GTG ATG GGT ATC CTC ACC TTC TGG CTG	Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val Thr Phe Trp Leu	2640		
865	870		875	880
GAG AAA ATC ACA ATC GGT AGT GAT AAT CCT TTT ACA TTG GCA AAC TAC	Glu Lys Ile Thr Ile Gly Ser Asp Asn Pro Phe Thr Leu Ala Asn Tyr	2688		
885	890		895	

TGG GAT GAT ATT TAA ACC CTG TTT ACC GAT GAC AAT GCC ACC TTA GAG 2714
 Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn Ala Thr Leu Glu
 900 905 910
 TTC TTA CAA ACC GAC ACT TCT CTG GTA ATT GCT ACT CAG CAA GTT ACC 2731
 Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr Gln Gln Leu Ser
 915 920 925
 CAG CTA GTG TTA ATT GTG AAA TGG CTG AGC CTG ACC GAG CAG GAT CTG 2832
 Gln Leu Val Leu Ile Val Lys Trp Leu Ser Leu Thr Glu Gln Asp Leu
 930 935 940
 CAA TTA CTG ACA ACC TAT CCC GAA CGT TTA ATC AAC GCC ATC ACC AAT 2889
 Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn Gly Ile Thr Asn
 945 950 955 960
 GTT CCT GTA CCC AAT CCG GAG CTA TTA CTC ACC CTA TCA CGT TTT AAG 2923
 Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu Ser Arg Phe Lys
 965 970 975
 CAG TGG GAA ACT CAA GTC ACC GTT TCC CGT GAT GAA GCG ATC CGC TGT 2976
 Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu Ala Met Arg Cys
 980 985 990
 TTC GAT CAA TTA AAT GCC AAT GAT ATG ACC ACT GAA AAT GCA CGT TCA 3024
 Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu Asn Ala Gly Ser
 995 1000 1005
 CTG ATC GCC ACA TTG TAT GAG ATG GAT AAA CGT ACC GGA GCG CAA GTT 3072
 Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr Gly Ala Gln Val
 1010 1015 1020
 AAT ACC TTG CTA TTA GGT GAA AAT AAC TGG CCG AAA AGT TTT ACC TCT 3120
 Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys Ser Phe Thr Ser
 1025 1030 1035 1040
 CTC TGG CAA CTT CTG ACC TGG TTA CGC GTC GCG CAA AGA CTG AAT GTC 3163
 Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln Arg Leu Asn Val
 1045 1050 1055
 GGT AGT ACC ACT CTG GCC AAT CTG TTG TCC ATG ATG CAA GCA GAC CCT 3216
 Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met Gln Ala Asp Pro
 1060 1065 1070
 GCT GCC GAG ACT AGC GCT TTA TTG GCA TCA GTA GCC CAA AAC TTA AGT 3264
 Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala Gln Asn Leu Ser
 1075 1080 1085
 GCC GCA ATC AGC AAT CGT CAG TAA 3285
 Ala Ala Ile Ser Asn Arg Gln ...
 1090 1095

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 34:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1095 амінокислот
- (B) ТИП: амінокислота
- (C) ТОПОЛОГІЯ: лінійна

(11) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 1:

Features	From	To	Description
	254	267	SEQ ID NO:15
	254	492	Tcaa11 peptide

U A 8 2 4 8 5 C 2

Met Val Thr Val Met Gln Asn Lys Ile Ser Phe Leu Ser Gly Thr Ser
1 5 10 15

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

Ser Ile Ser Arg Ala Thr Phe Val Gln Ser Val Pro Thr Leu Pro Val
35 40 45

Lys Glu Ala His Thr Val Tyr Arg Gln Ala Arg Gln Arg Ala Glu Asn
50 55 60

Leu Lys Ser Leu Tyr Arg Ala Trp Gln Leu Arg Gln Glu Pro Val Ile
65 70 75 80

Lys Gly Leu Ala Lys Leu Asn Leu Gln Ser Asn Val Ser Val Leu Gln
85 90 95

Asp Ala Leu Val Glu Asn Ile Gly Gly Asp Gly Asp Phe Ser Asp Leu
100 105 110

Met Asn Arg Ala Ser Gln Tyr Ala Asp Ala Ala Ser Ile Gln Ser Leu
115 120 125

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

Leu His Lys Ser Asp Ser Ser Leu His Ile Asp Asn Arg Arg Ala Asp
145 150 155 160

Leu Lys Asp Leu Ile Leu Ser Glu Thr Thr Met Asn Lys Glu Val Thr
165 170 175

Ser Leu Asp Ile Leu Leu Asp Val Leu Gln Lys Gly Gly Lys Asp Ile
180 185 190

Thr Glu Leu Ser Gly Ala Phe Phe Pro Met Thr Leu Pro Tyr Asp Asp
195 200 205

His Leu Ser Gln Ile Asp Ser Ala Leu Ser Ala Gln Ala Arg Thr Leu
210 215 220

Asn Gly Val Trp Asn Thr Leu Thr Asp Thr Thr Ala Gln Ala Val Ser
225 230 235 240

Glu Gln Thr Ser Asn Thr Asn Thr Arg Lys Leu Phe Ala Ala Gln Asp
245 250 255

Gly Asn Gln Asp Thr Phe Phe Ser Gly Asn Thr Phe Tyr Phe Lys Ala
260 265 270

Val Gly Phe Ser Gly Gln Pro Met Val Tyr Leu Ser Gln Tyr Thr Ser
275 280 285

Gly Asn Gly Ile Val Gly Ala Gln Leu Ile Ala Gly Asn Pro Asp Gln
290 295 300

Ala Ala Ala Ala Ile Val Ala Pro Leu Lys Leu Thr Trp Ser Met Ala
305 310 315 320

Lys Gln Cys Tyr Tyr Leu Val Ala Pro Asp Gly Thr Thr Met Gly Asp
325 330 335

Gly Asn Val Leu Thr Gly Cys Phe Leu Arg Gly Asn Ser Pro Thr Asn

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

340	345	350
Pro Asp Lys Asp Gly Ile Phe Ala Gln Val Ala Asn Lys Ser Gly Ser 355 360 365		
Thr Gln Pro Leu Pro Ser Phe His Leu Pro Val Thr Leu Glu His Ser 370 375 380		
Glu Asn Lys Asp Gln Tyr Tyr Leu Lys Thr Glu Gln Gly Tyr Ile Thr 385 390 395 400		
Val Asp Ser Ser Gly Gln Ser Asn Trp Lys Asn Ala Leu Val Ile Asn 405 410 415		
Gly Thr Lys Asp Lys Gly Leu Leu Leu Thr Phe Cys Ser Asp Ser Ser 420 425 430		
Gly Thr Pro Thr Asn Pro Asp Asp Val Ile Pro Pro Ala Ile Asn Asp 435 440 445		
Ile Pro Ser Pro Pro Ala Arg Glu Thr Leu Ser Leu Thr Pro Val Ser 450 455 460		
Tyr Gln Leu Met Thr Asn Pro Ala Pro Thr Glu Asp Asp Ile Thr Asn 465 470 475 480		
His Tyr Gly Phe Asn Gly Ala Ser Leu Arg Ala Ser Pro Leu Ser Thr 485 490 W4 > 495		
Ser Glu Leu Thr Ser Lys Leu Asn Ser Ile Asp Thr Phe Cys Glu Lys 500 505 510		
Thr Arg Leu Ser Phe Asn Gln Leu Met Asp Leu Thr Ala Gln Gln Ser 515 520 525		
Tyr Ser Gln Ser Ser Ile Asp Ala Lys Ala Ala Ser Arg Tyr Val Arg 530 535 540		
Phe Gly Glu Thr Thr Pro Thr Arg Val Asn Val Tyr Gly Ala Ala Tyr 545 550 555 560		
Leu Asn Ser Thr Leu Ala Asp Ala Ala Asp Gly Gln Tyr Leu Trp Ile 565 570 575		
Gln Thr Asp Gly Lys Ser Leu Asn Phe Thr Asp Asp Thr Val Val Ala 580 585 590		
Leu Ala Gly Arg Ala Glu Lys Leu Val Arg Leu Ser Ser Gln Thr Gly 595 600 605		
Leu Ser Phe Glu Glu Leu Asp Trp Leu Ile Ala Asn Ala Ser Arg Ser 610 615 620		
Val Pro Asp His His Asp Lys Ile Val Leu Asp Lys Pro Val Leu Glu 625 630 635 640		
Ala Leu Ala Glu Tyr Val Ser Leu Lys Gln Arg Tyr Gly Leu Asp Ala 645 650 655		
Asn Thr Phe Ala Thr Phe Ile Ser Ala Val Asn Pro Tyr Thr Pro Asp 660 665 670		
Gln Thr Pro Ser Phe Tyr Glu Thr Ala Phe Arg Ser Ala Asp Gly Asn 675 680 685		
His Val Ile Ala Leu Gly Thr Glu Val Lys Tyr Ala Glu Asn Glu Gln		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

690	695	700
Asp Glu Leu Ala Ala Ile Cys Cys Lys Ala Leu Gly Val Thr Ser Asp 705 710 715 720		
Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn Ala Gly Ser Phe 725 730 735		
Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg Phe Gly Ala Ile 740 745 750		
Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu Ile Leu Trp Arg 755 760 765		
Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln Leu Gly Gln Ala 770 775 780		
Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr Glu Gln Val Leu 785 790 795 800		
Asp Trp Met Ser Ser Val Asn Leu Ser Leu Thr Tyr Leu Gln Gly Met 805 810 815		
Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu Met Phe Asn Phe 820 825 830		
Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala Ala Thr Lys Glu 835 840 845		
Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg Ala Leu Ser Ala 850 855 860		
Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val Thr Phe Trp Leu 865 870 875 880		
Glu Lys Ile Thr Ile Gly Ser Asp Asn Pro Phe Thr Leu Ala Asn Tyr 885 890 895		
Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn Ala Thr Leu Glu 900 905 910		
Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr Gln Gln Leu Ser 915 920 925		
Gln Leu Val Leu Ile Val Lys Trp Leu Ser Leu Thr Glu Gln Asp Leu 930 935 940		
Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn Gly Ile Thr Asn 945 950 955 960		
Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu Ser Arg Phe Lys 965 970 975		
Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu Ala Met Arg Cys 980 985 990		
Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu Asn Ala Gly Ser 995 1000 1005		
Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr Gly Ala Gln Val 1010 1015 1020		
Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys Ser Phe Thr Ser 1025 1030 1035 1040		
Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln Arg Leu Asn Val		

U A 8 2 4 8 5 C 2

1045

1050

1055

Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met Gln Ala Asp Pro
1060 1065 1070

Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala Gln Asn Leu Ser
1075 1080 1085

Ala Ala Ile Ser Asn Arg Gln ...
1090 1095

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 35:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 603 амінокислоти

(B) ТИП: амінокислота

(C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 35:

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

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Val Thr Ser Asp Glu Leu Leu Arg Ile Gly Arg Tyr Cys Phe Gly Asn
225 230 235 240

Ala Gly Arg Phe Thr Leu Asp Glu Tyr Thr Ala Ser Gln Leu Tyr Arg
245 250 255

Phe Gly Ala Ile Pro Arg Leu Phe Gly Leu Thr Phe Ala Gln Ala Glu
260 265 270

) Ile Leu Trp Arg Leu Met Glu Gly Gly Lys Asp Ile Leu Leu Gln Gln
275 280 285

Xxx Gly Gln Ala Lys Ser Leu Gln Pro Leu Ala Ile Leu Arg Arg Thr
290 295 300

{ Glu Gln Val Leu Asp Trp Met Ser Pro Val Asn Leu Ser Leu Thr Tyr
305 310 315 320

) Leu Gln Gly Met Val Ser Thr Gln Trp Ser Gly Thr Ala Thr Ala Glu
325 330 335

Met Phe Asn Phe Leu Glu Asn Val Cys Asp Ser Val Asn Ser Gln Ala
340 345 350

{ Xxx Thr Lys Glu Thr Met Asp Ser Ala Leu Gln Gln Lys Val Leu Arg
355 360 365

Ala Leu Ser Ala Gly Phe Gly Ile Lys Ser Asn Val Met Gly Ile Val
370 375 380

Thr Phe Trp Leu Glu Lys Ile Thr Ile Gly Arg Asp Asn Pro Phe Thr
385 390 395 400

Leu Ala Asn Tyr Trp His Asp Ile Gln Thr Leu Phe Ser His Asp Asn
405 410 415

Ala Thr Leu Glu Ser Leu Gln Thr Asp Thr Ser Leu Val Ile Ala Thr
420 425 430

Gln Gln Leu Ser Gln Leu Val Leu Ile Val Lys Trp Val Ser Leu Thr
435 440 445

Glu Gln Asp Leu Gln Leu Leu Thr Thr Tyr Pro Glu Arg Leu Ile Asn
450 455 460

Gly Ile Thr Asn Val Pro Val Pro Asn Pro Glu Leu Leu Leu Thr Leu
465 470 475 480

Ser Arg Phe Lys Gln Trp Glu Thr Gln Val Thr Val Ser Arg Asp Glu
485 490 495

Ala Met Arg Cys Phe Asp Gln Leu Asn Ala Asn Asp Met Thr Thr Glu
500 505 510

Asn Ala Gly Ser Leu Ile Ala Thr Leu Tyr Glu Met Asp Lys Gly Thr
515 520 525

Gly Ala Gln Val Asn Thr Leu Leu Leu Gly Glu Asn Asn Trp Pro Lys
530 535 540

Ser Phe Thr Ser Leu Trp Gln Leu Leu Thr Trp Leu Arg Val Gly Gln
545 550 555 560

Arg Leu Asn Val Gly Ser Thr Thr Leu Gly Asn Leu Leu Ser Met Met
565 570 575

U A 8 2 4 8 5 C 2

Gln Ala Asp Pro Ala Ala Glu Ser Ser Ala Leu Leu Ala Ser Val Ala
580 585 590

Gln Asn Leu Ser Ala Ala Ile Ser Asn Arg Gln *
595 600

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 36:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 2557 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 36:

```
GAATTCGGCT TCGGTTTAAT ATTGATGATG TCTCGGCTCTT CCGGCTGCTT AAAATTACCG 60
ACCATGATAA TAAAGATGGA AAAATTAAAA ATAACCTAAA GAATCTTTCC AATTTATATA 120
TTGGAAAAAT ACTGGCAGAT ATTCAATCAAT TAACCATGCA TGAACCTGGAT TTATTACTGA 180
TTGGCGTAGG TGAAGGAAAA ACTAATTTAT CCGCTATCAG TGATAAGCAA TTGGCTAGCC 240
TGATCAGAAA ACTCAATACT ATTACCAGCT CGCTACATAC ACAGAAGTGG AGTGATTTC 300
AGGTATTTAT CATGACCTCC ACCAGCTATA AAAAAACGCT AACGGCTGAA ATTAAGAATT 360
TGGTGGATAC CGTCTACCAC GGTTTACAAG GTTTTGATAA AGACAAAGCA CATTTGCTAC 420
ATGTCATGGC GCGCTATATT GCGGCGACCT TGCAATTATC ATGCGAAAAAT GTCGGCTGCT 480
CGGTACTGCT TTGGCCAGAT AAGTTACAGC CCGGCGACCG CGCAATGACA GCAGAGGGAN 540
TCTGGGACTG GTTGAATACT AAGTATACCG CCGGTTGATC GGAAGCCGTA GAAACCGAG 600
AACATATGCT TCAGTATTCT CAGGCTCTCG CACAATTGGA AATGGTTTAC CATTCGAGCG 660
GCATCAACGA AAACGCTTTC CGTCTATTTC TGACAAAACC AGAGATGTTT GCGGCTGCAA 720
CTGGAGCAGC GCGGCGCAT CATGCGCTTT CACTGATTAT GCTGACACGT TTTGGCGATT 780
GGGTGAACGC ACTAGCGGAA AAAGCGTCTT CGGTGCTAGC GGCATTGAA GCTAAGTCTT 840
TAACCGCAGA ACAACTGGCT CATGCCATGA ATCTTGATCG TAATTTGCTG TTGCAAGCCA 900
GTATTCAAGC AAAAAATCAT CAACATCTTC CCCCAGTAAC TCCAGAAAAT GCGTTGCTCT 960
GTTCCAGATC TATCAATACT ATCTGCAAT GGGTTAATGT CCGCACAACA TTGAAATGTC 1020
GCGCCACAGG GCGTTTCCGC TTTGGTCCGG CTGGATTATA TTCAATCAAT GAAACAGACA 1080
CCGACCTATC CCGAOTGGA AAACGCGGCA GCGCTATTAA CCGCGCGGTT GAATTCACCA 1140
ACAGGCTAAT ACATTACAAC GCTTTTCTGG ATGAATCTCG CAGTCCCGCA TTAAGCACCT 1200
ACTATATCCG TCAAGTCCGC AAGGCAGCGG CCGCTATTAA AAGCCGTGAT GACTGTATC 1260
AATACTTACT GATTGATAAT CAGGTTTCTG CCGCAATAAA AACCAACCGG ATCGCGGAAG 1320
CCATTGCCAG TATTCAACTG TACGTCAACC GGCATTGGA AAATGTGGA GAAATGCCA 1380
ATTCCGGGCT TATCAGCGCG CAATTCTTTA TCGACTGGGA CAAATACAAT AAACCTACA 1440
GCACTTGGCG GGGTGTTCCT CAATTAGTTT ACTACCGGA AAATATATT CATCGGACCA 1500
TGGGTATCGG ACAAACCAA ATGATGGAGC CATTAAGTCA ATCGCTCAGC CAAAGCGAAT 1560
TAAACCGCGA TACCCTGGA CATGCCCTTA TGTCTTATCT GACATCGTTT CAACAGTGG 1620
CTAATCTTAA AGTTATTAGC GCATATCAGC ATAATATTA TAACGATCAA GCGCTCAGCT 1680
ATTTTATCGG ACTCACTGAA ACTGATGCGG GTCAATTTA TTGGCGGACT CTCGATCACA 1740
GTAAATTCAA CGACCGTAAA TTGGCGGCTA ATGCGCTCAG TGAATGGCAT AAAATTGATT 1800
GTCCAATTAA CCGTTATAAA ACCACTATCC GTCCAGTCA ATATAAATCC CCGCTGTATC 1860
```

TCCTCTCGTT CGAACAAAAAG GAGATCACCA AACAGACAGG AAATAGTAAA SATCGCTATC 1920
 AAATCGAAAC CGATTATCGT TATGAATAA AATTGGCCCA TATCCGCTAT SATGGCACTT 1960
 GGAATACGCC AATCACCTTT CATGTCAATA AAAAAATATC CGAGCTAAAA CTGCAAAAAA 2040
 ATAGAGCGCC CGGACTCTAT TGTGCGCGTT ATCAAGGTGA AGATACGTTG CTGCTGATGT 2100
 TTTATAACCA ACAAGACACA CTAGATAGTT ATAAAAACGC TTCAATGCCA GCACTATATA 2160
 TCTTTGCTGA TATGGCATCC AAAGATATGA CCCCAGAACA GAGCAATGTT TATCGGATA 2220
 ATAGCTATCA ACAATTTGAT ACCAATAATG TCAGAAGAGT GAATAACCGC TATGCAGAGG 2280
 ATTATGAGAT TCCTTCTTCG GTAAGTAGCC GTAAAGACTA TGGTTGGCGA GATTATTACC 2340
 TCAGCATGCT ATATAACCGA GATATTCCAA CTATCAATTA CAAAGCCGCA TCAAGTGATT 2400
 TAAAAATTTA TATTTCACCA AAATTAAGAA TTATTTCATA TGGATATGAA CGACAGAAGC 2460
 GCAATCAATG CAATTTGATG AATAAATATG GCAAACTAGG TGATAAATTT ATTGTGTATA 2520
 CCAGCCTGCC CGTTAATCCG AATAATAAGC CGAATTC 2557

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 37:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 845 амінокислот
- (B) ТИП: амінокислоти
- (C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ:

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 37:

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

U A 8 2 4 8 5 C 2

180	185	190
Val Glu Thr Gln Glu His Ile	Val Gln Tyr Cys Gln	Ala Leu Ala Gln
195	200	205
Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe Arg		
210	215	220
Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala Ala		
225	230	235
Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala Asp		
245	250	255
Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala Phe		
260	265	270
Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn Leu		
275	280	285
Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His Gln		
290	295	300
His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr Ser		
305	310	315
Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Lys Cys		
325	330	335
Arg Pro Thr Gly Arg Phe Arg Phe Gly Arg Ala Gly Leu Tyr Ser Ile		
340	345	350
Asn Glu Arg Asp Thr Asp Leu Cys Pro Val Gly Lys Arg Gly Arg Arg		
355	360	365
Ile Asn Arg Arg Val Glu Phe Asn Asn Arg Leu Ile His Tyr Asn Ala		
370	375	380
Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg		
385	390	395
Gln Val Ala Lys Ala Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr		
405	410	415
Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr		
420	425	430
Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala		
435	440	445
Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln		
450	455	460
Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala		
465	470	475
Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr		
485	490	495
Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val		
500	505	510
Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser		
515	520	525
Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala		

U A 8 2 4 8 5 C 2

530	535	540
Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly 545 550 555 560		
Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His 565 570 575		
Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp 580 585 590		
His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro 595 600 605		
Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu 610 615 620		
Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr 625 630 635 640		
Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr 645 650 655		
Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu 660 665 670		
Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln 675 680 685		
Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu 690 695 700		
Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp 705 710 715 720		
Met Ala Ser Lys Asp Met Thr Pro Glu Gln Ser Asn Val Tyr Arg Asp 725 730 735		
Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn 740 745 750		
Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Ser Ser Arg Lys 755 760 765		
Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp 770 775 780		
Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr 785 790 795 800		
Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys 805 810 815		
Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys 820 825 830		
Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn 835 840 845		

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 38:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 16 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один

(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 38:

Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly
1 5 10 15
Lys

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 39:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 20 амінокислот

(B) ТИП: амінокислота

(C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один

(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 39:

Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala
1 5 10 15
Ile Ser Pro Ala Lys
20

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 40:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 11 амінокислот

(B) ТИП: амінокислота

(C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один

(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 40:

Ala Asn Ser Leu Tyr Ala Leu Phe Leu Pro Gln
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 41:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

(A) ДОВЖИНА: 14 амінокислот

(B) ТИП: амінокислота

(C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один

(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 41:

Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 42:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 19 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 42:

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

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 43:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 11 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 43:

Ile Arg Glu Asp Tyr Pro Ala Ser Leu Gly Lys
1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 44:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 16 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 44:

Asp Asp Ser Gly Asp Asp Asp Lys Val Thr Asn Thr Asp Ile His
 1 5 10 15
 Arg

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 45:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 13 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 45:

Asp Val Xaa Gly Ser Glu Lys Ala Asn Glu Lys Leu Lys
 1 5 10

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 46:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 7551 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 25:

ATG AAC CAG TCT GTA AAA CAG ATA CCT GAT GTA TTA AAA ACC CAG TGT	48
Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys	
1 5 10 15	
GGT TTT AAT TGT CTG ACA GAT ATT AGC CAC AGC TCT TTT AAT GAA TTT	96
Gly Phe Asn Cys Leu Thr Asp Ile Ser His Ser Ser Phe Asn Glu Phe	
20 25 30	
CGC CAG CAA GTA TCT GAG CAC CTC TCC TCG TCC GAA ACA CAC GAC TTA	144
Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu	
35 40 45	
TAT CAT GAT GCA CAA CAG GCA CAA AAG GAT AAT CGC CTG TAT GAA CCG	192
Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala	
50 55 60	
CGT ATT CTC AAA CGC GCC AAT CCC CAA TTA CAA AAT GCG GTG CAT CTT	240
Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu	
65 70 75 80	

GGC ATT CTC GCT CCC AAT GGT GAA CTG ATA GGC TAT AAC AAT CAA TTT 288
 Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe 85 90 95
 AGC GGT AGA GCC AGT CAA TAT GTT GCG CCG GGT ACC GTT TCT TCC ATG 336
 Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met 100 105 110
 TTC TCC CCC GCG GCT TAT TTG ACT GAA CTT TAT CGT GAA GCA CCC AAT 384
 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn 115 120 125
 TTA CAC GCA AGT GAC TCC GTT TAT TAT CTG GAT ACC CCG CCG CCA GAT 432
 Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp 130 135 140
 CTC AAA TCA ATG GCG CTC AGT CAG GAA AAT ATG GAT ATA GAA TTA TCC 480
 Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser 145 150 155 160
 ACA CTC TCT TTG TCC AAT GAG CTG TTA TTG GAA AGC ATT AAA ACT GAA 528
 Thr Leu Ser Leu Ser Asn Glu Leu Leu Leu Glu Ser Ile Lys Thr Glu 165 170 175
 TCT AAA CTG GAA AAC TAT ACT AAA GTG ATG GAA ATG CTC TCC ACT TTC 576
 Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe 180 185 190
 CGT CCT TCC GCG GCA ACC CCT TAT CAT GAT GCT TAT GAA AAT GTG CGT 624
 Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg 195 200 205
 GAA GTT ATC CAG CTA CAA GAT CCT GGA CTT GAG CAA CTC AAT GCA TCA 672
 Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser 210 215 220
 CCG GCA ATT GCC GCG TTG ATG CAT CAA GCC TCC CTA TTG GGT ATT AAC 720
 Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn 225 230 235 240
 GGT TCA ATC TCG CCT GAG CTA TTT AAT ATT CTG ACC GAG GAG ATT ACC 768
 Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr 245 250 255
 GAA GGT AAT GCT GAG GAA CTT TAT AAG AAA AAT TTT GGT AAT ATC GAA 816
 Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu 260 265 270
 CCG GCC TCA TTG GCT ATG CCG GAA TAC CTT AAA GGT TAT TAT AAT TTA 864
 Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu 275 280 285
 AGC GAT GAA GAA CTT AGT CAG TTT ATT GGT AAA GCC AGC AAT TTT GGT 912
 Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly 290 295 300
 CAA CAG GAA TAT AGT AAT AAC CAA CTT ATT ACT CCG GTA CTC AAC AGC 960
 Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser 305 310 315 320
 AGT GAT GGC ACG GTT AAG GTA TAT CCG ATC ACC CCG GAA TAT ACA ACC 1008
 Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr 325 330 335
 AAT GCT TAT CAA ATG GAT CTG GAG CTA TTT CCC TTC GGT GGT GAG AAT 1056
 Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn 340 345 350
 TAT CCG TTA GAT TAT AAA TTC AAA AAT TTT TAT AAT GCC TCT TAT TTA 1104
 Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu 355 360 365

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

TCC	ATC	AAG	TTA	AAT	GAT	AAA	AGA	GAA	CTT	GTT	CGA	ACT	GAA	GCC	GCT	1152
Ser	Ile	Lys	Leu	Asn	Asp	Lys	Arg	Glu	Leu	Val	Arg	Thr	Glu	Gly	Ala	
370						375					380					
CCT	CAA	GTC	AAT	ATA	GAA	TAC	TCC	GCA	AAT	ATC	ACA	TTA	AAT	ACC	GCT	1200
Pro	Gln	Val	Asn	Ile	Glu	Tyr	Ser	Ala	Asn	Ile	Thr	Leu	Asn	Thr	Ala	
385					390					395					400	
GAT	ATC	AGT	CAA	CCT	TTT	GAA	ATT	GGC	CTG	ACA	CGA	GTA	CTT	CCT	TCC	1248
Asp	Ile	Ser	Gln	Pro	Phe	Glu	Ile	Gly	Leu	Thr	Arg	Val	Leu	Pro	Ser	
				405				410						415		
GGT	TCT	TGG	GCA	TAT	GCC	GCC	GCA	AAA	TTT	ACC	GTT	GAA	GAG	TAT	AAC	1296
Gly	Ser	Trp	Ala	Tyr	Ala	Ala	Ala	Lys	Phe	Thr	Val	Glu	Glu	Tyr	Asn	
			420					425					430			
CAA	TAC	TCT	TTT	CTG	CTA	AAA	CTT	AAC	AAG	GCT	ATT	CGT	CTA	TCA	CGT	1344
Gln	Tyr	Ser	Phe	Leu	Leu	Lys	Leu	Asn	Lys	Ala	Ile	Arg	Leu	Ser	Arg	
		435					440					445				
CCG	ACA	GAA	TTG	TCA	CCC	ACG	ATT	CTG	GAA	GGC	ATT	GTG	CGC	AGT	GTT	1392
Ala	Thr	Glu	Leu	Ser	Pro	Thr	Ile	Leu	Glu	Gly	Ile	Val	Arg	Ser	Val	
	450					455					460					
AAT	CTA	CAA	CTG	GAT	ATC	AAC	ACA	GAC	GTA	TTA	GGT	AAA	GTT	TTT	CTG	1440
Asn	Leu	Gln	Leu	Asp	Ile	Asn	Thr	Asp	Val	Leu	Gly	Lys	Val	Phe	Leu	
465					470					475					480	
ACT	AAA	TAT	TAT	ATG	CAG	CGT	TAT	CCT	ATT	CAT	GCT	GAA	ACT	GCC	CTG	1488
Thr	Lys	Tyr	Tyr	Met	Gln	Arg	Tyr	Ala	Ile	His	Ala	Glu	Thr	Ala	Leu	
				485				490						495		
ATA	CTA	TGC	AAC	CCG	CCT	ATT	TCA	CAA	CGT	TCA	TAT	GAT	AAT	CAA	CGT	1536
Ile	Leu	Cys	Asn	Ala	Pro	Ile	Ser	Gln	Arg	Ser	Tyr	Asp	Asn	Gln	Pro	
			500					505						510		
AGC	CAA	TTT	GAT	CGC	CTG	TTT	AAT	ACG	CCA	TTA	CTG	AAC	GGA	CAA	TAT	1584
Ser	Gln	Phe	Asp	Arg	Leu	Phe	Asn	Thr	Pro	Leu	Leu	Asn	Gly	Gln	Tyr	
		515					520					525				
TTT	TCT	ACC	GGC	GAT	GAG	GAG	ATT	GAT	TTA	AAT	TCA	GCT	AGC	ACC	GGC	1632
Phe	Ser	Thr	Gly	Asp	Glu	Glu	Ile	Asp	Leu	Asn	Ser	Gly	Ser	Thr	Gly	
	530					535					540					
GAT	TGG	CGA	AAA	ACC	ATA	CTT	AAG	CGT	GCA	TTT	AAT	ATT	GAT	GAT	GTC	1680
Asp	Trp	Arg	Lys	Thr	Ile	Leu	Lys	Arg	Ala	Phe	Asn	Ile	Asp	Asp	Val	
545					550					555					560	
TGG	CTC	TTC	CGC	CTG	CTT	AAA	ATT	ACC	GAC	CAT	GAT	AAT	AAA	GAT	GGA	1728
Ser	Leu	Phe	Arg	Leu	Leu	Lys	Ile	Thr	Asp	His	Asp	Asn	Lys	Asp	Gly	
			565					570						575		
AAA	ATT	AAA	AAT	AAC	CTA	AAG	AAT	CTT	TCC	AAT	TTA	TAT	ATT	GGA	AAA	1776
Lys	Ile	Lys	Asn	Asn	Leu	Lys	Asn	Leu	Ser	Asn	Leu	Tyr	Ile	Gly	Lys	
			580				585						590			
TTA	CTG	GCA	GAT	ATT	CAT	CAA	TTA	ACC	ATT	GAT	GAA	CTG	GAT	TTA	TTA	1824
Leu	Leu	Ala	Asp	Ile	His	Gln	Leu	Thr	Ile	Asp	Glu	Leu	Asp	Leu	Leu	
		595					600					605				
CTG	ATT	GGC	GTA	GGT	GAA	GGA	AAA	ACT	AAT	TTA	TCC	GCT	ATC	AGT	GAT	1872
Leu	Ile	Ala	Val	Gly	Glu	Gly	Lys	Thr	Asn	Leu	Ser	Ala	Ile	Ser	Asp	
	610					615					620					
AAG	CAA	TTG	GCT	ACC	CTG	ATC	AGA	AAA	CTC	AAT	ACT	ATT	ACC	AGC	TGG	1920
Lys	Gln	Leu	Ala	Thr	Leu	Ile	Arg	Lys	Leu	Asn	Thr	Ile	Thr	Ser	Trp	
625					630					635					640	
CTA	CAT	ACA	CAG	AAG	TGG	AGT	GTA	TTC	CAG	CTA	TTT	ATC	ATG	ACC	TCC	1968
Leu	His	Thr	Gln	Lys	Trp	Ser	Val	Phe	Gln	Leu	Phe	Ile	Met	Thr	Ser	

U A 8 2 4 8 5 C 2

ACC AGC TAT AAC AAA ACC CTA ACC CCT GAA ATT AAG AAT TTG CTG GAT	2013
Thr. Ser Tyr Asn Lys Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp	
660 665 670	
ACC GTC TAC CAC GGT TTA CAA GGT TTT GAT AAA GAC AAA GCA GAT TTG	2064
Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asp Leu	
675 680 685	
CTA CAT GTC ATG GCG CCC TAT ATT GCG GCC ACC TTG CAA TTA TCA TCG	2112
Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser	
690 695 700	
GAA AAT GTC GCC CAC TCG GTA CTC CTT TCG GCA GAT AAG TTA CAG CCC	2160
Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Gln Pro	
705 710 715 720	
GGC GAC GCG GCA ATG ACA GCA GAA AAA TTC TCG GAC TCG TTG AAT ACT	2208
Gly Asp Gly Ala Met Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr	
725 730 735	
AAG TAT ACG CCG GGT TCA TCG GAA GCC GTA GAA ACG CAG GAA CAT ATC	2256
Lys Tyr Thr Pro Gly Ser Ser Glu Ala Val Glu Thr Gln Glu His Ile	
740 745 750	
GTT CAG TAT TGT CAG GCT CTG GCA CAA TTG GAA ATG GTT TAC CAT TCC	2304
Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser	
755 760 765	
ACC GGC ATC AAC GAA AAC GCC TTC CGT CTA TTT GTC ACA AAA CCA GAG	2352
Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu	
770 775 780	
ATG TTT GGC GCT GCA ACT GGA GCA GCG CCC GCG CAT GAT GCC CTT TCA	2400
Met Phe Gly Ala Ala Thr Gly Ala Ala Pro Ala His Asp Ala Leu Ser	
785 790 800	
CTG ATT ATG CTG ACA CGT TTT GCG GAT TGG GTG AAC GCA CTA GCG GAA	2448
Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu	
805 810 815	
AAA GCG TCC TCG GTG CTA GCG GCA TTT GAA GCT AAC TCG TTA ACC GCA	2496
Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala	
820 825 830	
GAA CAA CTG GCT GAT GCC ATG AAT CTT GAT GCT AAT TTG CTG TTG CAA	2544
Glu Gln Leu Ala Asp Ala Met Asn Leu Asp Ala Asn Leu Leu Gln	
835 840 845	
GCC ACT ATT CAA GCA CAA AAT CAT CAA CAT CTT CCC CCA GTA ACT CCA	2592
Ala Ser Ile Gln Ala Gln Asn His Gln His Leu Pro Pro Val Thr Pro	
850 855 860	
GAA AAT GCG TTC TCC TGT TGG ACA TCT ATC AAT ACT ATC CTG CAA TGG	2640
Glu Asn Ala Phe Ser Cys Trp Thr Ser Ile Asn Thr Ile Leu Gln Trp	
865 870 875 880	
GTT AAT GTC GCA CAA CAA TTG AAT GTC GCC CCA CAG GCG GTT TCC GCT	2688
Val Asn Val Ala Gln Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala	
885 890 895	
TTG GTC GCG CTG GAT TAT ATT CAA TCA ATG AAA GAG ACA CCG ACC TAT	2736
Leu Val Gly Leu Asp Tyr Ile Gln Ser Met Lys Glu Thr Pro Thr Tyr	
900 905 910	
GCC CAG TGG GAA AAC GCG GCA GCG GTA TTA ACC GCC GCG TTG AAT TCA	2784
Ala Gln Trp Glu Asn Ala Ala Gly Val Leu Thr Ala Gly Leu Asn Ser	
915 920 925	
CAA CAG GCT AAT ACA TTA CAC GCT TTT CTG GAT GAA TCT CGC ACT GCC	2832

U A 8 2 4 8 5 C 2

Gln Gln Ala Asn Thr Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala	930	935	940	
GCA TTA ACC ACC TAC TAT ATC CGT CAA GTC GCC AAG GCA GCG GCG GCT	945	950	955	960
Ala Leu Ser Thr Tyr Tyr Ile Arg Gln Val Ala Lys Ala Ala Ala Ala				
ATT AAA ACC CGT GAT GAC TTG TAT CAA TAC TTA CTG ATT GAT AAT CAG	965	970	975	
Ile Lys Ser Arg Asp Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln				
GTT TCT GCG GCA ATA AAA ACC ACC CCG ATC GCC GAA GCC ATT GCC AGT	980	985	990	
Val Ser Ala Ala Ile Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser				
ATT CAA CTG TAC GTC AAC CCG GCA TTG GAA AAT CTG GAA GAA AAT GCC	995	1000	1005	
Ile Gln Leu Tyr Val Asn Arg Ala Leu Glu Asn Val Glu Glu Asn Ala				
AAT TCG GCG GTT ATC AGC CCG CAA TTC TTT ATC GAC TGG GAC AAA TAC	1010	1015	1020	
Asn Ser Gly Val Ile Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr				
AAT AAA CCG TAC AGC ACT TCG CCG CGT GTT TCT CAA TTA GTT TAC TAC	1025	1030	1035	
Asn Lys Arg Tyr Ser Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr				
CCG GAA AAC TAT ATT GAT CCG ACC ATG CGT ATC GGA CAA ACC AAA ATC	1045	1050	1055	
Pro Glu Asn Tyr Ile Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met				
ATG GAC GCA TTA CTG CAA TCC GTC AGC CAA AGC CAA TTA AAC GCC GAT	1060	1065	1070	
Met Asp Ala Leu Leu Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp				
ACC GTC GAA GAT GCC TTT ATG TCT TAT CTG ACA TCG TTT GAA CAA GTC	1075	1080	1085	
Thr Val Glu Asp Ala Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val				
GCT AAT CTT AAA GTT ATT AGC GCA TAT CAC GAT AAT ATT AAT AAC GAT	1090	1095	1100	
Ala Asn Leu Lys Val Ile Ser Ala Tyr His Asp Asn Ile Asn Asn Asp				
CAA GCG CTG ACC TAT TTT ATC GGA CTC AGT GAA ACT GAT GCC GGT GAA	1105	1110	1115	1120
Gln Gly Leu Thr Tyr Phe Ile Gly Leu Ser Glu Thr Asp Ala Gly Glu				
TAT TAT TCG CCG AGT GTC GAT CAC AGT AAA TTC AAC GAC GGT AAA TTC	1125	1130	1135	
Tyr Tyr Trp Arg Ser Val Asp His Ser Lys Phe Asn Asp Gly Lys Phe				
GCG GCT AAT GCC TGG AGT GAA TGG CAT AAA ATT GAT TGT CCA ATT AAC	1140	1145	1150	
Ala Ala Asn Ala Trp Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn				
CCT TAT AAA AGC ACT ATC CGT CCA GTC ATA TAT AAA TCC CCG CTG TAT	1155	1160	1165	
Pro Tyr Lys Ser Thr Ile Arg Pro Val Ile Tyr Lys Ser Arg Leu Tyr				
CTG CTC TGG TTG GAA CAA AAG GAG ATC ACC AAA CAG ACA GGA AAT AGT	1170	1175	1180	
Leu Leu Trp Leu Glu Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser				
AAA GAT GCG TAT CAA ACT GAA ACG GAT TAT CGT TAT GAA CTA AAA TTG	1185	1190	1195	1200
Lys Asp Gly Tyr Gln Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu				
GCG CAT ATC CCG TAT GAT GCG ACT TGG AAT ACG CCA ATC ACC TTT CAT	1205	1210	1215	
Ala His Ile Arg Tyr Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp				

U A 8 2 4 8 5 C 2

UTC AAT AAA AAA ATA TCC CAG CTA AAA CTG GAA AAA AAT AGA GCG CCG	1220	1225	1230	1235
Val Asn Lys Lys Ile Ser Glu Leu Lys Leu Glu Lys Asn Arg Ala Pro				
GGA CTC TAT TGT CCC GGT TAT CAA GGT GAA GAT ACG TTG CTG CTG ATC	1235	1240	1245	
Gly Leu Tyr Cys Ala Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met				
TTT TAT AAC CAA CAA GAC ACA CTA GAT AGT TAT AAA AAC GCT TCA ATC	1250	1255	1260	
Phe Tyr Asn Gln Gln Asp Thr Leu Asp Ser Tyr Lys Asn Ala Ser Met				
CAA GGA CTA TAT ATC TTT GCT GAT ATG GCA TCC AAA GAT ATG ACC CCA	1265	1270	1275	1280
Gln Gly Leu Tyr Ile Phe Ala Asp Met Ala Ser Lys Asp Met Thr Pro				
GAA CAG AGC AAT GTT TAT CCG GAT AAT AGC TAT CAA CAA TTT GAT ACC	1285	1290	1295	
Glu Gln Ser Asn Val Tyr Arg Asp Asn Ser Tyr Gln Gln Phe Asp Thr				
AAT AAT CTC AGA AGA GTG AAT AAC CCG TAT GCA GAG GAT TAT GAG ATT	1300	1305	1310	
Asn Asn Val Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr Glu Ile				
CCT TCC TCG GTA AGT AGC CCT AAA GAC TAT CGT TGG GGA GAT TAT TAC	1315	1320	1325	
Pro Ser Ser Val Ser Ser Arg Lys Asp Tyr Gly Trp Gly Asp Tyr Tyr				
CTC AGC ATG GTA TAT AAC GGA GAT ATT CCA ACT ATC AAT TAC AAA GCC	1330	1335	1340	
Leu Ser Met Val Tyr Asn Gly Asp Ile Pro Thr Ile Asn Tyr Lys Ala				
GCA TCA AGT GAT TTA AAA ATC TAT ATC TCA CCA AAA TTA AGA ATT ATT	1345	1350	1355	1360
Ala Ser Ser Asp Leu Lys Ile Tyr Ile Ser Pro Lys Leu Arg Ile Ile				
GAT AAT GGA TAT GAA GGA CAG AAG CGC AAT CAA TGC AAT CTG ATG AAT	1365	1370	1375	
His Asn Gly Tyr Glu Gly Gln Lys Arg Asn Gln Cys Asn Leu Met Asn				
AAA TAT GGC AAA CTA GGT GAT AAA TTT ATT GTT TAT ACT AGC TTG GCG	1380	1385	1390	
Lys Tyr Gly Lys Leu Gly Asp Lys Phe Ile Val Tyr Thr Ser Leu Gly				
UTC AAT CCA AAT AAC TCG TCA AAT AAG CTC ATG TTT TAC CCC GTC TAT	1395	1400	1405	
Val Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val Tyr				
GAA TAT AGC GGA AAC ACC AGT GGA CTC AAT CAA GCG AGA CTA CTA TTC	1410	1415	1420	
Gln Tyr Ser Gly Asn Thr Ser Gly Leu Asn Gln Gly Arg Leu Leu Phe				
CAC CGT GAC ACC ACT TAT CCA TCT AAA GTA GAA GCT TGG ATT CCT GGA	1425	1430	1435	1440
His Arg Asp Thr Thr Tyr Pro Ser Lys Val Glu Ala Trp Ile Pro Gly				
GCA AAA CGT TCT CTA ACC AAC CAA AAT GCC GCC ATT GGT GAT GAT TAT	1445	1450	1455	
Ala Lys Arg Ser Leu Thr Asn Gln Asn Ala Ala Ile Gly Asp Asp Tyr				
GCT ACA GAC TCT CTG AAT AAA CCG GAT GAT CTT AAG CAA TAT ATC TTT	1460	1465	1470	
Ala Thr Asp Ser Leu Asn Lys Pro Asp Asp Leu Lys Gln Tyr Ile Phe				
ATG ACT GAC AGT AAA GCG ACT GCT ACT GAT GTC TCA GCG CCA GTA GAG	1475	1480	1485	
Met Thr Asp Ser Lys Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu				
ATT AAT ACT GCA ATT TCT CCA GCA AAA GTT CAG ATA ATA GTC AAA GCG	1490	1495	1500	
Ile Asn Thr Ala Ile Ser Pro Ala Lys Val Gln Ile Ile Val Lys Ala				

U A 8 2 4 8 5 C 2

GGT GGC AAG GAG CAA ACT TTT ACC GCA GAT AAA GAT GTC TCC ATT CAG 4560
 Gly Gly Lys Glu Gln Thr Phe Thr Ala Asp Lys Asp Val Ser Ile Gln 1505 1510 1515 1520
 CCA TCA CCT AGC TTT GAT GAA ATG AAT TAT CAA TTT AAT GCC CTT GAA 4608
 Pro Ser Pro Ser Phe Asp Glu Met Asn Tyr Gln Phe Asn Ala Leu Glu 1525 1530 1535
 ATA GAC GGT TCT GGT CTG AAT TTT ATT AAC AAC TCA GCC AGT ATT GAT 4656
 Ile Asp Gly Ser Gly Leu Asn Phe Ile Asn Asn Ser Ala Ser Ile Asp 1540 1545 1550
 GTT ACT TTT ACC GCA TTT GCG GAG GAT GGC CGC AAA CTG GGT TAT GAA 4704
 Val Thr Phe Thr Ala Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu 1555 1560 1565
 AGT TTC ACT ATT CCT GTT ACC CTC AAG GTA AGT ACC GAT AAT GCC CTG 4752
 Ser Phe Ser Ile Pro Val Thr Leu Lys Val Ser Thr Asp Asn Ala Leu 1570 1575 1580
 ACC CTG CAC CAT AAT GAA AAT GGT GCG CAA TAT ATG CAA TCG CAA TCC 4800
 Thr Leu His His Asn Glu Asn Gly Ala Gln Tyr Met Gln Trp Gln Ser 1585 1590 1595 1600
 TAT GGT ACC CGC CTG AAT ACT CTA TTT GCC CGC CAG TTG GTT GCA CGC 4848
 Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg 1605 1610 1615
 GCC ACC ACC GGA ATC GAT ACA ATT CTG AGT ATG GAA ACT CAG AAT ATT 4896
 Ala Thr Thr Gly Ile Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile 1620 1625 1630
 CAG GAA CCG CAG TTA GGC AAA GGT TTC TAT GGT ACC TTC CTG ATA CCT 4944
 Gln Glu Pro Gln Leu Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro 1635 1640 1645
 CCC TAT AAC CTA TCA ACT CAT GGT GAT GAA CGT TCG TTT AAG CTT TAT 4992
 Pro Tyr Asn Leu Ser Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr 1650 1655 1660
 ATC AAA CAT GTT GTT GAT AAT AAT TCA CAT ATT ATC TAT TCA GGC CAG 5040
 Ile Lys His Val Val Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln 1665 1670 1675 1680
 CTA ACA GAT ACA AAT ATA AAC ATC ACA TTA TTT ATT CCT CTT GAT GAT 5088
 Leu Thr Asp Thr Asn Ile Asn Ile Thr Leu Phe Ile Pro Leu Asp Asp 1685 1690 1695
 GTC CCA TTG AAT CAA GAT TAT CAC GCC AAG GTT TAT ATG ACC TTC AAG 5136
 Val Pro Leu Asn Gln Asp Tyr His Ala Lys Val Tyr Met Thr Phe Lys 1700 1705 1710
 AAA TCA CCA TCA GAT GGT ACC TGG TGG GGC CCT CAC TTT GTT AGA GAT 5184
 Lys Ser Pro Ser Asp Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp 1715 1720 1725
 GAT AAA CGA ATA GTA ACA ATA AAC CCT AAA TCC ATT TTG ACC CAT TTT 5232
 Asp Lys Gly Ile Val Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe 1730 1735 1740
 GAG AGC GTC AAT GTC CTG AAT AAT ATT AGT AGC GAA CCA ATG CAT TTC 5280
 Glu Ser Val Asn Val Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe 1745 1750 1755 1760
 AGC GGC GCT AAC AGC CTC TAT TTC TCG GAA CTG TTC TAC TAT ACC CGC 5328
 Ser Gly Ala Asn Ser Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro 1765 1770 1775
 ATG CTG GTT GCT CAA CGT TTG CTG CAT GAA CAG AAC TTC GAT GAA GCC 5376
 Met Leu Val Ala Gln Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala

1730	1785	1790	
AAC CGT TGG CTG AAA TAT GTC Asn Arg Trp Leu Lys Tyr Val 1795	TGG AGT CCA TCC GGT TAT ATT GTC CAC Trp Ser Pro Ser Gly Tyr Ile Val His 1800	5424	
GGC CAG ATT CAG AAC TAC CAG Gly Gln Ile Gln Asn Tyr Gln Trp 1810	TGG AAC GTC CGC CCG TTA CTG GAA GAC Trp Asn Val Arg Pro Leu Leu Glu Asp 1815	5472	
ACC AGT TGG AAC AGT GAT CCT TTG Thr Ser Trp Asn Ser Asp Pro Leu 1825	GAT TCC GTC GAT CCT GAC CGC GTA Asp Ser Val Asp Pro Asp Ala Val 1830	5520	
GCA CAG CAC GAT CCA ATG CAC TAC Ala Gln His Asp Pro Met His Tyr 1845	AAA GTT TCA ACT TTT ATG CGT ACC Lys Val Ser Thr Phe Met Arg Thr 1850	5568	
TTG GAT CTA TTG ATA GCA CGC GGC Leu Asp Leu Leu Ile Ala Arg Gly 1860	GAT CAT GCT TAT CGC CAA CTG GAA Asp His Ala Tyr Arg Gln Leu Glu 1865	5616	
CGA GAT ACA CTC AAC GAA GCG AAG Arg Asp Thr Leu Asn Glu Ala Lys 1875	ATG TGG TAT ATG CAA GCG CTG CAT Met Trp Tyr Met Gln Ala Leu His 1880	5664	
CTA TTA GGT GAC AAA CCT TAT CTA Leu Leu Gly Asp Lys Pro Tyr Leu 1890	CCG CTG AGT ACG ACA TGG AGT GAT Pro Leu Ser Thr Thr Trp Ser Asp 1895	5712	
CCA CGA CTA GAC AGA GCC GCG GAT Pro Arg Leu Asp Arg Ala Ala Asp 1905	ATC ACT ACC CAA AAT GCT CAC GAC Ile Thr Thr Gln Asn Ala His Asp 1910	5760	
AGC GCA ATA GTC GCT CTG CCG CAG Ser Ala Ile Val Ala Leu Arg Gln 1925	AAT ATA CCT ACA CCG GCA CCT TTA Asn Ile Pro Thr Pro Ala Pro Leu 1930	5808	
TCA TTG CGC AGC GCT AAT ACC CTG Ser Leu Arg Ser Ala Asn Thr Leu 1940	ACT GAT CTC TTC CTG CCG CAA ATC Thr Asp Leu Phe Leu Pro Gln Ile 1945	5856	
AAT GAA GTG ATG ATG AAT TAC TGG Asn Glu Val Met Met Asn Tyr Trp 1955	CAG ACA TTA GCT CAG AGA GTA TAC Gln Thr Leu Ala Gln Arg Val Tyr 1960	5904	
AAT CTG CGT CAT AAC CTC TCT ATC Asn Leu Arg His Asn Leu Ser Ile 1970	GAC GGC CAG CCG TTA TAT CTG CCA Asp Gly Gln Pro Leu Tyr Leu Pro 1975	5952	
ATC TAT GCC ACA CCG GCC GAT CCG Ile Tyr Ala Thr Pro Ala Asp Pro 1985	AAA GCG TTA CTC AGC GCC GCC GTT Lys Ala Leu Leu Ser Ala Ala Val 1990	6000	
GGC ACT TCT CAA GGT GGA GGC AAG Ala Thr Ser Gln Gly Gly Lys Leu 2005	CTA CCG GAA TCA TTT ATG TCC CTG Leu Pro Glu Ser Phe Met Ser Leu 2010	6048	
TGG CGT TTC CCG CAC ATG CTG GAA Trp Arg Phe Pro His Met Leu Glu 2020	AAT GCG CGC GGC ATG GTT AGC CAG Asn Ala Arg Gly Met Val Ser Gln 2025	6096	
CTC ACC CAG TTC GGC TCC ACG TTA Leu Thr Gln Phe Gly Ser Thr Leu 2035	CAA AAT ATT ATC GAA CGT CAG GAC Gln Asn Ile Ile Glu Arg Gln Asp 2040	6144	
GGC GAA CCG CTC AAT GCG TTA TTA Ala Glu Ala Leu Asn Ala Leu Leu 2050	CAA AAT CAG GCC GCC CAG CTG ATA Gln Asn Gln Ala Ala Glu Leu Ile 2055	6192	
TTG ACT AAC CTG AGC ATT CAG CAC Trp Ser Trp Asn Ser Asp Pro Leu 2065	AAA ACC ATT GAA GAA TTG GAT GCC Lys Val Ser Thr Phe Met Ser Leu 2070	6240	

U A 8 2 4 8 5 C 2

Leu Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala	2065	2070	2075	2080
GAG AAA ACG GTG TTG GAA AAA TCC AAA GCG GGA GCA CAA TCG CGC TTT	6283			
Glu Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe	2085	2090	2095	
GAT AGC TAC GGC AAA CTG TAC GAT GAG AAT ATC AAC GCC GGT GAA AAC	6336			
Asp Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Thr Thr Ala Asn	2100	2105	2110	
CAA GCG ATG ACG CTA CGA GCG TCC GCG GCG GCG CTT ACC ACG GCA GTT	6384			
Gln Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val	2115	2120	2125	
CAG GCA TCC CGT CTG GCC GGT GCG GCG GCT GAT CTG GTG CCT AAC ATC	6432			
Gln Ala Ser Arg Leu Ala Gly Ala Ala Asp Leu Val Pro Asn Ile	2130	2135	2140	
TTC GCG TTT GCG GGT GCG GCG AGC CGT TCG GCG GCT ATC GCT GAG GCG	6480			
Phe Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala	2145	2150	2155	2160
ACA GGT TAT GTG ATG GAA TTC TCC GCG AAT GTT ATG AAC ACC GAA GCG	6528			
Thr Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala	2165	2170	2175	
GAT AAA ATT AGC CAA TCT GAA ACC TAC CGT CGT GCG CGT CAG GAG TGG	6576			
Asp Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp	2180	2185	2190	
GAG ATC CAG CCG AAT AAT GCC GAA GCG GAA TTC AAG CAA ATC GAT GCT	6624			
Glu Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala	2195	2200	2205	
CAG CTC AAA TCA CTC GCT GTA CCG CCG GAA GCC GCC GTA TTG CAG AAA	6672			
Gln Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys	2210	2215	2220	
ACC AGT CTG AAA ACC CAA CAA GAA CAG ACC CAA TCT CAA TTG GCC TTC	6720			
Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe	2225	2230	2235	2240
CTG CAA CGT AAG TTC AGC AAT CAG CCG TTA TAC AAC TCG CTC CGT GGT	6768			
Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly	2245	2250	2255	
CGA CTG GCG GCG ATT TAC TTC CAG TTC TAC GAT TTG GCC GTC GCG CGT	6816			
Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg	2260	2265	2270	
TGC CTC ATG GCA GAA CAA GCT TAC CGT TGG GAA CTC AAT GAT GAC TCT	6864			
Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser	2275	2280	2285	
GCC CCG TTC ATT AAA CCG GCG GCC TGG CAG GGA ACC TAT GCC GGT CTG	6912			
Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu	2290	2295	2300	
CTT GCA GGT GAA ACC TTG ATG CTG AGT CTG GCA CAA ATG GAA GAC GCT	6960			
Leu Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala	2305	2310	2315	2320
CAT CTG AAA CCG GAT AAA CCG GCA TTA CAG GTT GAA CCG ACA GTA TCG	7008			
His Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser	2325	2330	2335	
CTG GCC GAA GTT TAT GCA GGA TTA CCA AAA GAT AAC GGT CCA TTT TCC	7056			
Leu Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser	2340	2345	2350	

U A 8 2 4 8 5 C 2

CTG GGT CAG GAA ATT GAC AAG CTG CTC AGT CAA GGT TCA GGC AGT GGC 7134
 Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala
 2355 2360 2365
 GGC AGT GGT AAT AAT AAT TTG GCG TTC GCG GCG GCG ACC GAC ACT AAA 7152
 Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys
 2370 2375 2380
 ACC TCT TTG CAG GCA TCA GTT TCA TTC GCT GAT TTG AAA ATT CGT GAA 7200
 Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu
 2385 2390 2395 2400
 GAT TAC CCG GCA TCG CTT GGC AAA ATT CGA CGT ATC AAA CAG ATC ACC 7248
 Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser
 2405 2410 2415
 GTC ACT TTG CCC GCG CTA CTG GGA CCG TAT CAG GAT GTA CAG GCA ATA 7296
 Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile
 2420 2425 2430
 TTG TCT TAC GGC GAT AAA GCG GGA TTA GCT AAC GCG TGT GAA GCG CTG 7344
 Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu
 2435 2440 2445
 GCA GTT TCT CAC GGT ATG AAT GAC AGC GCG CAA TTC CAG CTC GAT TTC 7392
 Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe
 2450 2455 2460
 AAC GAT GCG AAA TTC CTG CCA TTC GAA GCG ATC GCG ATT GAT CAA GCG 7440
 Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly
 2465 2470 2475 2480
 ACC CTG ACA CTG AGC TTC CCA AAT GCA TCT ATG CCG CAG AAA GGT AAA 7488
 Thr Leu Thr Met Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys
 2485 2490 2495
 CAA GCC ACT ATG TTA AAA ACC CTG AAC GAT ATC ATT TTG CAT ATT CCG 7536
 Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg
 2500 2505 2510
 TAC ACC ATT AAA TAA 7551
 Tyr Thr Ile Lys ...
 2516

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 47:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 2516 амінокислот
- (B) ТИП: амінокислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 47 (TcdA):

Features	From	To	Description
Peptide	1	2516	TcdA proteins
Peptide	89	1937	TcdA ₁₁ peptide
Fragment	89	100	S2-N-terminus (SEQ ID NO:13)
Fragment	284	299	(SEQ ID NO:38)
Fragment	554	563	(SEQ ID NO:17)
Fragment	1080	1092	(SEQ ID NO:23; 12/13)
Fragment	1385	1400	(SEQ ID NO:18)
Fragment	1478	1497	(SEQ ID NO:39)
Fragment	1620	1642	(SEQ ID NO:21; 19/23)
Fragment	1938	1948	(SEQ ID NO:41)
Peptide	1938	2516	TcdA ₁₁₁ peptide
Fragment	2327	2345	(SEQ ID NO:42)
Fragment	2398	2408	(SEQ ID NO:43)

Met Asn Glu Ser Val Lys Glu Ile Pro Asp Val Leu Lys Ser Gln Cys
 1 5 10 15
 Gly Phe Asn Cys Leu Thr Asp Ile Ser His Ser Ser Phe Asn Glu Phe
 20 25 30
 Arg Gln Gln Val Ser Glu His Leu Ser Trp Ser Glu Thr His Asp Leu
 35 40 45
 Tyr His Asp Ala Gln Gln Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala
 50 55 60
 Arg Ile Leu Lys Arg Ala Asn Pro Gln Leu Gln Asn Ala Val His Leu
 65 70 75 80
 Ala Ile Leu Ala Pro Asn Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe
 85 90 95
 Ser Gly Arg Ala Ser Gln Tyr Val Ala Pro Gly Thr Val Ser Ser Met
 100 105 110
 Phe Ser Pro Ala Ala Tyr Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn
 115 120 125
 Leu His Ala Ser Asp Ser Val Tyr Tyr Leu Asp Thr Arg Arg Pro Asp
 130 135 140
 Leu Lys Ser Met Ala Leu Ser Gln Gln Asn Met Asp Ile Glu Leu Ser
 145 150 155 160
 Thr Leu Ser Leu Ser Asn Glu Leu Leu Leu Glu Ser Ile Lys Thr Glu
 165 170 175
 Ser Lys Leu Glu Asn Tyr Thr Lys Val Met Glu Met Leu Ser Thr Phe
 180 185 190
 Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asn Val Arg
 195 200 205
 Glu Val Ile Gln Leu Gln Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser
 210 215 220
 Pro Ala Ile Ala Gly Leu Met His Gln Ala Ser Leu Leu Gly Ile Asn
 225 230 235 240
 Ala Ser Ile Ser Pro Glu Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr
 245 250 255
 Glu Gly Asn Ala Glu Glu Leu Tyr Lys Lys Asn Phe Gly Asn Ile Glu
 260 265 270
 Pro Ala Ser Leu Ala Met Pro Glu Tyr Leu Lys Arg Tyr Tyr Asn Leu
 275 280 285
 Ser Asp Glu Glu Leu Ser Gln Phe Ile Gly Lys Ala Ser Asn Phe Gly
 290 295 300
 Gln Gln Glu Tyr Ser Asn Asn Gln Leu Ile Thr Pro Val Val Asn Ser
 305 310 315 320
 Ser Asp Gly Thr Val Lys Val Tyr Arg Ile Thr Arg Glu Tyr Thr Thr
 325 330 335
 Asn Ala Tyr Gln Met Asp Val Glu Leu Phe Pro Phe Gly Gly Glu Asn
 340 345 350
 Tyr Arg Leu Asp Tyr Lys Phe Lys Asn Phe Tyr Asn Ala Ser Tyr Leu
 355 360 365
 Ser Ile Lys Leu Asn Asp Lys Arg Glu Leu Val Arg Thr Glu Gly Ala

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

370	375	380
Pro Gln Val Asn Ile Glu Tyr Ser Ala Asn Ile Thr Leu Asn Thr Ala 385 390 395 400		
Asp Ile Ser Gln Pro Phe Glu Ile Gly Leu Thr Arg Val Leu Pro Ser 405 410 415		
Gly Ser Trp Ala Tyr Ala Ala Ala Lys Phe Thr Val Glu Glu Tyr Asn 420 425 430		
Gln Tyr Ser Phe Leu Leu Lys Leu Asn Lys Ala Ile Arg Leu Ser Arg 435 440 445		
Ala Thr Glu Leu Ser Pro Thr Ile Leu Glu Gly Ile Val Arg Ser Val 450 455 460		
Asn Leu Gln Leu Asp Ile Asn Thr Asp Val Leu Gly Lys Val Phe Leu 465 470 475 480		
Thr Lys Tyr Tyr Met Gln Arg Tyr Ala Ile His Ala Glu Thr Ala Leu 485 490 495		
Ile Leu Cys Asn Ala Pro Ile Ser Gln Arg Ser Tyr Asp Asn Cln Pro 500 505 510		
Ser Gln Phe Asp Arg Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr 515 520 525		
Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Ser Gly Ser Thr Gly 530 535 540		
Asp Trp Arg Lys Thr Ile Leu Lys Arg Ala Phe Asn Ile Asp Asp Val 545 550 555 560		
Ser Leu Phe Arg Leu Leu Lys Ile Thr Asp His Asp Asn Lys Asp Gly 565 570 575		
Lys Ile Lys Asn Asn Leu Lys Asn Leu Ser Asn Leu Tyr Ile Gly Lys 580 585 590		
Leu Leu Ala Asp Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu 595 600 605		
Leu Ile Ala Val Gly Glu Gly Lys Thr Asn Leu Ser Ala Ile Ser Asp 610 615 620		
Lys Gln Leu Ala Thr Leu Ile Arg Lys Leu Asn Thr Ile Thr Ser Trp 625 630 635 640		
Leu His Thr Gln Lys Trp Ser Val Phe Gln Leu Phe Ile Met Thr Ser 645 650 655		
Thr Ser Tyr Asn Lys Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp 660 665 670		
Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asp Leu 675 680 685		
Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser 690 695 700		
Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Gln Pro 705 710 715 720		
Gly Asp Gly Ala Met Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr 725 730 735		
Lys Tyr Thr Pro Gly Ser Ser Glu Ala Val Glu Thr Gln Glu His Ile 740 745 750		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Val	Gln	Tyr	Cys	Gln	Ala	Leu	Ala	Gln	Leu	Glu	Met	Val	Tyr	His	Ser	
	755						760					765				
Thr	Gly	Ile	Asn	Glu	Asn	Ala	Phe	Arg	Leu	Phe	Val	Thr	Lys	Pro	Glu	
	770					775					780					
Met	Phe	Gly	Ala	Ala	Thr	Gly	Ala	Ala	Pro	Ala	His	Asp	Ala	Leu	Ser	
	785				790				795						800	
Leu	Ile	Met	Leu	Thr	Arg	Phe	Ala	Asp	Trp	Val	Asn	Ala	Leu	Gly	Glu	
				805				810						815		
Lys	Ala	Ser	Ser	Val	Leu	Ala	Ala	Phe	Glu	Ala	Asn	Ser	Leu	Thr	Ala	
			820					825					830			
Glu	Gln	Leu	Ala	Asp	Ala	Met	Asn	Leu	Asp	Ala	Asn	Leu	Leu	Leu	Gln	
		835					840					845				
Ala	Ser	Ile	Gln	Ala	Gln	Asn	His	Gln	His	Leu	Pro	Pro	Val	Thr	Pro	
	850					855					860					
Glu	Asn	Ala	Phe	Ser	Cys	Trp	Thr	Ser	Ile	Asn	Thr	Ile	Leu	Gln	Trp	
	865				870					875					880	
Val	Asn	Val	Ala	Gln	Gln	Leu	Asn	Val	Ala	Pro	Gln	Gly	Val	Ser	Ala	
			885						890					895		
Leu	Val	Gly	Leu	Asp	Tyr	Ile	Gln	Ser	Met	Lys	Glu	Thr	Pro	Thr	Tyr	
			900					905					910			
Ala	Gln	Trp	Glu	Asn	Ala	Ala	Gly	Val	Leu	Thr	Ala	Gly	Leu	Asn	Ser	
		915					920					925				
Gln	Gln	Ala	Asn	Thr	Leu	His	Ala	Phe	Leu	Asp	Glu	Ser	Arg	Ser	Ala	
		930				935					940					
Ala	Leu	Ser	Thr	Tyr	Tyr	Ile	Arg	Gln	Val	Ala	Lys	Ala	Ala	Ala	Ala	
	945				950					955					960	
Ile	Lys	Ser	Arg	Asp	Asp	Leu	Tyr	Gln	Tyr	Leu	Leu	Ile	Asp	Asn	Gln	
				965				970						975		
Val	Ser	Ala	Ala	Ile	Lys	Thr	Thr	Arg	Ile	Ala	Glu	Ala	Ile	Ala	Ser	
			980					985					990			
Ile	Gln	Leu	Tyr	Val	Asn	Arg	Ala	Leu	Glu	Asn	Val	Glu	Glu	Asn	Ala	
		995				1000						1005				
Asn	Ser	Gly	Val	Ile	Ser	Arg	Gln	Phe	Phe	Ile	Asp	Trp	Asp	Lys	Tyr	
	1010					1015					1020					
Asn	Lys	Arg	Tyr	Ser	Thr	Trp	Ala	Gly	Val	Ser	Gln	Leu	Val	Tyr	Tyr	
	1025				1030					1035					1040	
Pro	Glu	Asn	Tyr	Ile	Asp	Pro	Thr	Met	Arg	Ile	Gly	Gln	Thr	Lys	Met	
				1045					1050					1055		
Met	Asp	Ala	Leu	Leu	Gln	Ser	Val	Ser	Gln	Ser	Gln	Leu	Asn	Ala	Asp	
			1060					1065					1070			
Thr	Val	Glu	Asp	Ala	Phe	Met	Ser	Tyr	Leu	Thr	Ser	Phe	Glu	Gln	Val	
		1075					1080					1085				
Ala	Asn	Leu	Lys	Val	Ile	Ser	Ala	Tyr	His	Asp	Asn	Ile	Asn	Asn	Asp	
		1090				1095					1100					
Gln	Gly	Leu	Thr	Tyr	Phe	Ile	Gly	Leu	Ser	Glu	Thr	Asp	Ala	Gly	Glu	
	1105				1110					1115					1120	
Tyr	Tyr	Trp	Arg	Ser	Val	Asp	His	Ser	Lys	Phe	Asn	Asp	Gly	Lys	Phe	
				1125					1130					1135		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Ala Ala Asn Ala Trp Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn
1140 1145 1150

Pro Tyr Lys Ser Thr Ile Arg Pro Val Ile Tyr Lys Ser Arg Leu Tyr
1155 1160 1165

Leu Leu Trp Leu Glu Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser
1170 1175 1180

Lys Asp Gly Tyr Gln Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu
1185 1190 1195 1200

Ala His Ile Arg Tyr Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp
1205 1210 1215

Val Asn Lys Lys Ile Ser Glu Leu Lys Leu Glu Lys Asn Arg Ala Pro
1220 1225 1230

Gly Leu Tyr Cys Ala Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met
1235 1240 1245

Phe Tyr Asn Gln Gln Asp Thr Leu Asp Ser Tyr Lys Asn Ala Ser Met
1250 1255 1260

Gln Gly Leu Tyr Ile Phe Ala Asp Met Ala Ser Lys Asp Met Thr Pro
1265 1270 1275 1280

Glu Gln Ser Asn Val Tyr Arg Asp Asn Ser Tyr Gln Gln Phe Asp Thr
1285 1290 1295

Asn Asn Val Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr Glu Ile
1300 1305 1310

Pro Ser Ser Val Ser Ser Arg Lys Asp Tyr Gly Trp Gly Asp Tyr Tyr
1315 1320 1325

Leu Ser Met Val Tyr Asn Gly Asp Ile Pro Thr Ile Asn Tyr Lys Ala
1330 1335 1340

Ala Ser Ser Asp Leu Lys Ile Tyr Ile Ser Pro Lys Leu Arg Ile Ile
1345 1350 1355 1360

His Asn Gly Tyr Glu Gly Gln Lys Arg Asn Gln Cys Asn Leu Met Asn
1365 1370 1375

Lys Tyr Gly Lys Leu Gly Asp Lys Phe Ile Val Tyr Thr Ser Leu Gly
1380 1385 1390

Val Asn Pro Asn Asn Ser Ser Asn Lys Leu Met Phe Tyr Pro Val Tyr
1395 1400 1405

Gln Tyr Ser Gly Asn Thr Ser Gly Leu Asn Gln Gly Arg Leu Leu Phe
1410 1415 1420

His Arg Asp Thr Thr Tyr Pro Ser Lys Val Glu Ala Trp Ile Pro Gly
1425 1430 1435 1440

Ala Lys Arg Ser Leu Thr Asn Gln Asn Ala Ala Ile Gly Asp Asp Tyr
1445 1450 1455

Ala Thr Asp Ser Leu Asn Lys Pro Asp Asp Leu Lys Gln Tyr Ile Phe
1460 1465 1470

Met Thr Asp Ser Lys Gly Thr Ala Thr Asp Val Ser Gly Pro Val Glu
1475 1480 1485

Ile Asn Thr Ala Ile Ser Pro Ala Lys Val Gln Ile Ile Val Lys Ala
1490 1495 1500

Gly Gly Lys Glu Gln Thr Phe Thr Ala Asp Lys Asp Val Ser Ile Gln

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

1505	1510	1515	1520
Pro Ser Pro Ser Phe Asp Glu Met Asn Tyr Gln Phe Asn Ala Leu Glu	1525	1530	1535
Ile Asp Gly Ser Gly Leu Asn Phe Ile Asn Asn Ser Ala Ser Ile Asp	1540	1545	1550
Val Thr Phe Thr Ala Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu	1555	1560	1565
Ser Phe Ser Ile Pro Val Thr Leu Lys Val Ser Thr Asp Asn Ala Leu	1570	1575	1580
Thr Leu His His Asn Glu Asn Gly Ala Gln Tyr Met Gln Trp Gln Ser	1585	1590	1595
Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg	1605	1610	1615
Ala Thr Thr Gly Ile Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile	1620	1625	1630
Gln Glu Pro Gln Leu Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro	1635	1640	1645
Pro Tyr Asn Leu Ser Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr	1650	1655	1660
Ile Lys His Val Val Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln	1665	1670	1675
Leu Thr Asp Thr Asn Ile Asn Ile Thr Leu Phe Ile Pro Leu Asp Asp	1685	1690	1695
Val Pro Leu Asn Gln Asp Tyr His Ala Lys Val Tyr Met Thr Phe Lys	1700	1705	1710
Lys Ser Pro Ser Asp Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp	1715	1720	1725
Asp Lys Gly Ile Val Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe	1730	1735	1740
Glu Ser Val Asn Val Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe	1745	1750	1755
Ser Gly Ala Asn Ser Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro	1765	1770	1775
Met Leu Val Ala Gln Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala	1780	1785	1790
Asn Arg Trp Leu Lys Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val His	1795	1800	1805
Gly Gln Ile Gln Asn Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp	1810	1815	1820
Thr Ser Trp Asn Ser Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val	1825	1830	1835
Ala Gln His Asp Pro Met His Tyr Lys Val Ser Thr Phe Met Arg Thr	1845	1850	1855
Leu Asp Leu Leu Ile Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu	1860	1865	1870
Arg Asp Thr Leu Asn Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His	1875	1880	1885

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Leu Leu Gly Asp Lys Pro Tyr Leu Pro Leu Ser Thr Thr Trp Ser Asp
1390 1395 1900

Pro Arg Leu Asp Arg Ala Ala Asp Ile Thr Thr Gln Asn Ala His Asp
1905 1910 1915 1920

Ser Ala Ile Val Ala Leu Arg Gln Asn Ile Pro Thr Pro Ala Pro Leu
1925 1930 1935

Ser Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile
1940 1945 1950

Asn Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr
1955 1960 1965

Asn Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro
1970 1975 1980

Ile Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val
1985 1990 1995 2000

Ala Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu
2005 2010 2015

Trp Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln
2020 2025 2030

Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp
2035 2040 2045

Ala Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile
2050 2055 2060

Leu Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala
2065 2070 2075 2080

Glu Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe
2085 2090 2095

Asp Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn
2100 2105 2110

Gln Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val
2115 2120 2125

Gln Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile
2130 2135 2140

Phe Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala
2145 2150 2155 2160

Thr Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala
2165 2170 2175

Asp Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp
2180 2185 2190

Glu Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala
2195 2200 2205

Gln Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys
2210 2215 2220

Thr Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe
2225 2230 2235 2240

Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly
2245 2250 2255

Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg
2260 2265 2270

Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser
 2275 2280 2285
 Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu
 2290 2295 2300
 Leu Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala
 2305 2310 2315 2320
 His Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser
 2325 2330 2335
 Leu Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser
 2340 2345 2350
 Leu Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala
 2355 2360 2365
 Gly Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys
 2370 2375 2380
 Thr Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu
 2385 2390 2395 2400
 Asp Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser
 2405 2410 2415
 Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile
 2420 2425 2430
 Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu
 2435 2440 2445
 Ala Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe
 2450 2455 2460
 Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly
 2465 2470 2475 2480
 Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys
 2485 2490 2495
 Gln Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg
 2500 2505 2510
 Tyr Thr Ile Lys
 2516

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 48:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 5547 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 48 (кодуючий район tcdA₁₁):

CTC ATA GCC TAT AAC AAT CAA TTT ACC GGT AGA GCC ACT CAA TAT GTT 48
 Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val
 1 5 10 15
 GCG CCG GGT ACC GTT TCT TCC ATG TTC TCC CCC GCC GCT TAT TTG ACT 96
 Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr
 20 25 30

GAA CTT TAT COT GAA CCA CCG AAT TTA CAC GCA AGT GAC TCC GTT TAT 144
 Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr
 35 40 45
 TAT CTG GAT ACC CGC CGC CCA GAT CTC AAA TCA ATG GCG CTC ACT CAG 192
 Tyr Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln
 50 55 60
 CAA AAT ATG GAT ATA GAA TTA TCC ACA CTC TCT TTG TCC AAT GAG CTG 240
 Gln Asn Met Asp Ile Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu
 65 70 75 80
 TTA TTG GAA AGC ATT AAA ACT GAA TCT AAA CTG GAA AAC TAT ACT AAA 288
 Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Glu Asn Tyr Thr Lys
 85 90 95
 GTG ATG GAA ATG CTC TCC ACT TTC CGT CCT TCC GGC CCA ACG CCT TAT 336
 Val Met Glu Met Leu Ser Thr Phe Arg Pro Ser Gly Ala Thr Pro Tyr
 100 105 110
 CAT GAT COT TAT GAA AAT CTC CGT GAA GTT ATC CAG CTA CAA GAT CCT 384
 His Asp Ala Tyr Glu Asn Val Arg Glu Val Ile Gln Leu Gln Asp Pro
 115 120 125
 GGA CTT GAG CAA CTC AAT GCA TCA CCG GCA ATT GCC GCG TTG ATG CAT 432
 Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu Met His
 130 135 140
 CAA GCC TCC CTA TTG GGT ATT AAC GCT TCA ATC TCG CCT GAG CTA TTT 480
 Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu Leu Phe
 145 150 155 160
 AAT ATT CTG ACG GAG GAG ATT ACC GAA GGT AAT CCT GAG CAA CTT TAT 528
 Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu Leu Tyr
 165 170 175
 AAG AAA AAT TTT GGT AAT ATC GAA CCG GCC TCA TTG GCT ATG CCG GAA 576
 Lys Lys Asn Phe Gly Asn Ile Glu Pro Ala Ser Leu Ala Met Pro Glu
 180 185 190
 TAC CTT AAA COT TAT TAT AAT TTA AGC GAT GAA GAA CTT AGT CAG TTT 624
 Tyr Leu Lys Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Ser Gln Phe
 195 200 205
 ATT GGT AAA GCC AGC AAT TTT GGT CAA CAG GAA TAT AGT AAT AAC CAA 672
 Ile Gly Lys Ala Ser Asn Phe Gly Gln Gln Glu Tyr Ser Asn Asn Gln
 210 215 220
 GTT ATT ACT CCG GTA GTC AAC AGC AGT GAT GCC ACG GTT AAG GTA TAT 720
 Leu Ile Thr Pro Val Val Asn Ser Ser Asp Gly Thr Val Lys Val Tyr
 225 230 235 240
 CCG ATC ACC CGC GAA TAT ACA ACC AAT GCT TAT CAA ATG GAT GTG GAG 768
 Arg Ile Thr Arg Glu Tyr Thr Thr Asn Ala Tyr Gln Met Asp Val Glu
 245 250 255
 CTA TTT CCC TTC GGT GGT GAG AAT TAT CCG TTA GAT TAT AAA TTC AAA 816
 Leu Phe Pro Phe Gly Gly Glu Asn Tyr Arg Leu Asp Tyr Lys Phe Lys
 260 265 270
 AAT TTT TAT AAT GCC TCT TAT TTA TCC ATC AAG TTA AAT GAT AAA ACA 864
 Asn Phe Tyr Asn Ala Ser Tyr Leu Ser Ile Lys Leu Asn Asp Lys Arg
 275 280 285
 GAA CTT GTT CGA ACT GAA GCC GCT CCT CAA GTC AAT ATA GAA TAC TCC 912
 Glu Leu Val Arg Thr Glu Gly Ala Pro Gln Val Asn Ile Glu Tyr Ser
 290 295 300
 GCA AAT ATC ACA TTA AAT ACC GCT GAT ATC AGT CAA CCT TTT GAA ATT 960
 Ala Asn Ile Thr Leu Asn Thr Ala Asp Ile Ser Gln Pro Phe Glu Ile

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

305 GGC CTG ACA CGA GTA CTT CCT TCC GGT TCT TCG GCA TAT GCC GCC GCA 1003
 Gly Leu Thr Arg Val Leu Pro Ser Gly Ser Trp Ala Tyr Ala Ala Ala
 325 330 335
 310 AAA TTT ACC GTT GAA GAG TAT AAC CAA TAC TCT TTT CTG CTA AAA CTT 1056
 Lys Phe Thr Val Glu Glu Tyr Asn Gln Tyr Ser Phe Leu Leu Lys Leu
 340 345 350
 315 AAC AAG GCT ATT CCT CTA TCA CCT GCG ACA GAA TTG TCA CCC ACC ATT 1104
 Asn Lys Ala Ile Arg Leu Ser Arg Ala Thr Glu Leu Ser Pro Thr Ile
 355 360 365
 320 CTG GAA GGC ATT GTG CCC AGT GTT AAT CTA CAA CTG GAT ATC AAC ACA 1152
 Leu Glu Gly Ile Val Arg Ser Val Asn Leu Gln Leu Asp Ile Asn Thr
 370 375 380
 GAC GTA TTA GGT AAA GTT TTT CTG ACT AAA TAT TAT ATC CAG CGT TAT 1200
 Asp Val Leu Gly Lys Val Phe Leu Thr Lys Tyr Tyr Met Gln Arg Tyr
 385 390 395 400
 GCT ATT CAT GCT GAA ACT GCC CTG ATA CTA TCC AAC GCC CCT ATT TCA 1248
 Ala Ile His Ala Glu Thr Ala Leu Ile Leu Cys Asn Ala Pro Ile Ser
 405 410 415
 CAA CGT TCA TAT GAT AAT CAA CCT ACC CAA TTT GAT CGC CTG TTT AAT 1296
 Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn
 420 425 430
 ACC CGA TTA CTG AAC CGA CAA TAT TTT TCT ACC GGC GAT GAG GAG ATT 1344
 Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile
 435 440 445
 GAT TTA AAT TCA GGT ACC ACC GGC GAT TGG CGA AAA ACC ATA CTT AAG 1392
 Asp Leu Asn Ser Gly Ser Thr Gly Asp Trp Arg Lys Thr Ile Leu Lys
 450 455 460
 CGT GCA TTT AAT ATT GAT GAT CTC TCG CTC TTC GCC CTG CTT AAA ATT 1440
 Arg Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile
 465 470 475 480
 ACC GAC CAT GAT AAT AAA GAT GGA AAA ATT AAA AAT AAC CTA AAG AAT 1488
 Thr Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn
 485 490 495
 CTT TCC AAT TTA TAT ATT GGA AAA TTA CTG GCA GAT ATT CAT CAA TTA 1536
 Leu Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu
 500 505 510
 ACC ATT GAT GAA CTG GAT TTA TTA CTG ATT GCC GTA GGT GAA GGA AAA 1584
 Thr Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys
 515 520 525
 ACT AAT TTA TCC CCT ATC AGT GAT AAG CAA TTG GCT ACC CTG ATC AGA 1632
 Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg
 530 535 540
 AAA CTC AAT ACT ATT ACC ACC TGG CTA CAT ACA CAG AAG TCG AGT GTA 1680
 Lys Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val
 545 550 555 560
 TTC CAG CTA TTT ATC ATG ACC TCC ACC ACC TAT AAC AAA ACC CTA ACC 1728
 Phe Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr
 565 570 575
 CCT GAA ATT AAG AAT TTG CTG GAT ACC GTC TAC CAC GGT TTA CAA GGT 1776
 Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly
 580 585 590
 TTT GAT AAA GAC AAA GCA GAT TTG CTA CAT GTC ATC GCG CCC TAT ATT 1824

Phe Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile
 595 600 605
 GCG GCC ACC TTG CAA TTA TCA TCG GAA AAT GTC GCC CAC TCG GTA CTC 1372
 Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu
 610 615 620
 CTT TGG GCA GAT AAG TTA CAG CCC GCG GAC GCG GCA ATG ACA GCA GAA 1920
 Leu Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu
 625 630 635 640
 AAA TTC TGG GAC TGG TTG AAT ACT AAG TAT ACG CCG GGT TCA TCG GAA 1963
 Lys Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu
 645 650 655
 GCG GTA GAA ACG CAG GAA CAT ATC GTT CAG TAT TGT CAG GCT CTC GCA 2016
 Ala Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala
 660 665 670
 CAA TTG GAA ATG GTT TAC CAT TCC ACC GCG ATC AAC GAA AAC GCG TTC 2064
 Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe
 675 680 685
 CGT GTA TTT GTC ACA AAA CCA GAG ATG TTT GCG GCT GCA ACT GGA GCA 2112
 Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala
 690 695 700
 GCG CCC GCG CAT GAT GCC CTT TCA CTC ATT ATG CTG ACA CGT TTT GCG 2160
 Ala Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala
 705 710 715 720
 GAT TGG GTC AAC GCA CTA GCG GAA AAA GCG TCC TCC GTC CTA GCG GCA 2208
 Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala
 725 730 735
 TTT GAA GCT AAC TCG TTA ACG GCA GAA CAA CTG GCT GAT GCG ATG AAT 2256
 Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn
 740 745 750
 CTT GAT GCT AAT TTG CTC TTG CAA GCC AGT ATT CAA GCA CAA AAT CAT 2304
 Leu Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His
 755 760 765
 CAA CAT CTT CCC CCA GTA ACT CCA GAA AAT GCG TTC TCC TGT TCG ACA 2352
 Gln His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr
 770 775 780
 TCT ATC AAT ACT ATC CTC CAA TGG GTT AAT GTC GCA CAA CAA TTG AAT 2400
 Ser Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Asn
 785 790 795 800
 GTC GCC CCA CAG GCG GTT TCC GCT TTG GTC GCG CTG GAT TAT ATT CAA 2448
 Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp Tyr Ile Gln
 805 810 815
 TCA ATG AAA GAG ACA CCG ACC TAT GCG CAG TGG GAA AAC GCG GCA GCG 2496
 Ser Met Lys Glu Thr Pro Thr Tyr Ala Gln Trp Glu Asn Ala Ala Gly
 820 825 830
 GTA TTA ACC GCG GCG TTG AAT TCA CAA CAG GCT AAT ACA TTA CAC GCT 2544
 Val Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asn Thr Leu His Ala
 835 840 845
 TTT CTC GAT GAA TCT CCG AGT GCG GCA TTA AGC ACC TAC TAT ATC GGT 2592
 Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg
 850 855 860
 CAA CTC GCC AAG GCA GCG GCG GCT ATT AAA AGC CGT GAT GAC TTG TAT 2640
 Gln Val Ala Lys Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr
 865 870 875 880

CAA TAC TTA CTC ATT GAT AAT CAG GTT TCT GCG GCA ATA AAA ACC ACC 1888
Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr
385 890 895

CGG ATC GCC GAA GCG ATT GCC AGT ATT CAA CTC TAC CTC AAC CGG GCA 2735
Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala
900 905 910

TTG GAA AAT CTC GAA GAA AAT GCG AAT TCG GCG GTT ATC AGC CGC CAA 2784
Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln
915 920 925

TTC TTT ATC GAC TCG GAC AAA TAC AAT AAA CCG TAC AGC ACT TCG GCG 2832
Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala
930 935 940

GGT GTT TCT CAA TTA GTT TAC TAC CCG GAA AAC TAT ATT GAT CCG ACC 2880
Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr
945 950 955 960

ATG CGT ATC GGA CAA ACC AAA ATG ATG GAC GCA TTA CTC CAA TCC GTC 2928
Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val
965 970 975

AGC CAA AGC CAA TTA AAC GCC GAT ACC CTC GAA GAT GCC TTT ATG TCT 2976
Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser
980 985 990

TAT CTC ACA TCG TTT GAA CAA CTC GCT AAT CTT AAA GTT ATT AGC GCA 3024
Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala
995 1000 1005

TAT CAC GAT AAT ATT AAT AAC GAT CAA GCG CTC ACC TAT TTT ATC GGA 3072
Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly
1010 1015 1020

CTC AGT GAA ACT GAT GCC GGT GAA TAT TAT TCG CCG AGT CTC GAT CAC 3120
Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His
1025 1030 1035 1040

AGT AAA TTC AAC GAC GGT AAA TTC GCG GCT AAT GCC TCG AGT GAA TCG 3168
Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp
1045 1050 1055

CAT AAA ATT GAT TGT CCA ATT AAC CCT TAT AAA AGC ACT ATC CGT CCA 3216
His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro
1060 1065 1070

GTG ATA TAT AAA TCC CCG CTC TAT CTC CTC TCG TTC GAA CAA AAG GAG 3264
Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu
1075 1080 1085

ATC ACC AAA CAG ACA GGA AAT AGT AAA GAT GCG TAT CAA ACT GAA ACC 3312
Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr
1090 1095 1100

GAT TAT CGT TAT GAA CTA AAA TTG GCG CAT ATC CCG TAT GAT GCG ACT 3360
Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr
1105 1110 1115 1120

TGG AAT ACC CCA ATC ACC TTT GAT CTC AAT AAA AAA ATA TCC GAG CTA 3408
Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu
1125 1130 1135

AAA CTC GAA AAA AAT AGA GCG CCC GGA CTC TAT TGT GCC GGT TAT CAA 3456
Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln
1140 1145 1150

GGT GAA GAT ACC TTG CTC CTC ATC TTT TAT AAC CAA CAA GAC ACA CTA 3504
Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu
1155 1160 1165

GAT AGT TAT AAA AAC GCT TCA ATG CAA GCA CTA TAT ATC TTT GCT GAT 3552
 Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp
 1170 1175 1180

ATG GCA TCC AAA GAT ATG ACC CCA GAA CAG AGC AAT GTT TAT CGG GAT 3600
 Met Ala Ser Lys Asp Met Thr Pro Gln Gln Ser Asn Val Tyr Arg Asp
 1185 1190 1195 1200

AAT AGC TAT CAA CAA TTT GAT ACC AAT AAT GTC AGA AGA CTG AAT AAC 3648
 Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn
 1205 1210 1215

CGC TAT GCA GAG GAT TAT GAG ATT CCT TCC TCG GTA AGT ACC COT AAA 3696
 Arg Tyr Ala Glu Asp Tyr Gln Ile Pro Ser Ser Val Ser Ser Arg Lys
 1220 1225 1230

GAC TAT GGT TCG GGA GAT TAT TAC CTC AGC ATG GTA TAT AAC GCA GAT 3744
 Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp
 1235 1240 1245

ATT CCA ACT ATC AAT TAC AAA GCC CCA TCA AGT GAT TTA AAA ATC TAT 3792
 Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr
 1250 1255 1260

ATC TCA CCA AAA TTA AGA ATT ATT CAT AAT GCA TAT GAA GGA CAG AAG 3840
 Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys
 1265 1270 1275 1280

CGC AAT CAA TGC AAT CTC ATG AAT AAA TAT GGC AAA CTA GGT GAT AAA 3888
 Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys
 1285 1290 1295

TTT ATT GTT TAT ACT AGC TTG GGG GTC AAT CCA AAT AAC TCG TCA AAT 3936
 Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
 1300 1305 1310

AAG CTC ATG TTT TAC CCC GTC TAT CAA TAT ACC GGA AAC ACC ACT GGA 3984
 Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly
 1315 1320 1325

CTC AAT CAA GGG AGA CTA CTA TTC CAC CGT GAC ACC ACT TAT CCA TCT 4032
 Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Tyr Pro Ser
 1330 1335 1340

AAA GTA GAA GCT TGG ATT CCT GGA GCA AAA CGT TCT CTA ACC AAC CAA 4080
 Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln
 1345 1350 1355 1360

AAT GCC GCC ATT GGT GAT GAT TAT GCT ACA GAC TCT CTG AAT AAA CCG 4128
 Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro
 1365 1370 1375

GAT GAT CTT AAG CAA TAT ATC TTT ATG ACT GAC AGT AAA GGG ACT GCT 4176
 Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala
 1380 1385 1390

ACT GAT GTC TCA GGC CCA GTA GAG ATT AAT ACT CCA ATT TCT CCA GCA 4224
 Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala
 1395 1400 1405

AAA GTT CAG ATA ATA GTC AAA GCC GGT GGC AAG GAG CAA ACT TTT ACC 4272
 Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr
 1410 1415 1420

GCA GAT AAA GAT GTC TCC ATT CAC CCA TCA CCT AGC TTT GAT GAA ATG 4320
 Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met
 1425 1430 1435 1440

AAT TAT CAA TTT AAT GCC CTT GAA ATA GAC GGT TCT GGT CTG AAT TTT 4368
 Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe

1445										1450										1455									
ATT AAC AAC TCA GCC AGT ATT GAT GTT ACT TTT ACC GCA TTT GCG GAG 4416	Ile Asn Asn Ser Ala Ser Ile Asp Val Thr Phe Thr Ala Phe Ala Glu	1460	1465	1470																									
GAT GGC CCC AAA CTC GGT TAT GAA AGT TTC ACT ATT CCT GTT ACC CTC 4464	Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Val Thr Leu	1475	1480	1485																									
AAC GTA AGT ACC GAT AAT GCC CTC ACC CTC CAC CAT AAT GAA AAT GGT 4512	Lys Val Ser Thr Asp Asn Ala Leu Thr Leu His His Asn Glu Asn Gly	1490	1495	1500																									
GCG CAA TAT ATG CAA TGG CAA TCC TAT CGT ACC GCG CTC AAT ACT CTA 4560	Ala Gln Tyr Met Gln Trp Gln Ser Tyr Arg Thr Arg Leu Asn Thr Leu	1505	1510	1515																									
TTT GGC CCC CAG TTG GTT GCA CCC GCG ACC ACC GGA ATC GAT ACA ATT 4608	Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile	1525	1530	1535																									
CTG AGT ATG GAA ACT CAG AAT ATT CAG GAA CCG CAG TTA GGC AAA GGT 4656	Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly	1540	1545	1550																									
TTC TAT GCT ACG TTC GTG ATA CCT CCC TAT AAC CTA TCA ACT CAT GGT 4704	Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Leu Ser Thr His Gly	1555	1560	1565																									
GAT GAA CGT TCG TTT AAG CTT TAT ATC AAA CAT GTT GTT GAT AAT AAT 4752	Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn	1570	1575	1580																									
TCA CAT ATT ATC TAT TCA GGC CAG CTA ACA GAT ACA AAT ATA AAC ATC 4800	Ser His Ile Ile Tyr Ser Gly Gln Leu Thr Asp Thr Asn Ile Asn Ile	1585	1590	1595																									
ACA TTA TTT ATT CCT CTT GAT GAT GTC CCA TTG AAT CAA GAT TAT CAC 4848	Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr His	1605	1610	1615																									
GCC AAG GTT TAT ATG ACC TTC AAG AAA TCA CCA TCA GAT GGT ACC TCG 4896	Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp	1620	1625	1630																									
TGG GGC CCT CAC TTT GTT ACA GAT GAT AAA GGA ATA GTA ACA ATA AAC 4944	Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn	1635	1640	1645																									
CCT AAA TCC ATT TTG ACC CAT TTT GAG AGC GTC AAT GTC CTC AAT AAT 4992	Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn	1650	1655	1660																									
ATT AGT AGC GAA CCA ATG GAT TTC AGC GGC GCT AAC AGC CTC TAT TTC 5040	Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe	1665	1670	1675																									
TGG GAA CTC TTC TAC TAT ACC CCG ATG CTG GTT GCT CAA CGT TTG CTG 5088	Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu	1685	1690	1695																									
CAT GAA CAG AAC TTC GAT GAA GCC AAC CGT TGG CTG AAA TAT GTC TCG 5136	His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp	1700	1705	1710																									
AGT CCA TCC GGT TAT ATT GTC CAC GGC CAG ATT CAG AAC TAC CAG TCG 5184	Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp	1715	1720	1725																									
AAC CTC CCC CCG TTA CTC GAA GAC ACC AGT TCG AAC AGT GAT CCT TTG 5232																													

Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu
 1730 1735 1740
 GAT TCC CTC GAT CCT GAC GCG GTA GCA CAG CAC GAT CCA ATG CAC TAC 5230
 Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr
 1745 1750 1755 1760
 AAA GTT TCA ACT TTT ATG CGT ACC TTG GAT CTA TTG ATA GCA GCG GCG 5328
 Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly
 1765 1770 1775
 GAC CAT GCT TAT CCG CAA CTG GAA CGA GAT ACA CTC AAC GAA GCG AAG 5376
 Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys
 1780 1785 1790
 ATG TGG TAT ATG CAA GCG CTG CAT CTA TTA GGT GAC AAA CCT TAT CTA 5424
 Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu
 1795 1800 1805
 CCG CTG ACT ACC ACA TGG AGT GAT CCA CGA CTA GAC AGA GCC GCG GAT 5472
 Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp
 1810 1815 1820
 ATC ACT ACC CAA AAT GCT CAC GAC AGC GCA ATA CTC GCT CTG CCG CAG 5520
 Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln
 1825 1830 1835 1840
 AAT ATA CCT ACA CCG GCA CCT TTA TCA 5547
 Asn Ile Pro Thr Pro Ala Pro Leu Ser
 1845 1849

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 49:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1849 амінокислот
- (B) ТИП: амінокислоти
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 49 (TcdA₄₁):

Тип	Від	До	Опис
Peptide	1	1849	TcdA ₄₁ peptide
Fragment	1	12	S2 N-terminus (SEQ ID NO:13)
Fragment	196	211	(SEQ ID NO:38)
Fragment	466	475	(SEQ ID NO:17)
Fragment	993	1004	(SEQ ID NO:23; 12/13)
Fragment	1297	1312	(SEQ ID NO:18)
Fragment	1390	1409	(SEQ ID NO:39)
Fragment	1532	1554	(SEQ ID NO:21; 19/23)

Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val
 1 5 10 15
 Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr
 20 25 30
 Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr
 35 40 45
 Tyr Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln
 50 55 60
 Gln Asn Met Asp Ile Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu
 65 70 75 80
 Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Glu Asn Tyr Thr Lys

U A 8 2 4 8 5 C 2

85										90										95									
Val	Met	Glu	Met	Leu	Ser	Thr	Phe	Arg	Pro	Ser	Gly	Ala	Thr	Pro	Tyr														
			100					105																					
His	Asp	Ala	Tyr	Glu	Asn	Val	Arg	Glu	Val	Ile	Gln	Leu	Gln	Asp	Pro														
		115					120						125																
Gly	Leu	Glu	Gln	Leu	Asn	Ala	Ser	Pro	Ala	Ile	Ala	Gly	Leu	Met	His														
		130				135						140																	
Gln	Ala	Ser	Leu	Leu	Gly	Ile	Asn	Ala	Ser	Ile	Ser	Pro	Glu	Leu	Phe														
		145			150						155				160														
Asn	Ile	Leu	Thr	Glu	Glu	Ile	Thr	Glu	Gly	Asn	Ala	Glu	Glu	Leu	Tyr														
			165					170						175															
Lys	Lys	Asn	Phe	Gly	Asn	Ile	Glu	Pro	Ala	Ser	Leu	Ala	Met	Pro	Glu														
		180						185					190																
Tyr	Leu	Lys	Arg	Tyr	Tyr	Asn	Leu	Ser	Asp	Glu	Glu	Leu	Ser	Gln	Phe														
		195				200						205																	
Ile	Gly	Lys	Ala	Ser	Asn	Phe	Gly	Gln	Gln	Glu	Tyr	Ser	Asn	Asn	Gln														
		210				215					220																		
Leu	Ile	Thr	Pro	Val	Val	Asn	Ser	Ser	Asp	Gly	Thr	Val	Lys	Val	Tyr														
		225			230						235			240															
Arg	Ile	Thr	Arg	Glu	Tyr	Thr	Thr	Asn	Ala	Tyr	Gln	Met	Asp	Val	Glu														
			245					250					255																
Leu	Phe	Pro	Phe	Gly	Gly	Glu	Asn	Tyr	Arg	Leu	Asp	Tyr	Lys	Phe	Lys														
			260					265					270																
Asn	Phe	Tyr	Asn	Ala	Ser	Tyr	Leu	Ser	Ile	Lys	Leu	Asn	Asp	Lys	Arg														
		275					280					285																	
Glu	Leu	Val	Arg	Thr	Glu	Gly	Ala	Pro	Gln	Val	Asn	Ile	Glu	Tyr	Ser														
		290				295					300																		
Ala	Asn	Ile	Thr	Leu	Asn	Thr	Ala	Asp	Ile	Ser	Gln	Pro	Phe	Glu	Ile														
		305			310						315			320															
Gly	Leu	Thr	Arg	Val	Leu	Pro	Ser	Gly	Ser	Trp	Ala	Tyr	Ala	Ala	Ala														
			325					330					335																
Lys	Phe	Thr	Val	Glu	Glu	Tyr	Asn	Gln	Tyr	Ser	Phe	Leu	Leu	Lys	Leu														
			340				345						350																
Asn	Lys	Ala	Ile	Arg	Leu	Ser	Arg	Ala	Thr	Glu	Leu	Ser	Pro	Thr	Ile														
		355					360					365																	
Leu	Glu	Gly	Ile	Val	Arg	Ser	Val	Asn	Leu	Gln	Leu	Asp	Ile	Asn	Thr														
		370				375						380																	
Asp	Val	Leu	Gly	Lys	Val	Phe	Leu	Thr	Lys	Tyr	Tyr	Met	Gln	Arg	Tyr														
		385			390						395			400															
Ala	Ile	His	Ala	Glu	Thr	Ala	Leu	Ile	Leu	Cys	Asn	Ala	Pro	Ile	Ser														
			405					410					415																
Gln	Arg	Ser	Tyr	Asp	Asn	Gln	Pro	Ser	Gln	Phe	Asp	Arg	Leu	Phe	Asn														
			420				425					430																	
Thr	Pro	Leu	Leu	Asn	Gly	Gln	Tyr	Phe	Ser	Thr	Gly	Asp	Glu	Glu	Ile														
		435				440						445																	
Asp	Leu	Asn	Ser	Gly	Ser	Thr	Gly	Asp	Trp	Arg	Lys	Thr	Ile	Leu	Lys														
		450				455					460																		

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Arg Ala Phe Asn Ile Asp Asp Val Ser Leu Phe Arg Leu Leu Lys Ile
465 470 475 480

Thr Asp His Asp Asn Lys Asp Gly Lys Ile Lys Asn Asn Leu Lys Asn
485 490 495

Leu Ser Asn Leu Tyr Ile Gly Lys Leu Leu Ala Asp Ile His Gln Leu
500 505 510

Thr Ile Asp Glu Leu Asp Leu Leu Leu Ile Ala Val Gly Glu Gly Lys
515 520 525

Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Thr Leu Ile Arg
530 535 540

Lys Leu Asn Thr Ile Thr Ser Trp Leu His Thr Gln Lys Trp Ser Val
545 550 555 560

Phe Gln Leu Phe Ile Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr
565 570 575

Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly
580 585 590

Phe Asp Lys Asp Lys Ala Asp Leu Leu His Val Met Ala Pro Tyr Ile
595 600 605

Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu
610 615 620

Leu Trp Ala Asp Lys Leu Gln Pro Gly Asp Gly Ala Met Thr Ala Glu
625 630 635 640

Lys Phe Trp Asp Trp Leu Asn Thr Lys Tyr Thr Pro Gly Ser Ser Glu
645 650 655

Ala Val Glu Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala
660 665 670

Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe
675 680 685

Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ala Ala Thr Gly Ala
690 695 700

Ala Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala
705 710 715 720

Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala
725 730 735

Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn
740 745 750

Leu Asp Ala Asn Leu Leu Leu Gln Ala Ser Ile Gln Ala Gln Asn His
755 760 765

Gln His Leu Pro Pro Val Thr Pro Glu Asn Ala Phe Ser Cys Trp Thr
770 775 780

Ser Ile Asn Thr Ile Leu Gln Trp Val Asn Val Ala Gln Gln Leu Asn
785 790 795 800

Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp Tyr Ile Gln
805 810 815

Ser Met Lys Glu Thr Pro Thr Tyr Ala Gln Trp Glu Asn Ala Ala Gly
820 825 830

Val Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asn Thr Leu His Ala
835 840 845

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr Tyr Ile Arg
 350 355 360
 Gln Val Ala Lys Ala Ala Ala Ala Ile Lys Ser Arg Asp Asp Leu Tyr
 365 370 375 380
 Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile Lys Thr Thr
 385 390 395
 Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val Asn Arg Ala
 900 905 910
 Leu Glu Asn Val Glu Glu Asn Ala Asn Ser Gly Val Ile Ser Arg Gln
 915 920 925
 Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser Thr Trp Ala
 930 935 940
 Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile Asp Pro Thr
 945 950 955 960
 Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln Ser Val
 965 970 975
 Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe Met Ser
 980 985 990
 Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile Ser Ala
 995 1000 1005
 Tyr His Asp Asn Ile Asn Asn Asp Gln Gly Leu Thr Tyr Phe Ile Gly
 1010 1015 1020
 Leu Ser Glu Thr Asp Ala Gly Glu Tyr Tyr Trp Arg Ser Val Asp His
 1025 1030 1035 1040
 Ser Lys Phe Asn Asp Gly Lys Phe Ala Ala Asn Ala Trp Ser Glu Trp
 1045 1050 1055
 His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Lys Ser Thr Ile Arg Pro
 1060 1065 1070
 Val Ile Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln Lys Glu
 1075 1080 1085
 Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln Thr Glu Thr
 1090 1095 1100
 Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr
 1105 1110 1115 1120
 Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Lys Lys Ile Ser Glu Leu
 1125 1130 1135
 Lys Leu Glu Lys Asn Arg Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln
 1140 1145 1150
 Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu
 1155 1160 1165
 Asp Ser Tyr Lys Asn Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp
 1170 1175 1180
 Met Ala Ser Lys Asp Met Thr Pro Glu Gln Ser Asn Val Tyr Arg Asp
 1185 1190 1195 1200
 Asn Ser Tyr Gln Gln Phe Asp Thr Asn Asn Val Arg Arg Val Asn Asn
 1205 1210 1215
 Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Ser Ser Arg Lys
 1220 1225 1230

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Asp Tyr Gly Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp
1235 1240 1245

Ile Pro Thr Ile Asn Tyr Lys Ala Ala Ser Ser Asp Leu Lys Ile Tyr
1250 1255 1260

Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Lys
1265 1270 1275 1280

Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys
1285 1290 1295

Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn
1300 1305 1310

Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Ser Gly Asn Thr Ser Gly
1315 1320 1325

Leu Asn Gln Gly Arg Leu Leu Phe His Arg Asp Thr Thr Tyr Pro Ser
1330 1335 1340

Lys Val Glu Ala Trp Ile Pro Gly Ala Lys Arg Ser Leu Thr Asn Gln
1345 1350 1355 1360

Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro
1365 1370 1375

Asp Asp Leu Lys Gln Tyr Ile Phe Met Thr Asp Ser Lys Gly Thr Ala
1380 1385 1390

Thr Asp Val Ser Gly Pro Val Glu Ile Asn Thr Ala Ile Ser Pro Ala
1395 1400 1405

Lys Val Gln Ile Ile Val Lys Ala Gly Gly Lys Glu Gln Thr Phe Thr
1410 1415 1420

Ala Asp Lys Asp Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met
1425 1430 1435 1440

Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Gly Leu Asn Phe
1445 1450 1455

Ile Asn Asn Ser Ala Ser Ile Asp Val Thr Phe Thr Ala Phe Ala Glu
1460 1465 1470

Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Val Thr Leu
1475 1480 1485

Lys Val Ser Thr Asp Asn Ala Leu Thr Leu His His Asn Glu Asn Gly
1490 1495 1500

Ala Gln Tyr Met Gln Trp Gln Ser Tyr Arg Thr Arg Leu Asn Thr Leu
1505 1510 1515 1520

Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile
1525 1530 1535

Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly
1540 1545 1550

Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Leu Ser Thr His Gly
1555 1560 1565

Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn
1570 1575 1580

Ser His Ile Ile Tyr Ser Gly Gln Leu Thr Asp Thr Asn Ile Asn Ile
1585 1590 1595 1600

Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr His

U A 8 2 4 8 5 C 2

1605	1610	1615
Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp 1620 1625 1630		
Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn 1635 1640 1645		
Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn 1650 1655 1660		
Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe 1665 1670 1675 1680		
Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu 1685 1690 1695		
His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp 1700 1705 1710		
Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp 1715 1720 1725		
Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu 1730 1735 1740		
Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr 1745 1750 1755 1760		
Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly 1765 1770 1775		
Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys 1780 1785 1790		
Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu 1795 1800 1805		
Pro Leu Ser Thr Thr Trp Ser Asp Pro Arg Leu Asp Arg Ala Ala Asp 1810 1815 1820		
Ile Thr Thr Gln Asn Ala His Asp Ser Ala Ile Val Ala Leu Arg Gln 1825 1830 1835 1840		
Asn Ile Pro Thr Pro Ala Pro Leu Ser 1845 1849		

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 50:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1740 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лівійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 50 (кодуючий район TcdA₁₁₁):

TTG CGC AGC GCT AAT ACC CTC ACT CAT CTC TTC CTC CCG CAA ATC AAT 48
Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn
1 5 10 15
GAA CTC ATC ATG AAT TAC TCG CAG ACA TTA GCT CAG AGA GTA TAC AAT 96
Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn
20 25 30
CTG CGT CAT AAC CTC TCT ATC GAC GGC CAG CCG TTA TAT CTG CCA ATC 144

Leu	Arg	His	Asn	Leu	Ser	Ile	Asp	Gly	Gln	Pro	Leu	Tyr	Leu	Pro	Ile		
		35					40					45					
TAT	GCC	ACA	CCG	GCC	GAT	CCG	AAA	GCG	TTA	CTC	AGC	GCC	GCC	GTT	GCC	152	
Tyr	Ala	Thr	Pro	Ala	Asp	Pro	Lys	Ala	Leu	Leu	Ser	Ala	Ala	Val	Ala		
	50					55					50						
ACT	TCT	CAA	GCT	GGA	GCC	AAG	CTA	GCG	GAA	TCA	TTT	ATG	TCC	CTG	TGG	240	
Thr	Ser	Gln	Gly	Gly	Gly	Lys	Leu	Pro	Glu	Ser	Phe	Met	Ser	Leu	Trp		
	55				70				75					30			
CGT	TTC	CCG	CAC	ATG	CTG	GAA	AAT	GCG	GCG	GCG	ATG	GTT	AGC	CAG	CTC	288	
Arg	Phe	Pro	His	Met	Leu	Glu	Asn	Ala	Arg	Gly	Met	Val	Ser	Gln	Leu		
			85					90						95			
ACC	CAG	TTC	GCC	TCC	ACG	TTA	CAA	AAT	ATT	ATC	GAA	GCT	CAG	GAC	GCG	336	
Thr	Gln	Phe	Gly	Ser	Thr	Leu	Gln	Asn	Ile	Ile	Glu	Arg	Gln	Asp	Ala		
			100					105					110				
GAA	GCG	CTC	AAT	GCG	TTA	TTA	CAA	AAT	CAG	GCC	GCC	GAG	CTG	ATA	TTG	384	
Glu	Ala	Leu	Asn	Ala	Leu	Leu	Gln	Asn	Gln	Ala	Ala	Glu	Leu	Ile	Leu		
		115					120					125					
ACT	AAC	CTG	AGC	ATT	CAG	GAC	AAA	ACC	ATT	GAA	GAA	TTG	GAT	GCC	GAG	432	
Thr	Asn	Leu	Ser	Ile	Gln	Asp	Lys	Thr	Ile	Glu	Glu	Leu	Asp	Ala	Glu		
		130				135					140						
AAA	ACG	GTG	TTG	GAA	AAA	TCC	AAA	GCG	GGA	GCA	CAA	TCG	CGC	TTT	GAT	480	
Lys	Thr	Val	Leu	Glu	Lys	Ser	Lys	Ala	Gly	Ala	Gln	Ser	Arg	Phe	Asp		
		145			150				155					160			
AGC	TAC	GGC	AAA	CTG	TAC	GAT	GAG	AAT	ATC	AAC	GCC	GGT	GAA	AAC	CAA	528	
Ser	Tyr	Gly	Lys	Leu	Tyr	Asp	Glu	Asn	Ile	Asn	Ala	Gly	Glu	Asn	Gln		
			165					170						175			
GCC	ATG	ACG	CTA	CGA	GCG	TCC	GCC	GCC	GCG	CTT	ACC	ACG	GCA	GTT	CAG	576	
Ala	Met	Thr	Leu	Arg	Ala	Ser	Ala	Ala	Gly	Leu	Thr	Thr	Ala	Val	Gln		
			180					185					190				
GCA	TCC	GCT	CTG	GCC	GCT	GCG	GCG	GCT	GAT	CTG	GTG	CCT	AAC	ATC	TTC	624	
Ala	Ser	Arg	Leu	Ala	Gly	Ala	Ala	Ala	Asp	Leu	Val	Pro	Asn	Ile	Phe		
		195				200						205					
GGC	TTT	GCC	GCT	GCG	GCG	AGC	GCT	TGG	GCG	GCT	ATC	GCT	GAG	GCG	ACA	672	
Gly	Phe	Ala	Gly	Gly	Gly	Ser	Arg	Trp	Gly	Ala	Ile	Ala	Glu	Ala	Thr		
	210					215					220						
GGT	TAT	GTG	ATG	GAA	TTC	TCC	GCG	AAT	GTT	ATG	AAC	ACC	GAA	GCG	GAT	720	
Gly	Tyr	Val	Met	Glu	Phe	Ser	Ala	Asn	Val	Met	Asn	Thr	Glu	Ala	Asp		
	225				230					235				240			
AAA	ATT	AGC	CAA	TCT	GAA	ACC	TAC	GCT	GCT	GCG	GCT	CAG	GAG	TGG	GAG	768	
Lys	Ile	Ser	Gln	Ser	Glu	Thr	Tyr	Arg	Arg	Arg	Arg	Gln	Glu	Trp	Glu		
			245					250						255			
ATC	CAG	CGG	AAT	AAT	GCC	GAA	GCG	GAA	TTG	AAG	CAA	ATC	GAT	GCT	CAG	816	
Ile	Gln	Arg	Asn	Asn	Ala	Glu	Ala	Glu	Leu	Lys	Gln	Ile	Asp	Ala	Gln		
			260					265					270				
CTC	AAA	TCA	CTC	GCT	GTA	GCG	GCG	GAA	GCC	GCC	GTA	TTG	CAG	AAA	ACC	864	
Leu	Lys	Ser	Leu	Ala	Val	Arg	Arg	Glu	Ala	Ala	Val	Leu	Gln	Lys	Thr		
		275				280						285					
AGT	CTG	AAA	ACC	CAA	CAA	GAA	CAG	ACC	CAA	TCT	CAA	TTG	GCC	TTC	CTG	912	
Ser	Leu	Lys	Thr	Gln	Gln	Glu	Gln	Thr	Gln	Ser	Gln	Leu	Ala	Phe	Leu		
		290				295					300						
CAA	GCT	AAG	TTC	AGC	AAT	CAG	GCG	TTA	TAC	AAC	TCG	CTG	GCT	GCT	GCA	960	
Gln	Arg	Lys	Phe	Ser	Asn	Gln	Ala	Leu	Tyr	Asn	Trp	Leu	Arg	Gly	Arg		
					310					315					320		

CTG GCG GCG ATT TAC TTC CAG TTC TAC GAT TTG CCC CTC GCG CGT TCG 1008
 Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg Cys
 325 330 335

CTG ATG GCA GAA CAA GCT TAC CGT TCG GAA CTC AAT GAT GAC TCT GCC 1056
 Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser Ala
 340 345 350

CGC TTC ATT AAA CCG GCG GCC TCG CAG GGA ACC TAT GCC GGT CTG CTT 1104
 Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365

GCA GGT GAA ACC TTG ATG CTG AGT CTG GCA CAA ATG GAA GAC GCT CAT 1152
 Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala His
 370 375 380

CTG AAA CCG GAT AAA CCG GCA TTA GAG GTT GAA CCC ACA GTA TCG CTG 1200
 Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Arg Arg Thr Val Ser Leu
 385 390 395 400

GCC GAA GTT TAT GCA GGA TTA CCA AAA GAT AAC GGT CCA TTT TCC CTG 1248
 Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser Leu
 405 410 415

GCT CAG GAA ATT GAC AAG CTG GTG AGT CAA GGT TCA GCG AGT GCC GCG 1296
 Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala Gly
 420 425 430

AGT GGT AAT AAT AAT TTG GCG TTC GGC GCC GGC ACC GAC ACT AAA ACC 1344
 Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Thr Lys Thr
 435 440 445

TCT TTG CAG GCA TCA GTT TCA TTC GCT GAT TTG AAA ATT CGT GAA GAT 1392
 Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu Asp
 450 455 460

TAC CCG GCA TCG CTT GCG AAA ATT CGA CGT ATC AAA CAG ATC AGC GTC 1440
 Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser Val
 465 470 475 480

ACT TTG CCC CCG CTA CTG GGA CCG TAT CAG GAT GTA CAG GCA ATA TTG 1488
 Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile Leu
 485 490 495

TCT TAC GCG GAT AAA GCG GGA TTA GCT AAC GCG TGT GAA GCG CTG GCA 1536
 Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu Ala
 500 505 510

GTT TCT CAC GGT ATG AAT GAC AGC GCG CAA TTC CAG CTC GAT TTC AAC 1584
 Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn
 515 520 525

GAT GCG AAA TTC CTG CCA TTC GAA GCG ATC GCC ATT GAT CAA GCG ACC 1632
 Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly Thr
 530 535 540

CTG ACA CTG AGC TTC CCA AAT GCA TCT ATG CCG GAG AAA GGT AAA GAA 1680
 Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys Gln
 545 550 555 560

GCC ACT ATG TTA AAA ACC CTG AAC GAT ATC ATT TTG CAT ATT CCG TAC 1728
 Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg Tyr
 565 570 575

ACC ATT AAA TAA 1740
 Thr Ile Lys ...
 579

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 51:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 579 амінокислот
(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 51 (TcdA₁₁₁):

Leu Arg Ser Ala Asn Thr Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn
1 5 10 15
Glu Val Met Met Asn Tyr Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn
20 25 30
Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Tyr Leu Pro Ile
35 40 45
Tyr Ala Thr Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ala
50 55 60
Thr Ser Gln Gly Gly Gly Lys Leu Pro Glu Ser Phe Met Ser Leu Trp
65 70 75 80
Arg Phe Pro His Met Leu Glu Asn Ala Arg Gly Met Val Ser Gln Leu
85 90 95
Thr Gln Phe Gly Ser Thr Leu Gln Asn Ile Ile Glu Arg Gln Asp Ala
100 105 110
Glu Ala Leu Asn Ala Leu Leu Gln Asn Gln Ala Ala Glu Leu Ile Leu
115 120 125
Thr Asn Leu Ser Ile Gln Asp Lys Thr Ile Glu Glu Leu Asp Ala Glu
130 135 140
Lys Thr Val Leu Glu Lys Ser Lys Ala Gly Ala Gln Ser Arg Phe Asp
145 150 155 160
Ser Tyr Gly Lys Leu Tyr Asp Glu Asn Ile Asn Ala Gly Glu Asn Gln
165 170 175
Ala Met Thr Leu Arg Ala Ser Ala Ala Gly Leu Thr Thr Ala Val Gln
180 185 190
Ala Ser Arg Leu Ala Gly Ala Ala Ala Asp Leu Val Pro Asn Ile Phe
195 200 205
Gly Phe Ala Gly Gly Gly Ser Arg Trp Gly Ala Ile Ala Glu Ala Thr
210 215 220
Gly Tyr Val Met Glu Phe Ser Ala Asn Val Met Asn Thr Glu Ala Asp
225 230 235 240
Lys Ile Ser Gln Ser Glu Thr Tyr Arg Arg Arg Arg Gln Glu Trp Glu
245 250 255
Ile Gln Arg Asn Asn Ala Glu Ala Glu Leu Lys Gln Ile Asp Ala Gln
260 265 270
Leu Lys Ser Leu Ala Val Arg Arg Glu Ala Ala Val Leu Gln Lys Thr
275 280 285
Ser Leu Lys Thr Gln Gln Glu Gln Thr Gln Ser Gln Leu Ala Phe Leu
290 295 300
Gln Arg Lys Phe Ser Asn Gln Ala Leu Tyr Asn Trp Leu Arg Gly Arg
305 310 315 320

Leu Ala Ala Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ala Arg Cys
 325 330 335
 Leu Met Ala Glu Gln Ala Tyr Arg Trp Glu Leu Asn Asp Asp Ser Ala
 340 345 350
 Arg Phe Ile Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365
 Ala Gly Glu Thr Leu Met Leu Ser Leu Ala Gln Met Glu Asp Ala His
 370 375 380
 Leu Lys Arg Asp Lys Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400
 Ala Glu Val Tyr Ala Gly Leu Pro Lys Asp Asn Gly Pro Phe Ser Leu
 405 410 415
 Ala Gln Glu Ile Asp Lys Leu Val Ser Gln Gly Ser Gly Ser Ala Gly
 420 425 430
 Ser Gly Asn Asn Asn Leu Ala Phe Gly Ala Gly Thr Asp Thr Lys Thr
 435 440 445
 Ser Leu Gln Ala Ser Val Ser Phe Ala Asp Leu Lys Ile Arg Glu Asp
 450 455 460
 Tyr Pro Ala Ser Leu Gly Lys Ile Arg Arg Ile Lys Gln Ile Ser Val
 465 470 475 480
 Thr Leu Pro Ala Leu Leu Gly Pro Tyr Gln Asp Val Gln Ala Ile Leu
 485 490 495
 Ser Tyr Gly Asp Lys Ala Gly Leu Ala Asn Gly Cys Glu Ala Leu Ala
 500 505 510
 Val Ser His Gly Met Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn
 515 520 525
 Asp Gly Lys Phe Leu Pro Phe Glu Gly Ile Ala Ile Asp Gln Gly Thr
 530 535 540
 Leu Thr Leu Ser Phe Pro Asn Ala Ser Met Pro Glu Lys Gly Lys Gln
 545 550 555 560
 Ala Thr Met Leu Lys Thr Leu Asn Asp Ile Ile Leu His Ile Arg Tyr
 565 570 575
 Thr Ile Lys ...
 579

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 52:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 5532 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 52 (кодуючий район TcdA₁₁₁):

TTT ATA CAA GGT TAT ACT GAT CTC TTT GGT AAT CGT GCT GAT AAC TAT 48
 Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala Asp Asn Tyr
 1 5 10 15
 CCC GCG CCG GCG TCG GTT CCA TCG ATC TTC TCA CCG GCG GCT TAT TCG 96

U A 8 2 4 8 5 C 2

Ala Ala Pro Gly Ser Val Ala Ser Met Phe Ser Pro Ala Ala Tyr Leu
20 25 30

ACC GAA TTG TAC COT GAA GCC AAA AAC TTG CAT GAC AGC AGC TCA ATT 144
Thr Glu Leu Tyr Arg Glu Ala Lys Asn Leu His Asp Ser Ser Ser Ile
35 40 45

TAT TAC CTA GAT AAA CGT GGC CCG GAT TTA GCA AGC TTA ATG CTC AGC 192
Tyr Tyr Leu Asp Lys Arg Arg Pro Asp Leu Ala Ser Leu Met Leu Ser
50 55 60

CAG AAA AAT ATG GAT GAG GAA ATT TCA ACG CTG GCT CTC TCT AAT GAA 240
Gln Lys Asn Met Asp Glu Glu Ile Ser Thr Leu Ala Leu Ser Asn Glu
65 70 75 80

TTG TGC CTT GCC GGG ATC GAA ACA AAA ACA GCA AAA TCA CAA GAT GAA 288
Leu Cys Leu Ala Gly Ile Glu Thr Lys Thr Gly Lys Ser Gln Asp Glu
85 90 95

GTG ATG GAT ATG TTG TCA ACT TAT COT TTA AGT GGA GAG ACA CCT TAT 336
Val Met Asp Met Leu Ser Thr Tyr Arg Leu Ser Gly Glu Thr Pro Tyr
100 105 110

CAT CAC GCT TAT GAA ACT GTT COT GAA ATC GTT CAT GAA CGT GAT CCA 384
His His Ala Tyr Glu Thr Val Arg Glu Ile Val His Glu Arg Asp Pro
115 120 125

GGA TTT COT CAT TTG TCA CAG GCA CCC ATT GTT GCT GCT AAG CTC GAT 432
Gly Phe Arg His Leu Ser Gln Ala Pro Ile Val Ala Ala Lys Leu Asp
130 135 140

CCT GTG ACT TTG TTG GGT ATT AGC TCC CAT ATT TCG CCA GAA CTG TAT 480
Pro Val Thr Leu Leu Gly Ile Ser Ser His Ile Ser Pro Glu Leu Tyr
145 150 155 160

AAC TTG CTG ATT GAG GAG ATC CCG GAA AAA GAT GAA GCC GCG CTT GAT 528
Asn Leu Leu Ile Glu Glu Ile Pro Glu Lys Asp Glu Ala Ala Leu Asp
165 170 175

ACG CTT TAT AAA ACA AAC TTT GGC GAT ATT ACT ACT GCT CAG TTA ATG 576
Thr Leu Tyr Lys Thr Asn Phe Gly Asp Ile Thr Thr Ala Gln Leu Met
180 185 190

TCC CCA AGT TAT CTG GCC CCG TAT TAT GGC GTC TCA CCG GAA GAT ATT 624
Ser Pro Ser Tyr Leu Ala Arg Tyr Tyr Gly Val Ser Pro Glu Asp Ile
195 200 205

GCC TAC GTG ACG ACT TCA TTA TCA CAT GTT GGA TAT AGC AGT GAT ATT 672
Ala Tyr Val Thr Thr Ser Leu Ser His Val Gly Tyr Ser Ser Asp Ile
210 215 220

CTG GTT ATT CCG TTG GTC GAT GGT GTG GGT AAG ATG GAA GTA GTT CGT 720
Leu Val Ile Pro Leu Val Asp Gly Val Gly Lys Met Glu Val Val Arg
225 230 235 240

GTT ACC CGA ACA CCA TCG GAT AAT TAT ACC AGT CAG ACC AAT TAT ATT 768
Val Thr Arg Thr Pro Ser Asp Asn Tyr Thr Ser Gln Thr Asn Tyr Ile
245 250 255

GAG CTG TAT CCA CAG GGT GGC GAC AAT TAT TTG ATC AAA TAC AAT CTA 816
Glu Leu Tyr Pro Gln Gly Gly Asp Asn Tyr Leu Ile Lys Tyr Asn Leu
260 265 270

AGC AAT AGT TTT GGT TTG CAT GAT TTT TAT CTG CAA TAT AAA GAT GGT 864
Ser Asn Ser Phe Gly Leu Asp Asp Phe Tyr Leu Gln Tyr Lys Asp Gly
275 280 285

TCC GCT GAT TCG ACT GAG ATT GCC CAT AAT CCC TAT CCT GAT ATG CTC 912
Ser Ala Asp Trp Thr Glu Ile Ala His Asn Pro Tyr Pro Asp Met Val
290 295 300

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

ATA AAT CAA AAG TAT GAA TCA CAG GCG ACA ATC AAA CGT AGT GAC TTT 350
 Ile Asn Gln Lys Tyr Glu Ser Gln Ala Thr Ile Lys Arg Ser Asp Ser
 305 310 315 320

GAC AAT ATA CTC AGT ATA GCG TTA CAA AGA TCG CAT AGC CGT AGT TAT 1008
 Asp Asn Ile Leu Ser Ile Gly Leu Gln Arg Trp His Ser Gly Ser Tyr
 325 330 335

AAT TTT GCC GCC GCC AAT TTT AAA ATT GAC CAA TAC TCC CCG AAA GCT 1056
 Asn Phe Ala Ala Ala Asn Phe Lys Ile Asp Gln Tyr Ser Pro Lys Ala
 340 345 350

TTC CTG CTT AAA ATG AAT AAG GCT ATT CCG TTG CTC AAA GCT ACC GGC 1104
 Phe Leu Leu Lys Met Asn Lys Ala Ile Arg Leu Leu Lys Ala Thr Gly
 355 360 365

CTC TCT TTT GCT ACC TTG GAG CGT ATT GTT GAT AGT GTT AAT AGC ACC 1152
 Leu Ser Phe Ala Thr Leu Glu Arg Ile Val Asp Ser Val Asn Ser Thr
 370 375 380

AAA TCC ATC ACC GTT GAG GTA TTA AAC AAG GTT TAT CCG GTA AAA TTC 1200
 Lys Ser Ile Thr Val Glu Val Leu Asn Lys Val Tyr Arg Val Lys Phe
 385 390 395 400

TAT ATT GAT CGT TAT GCC ATC AGT GAA GAG ACA GCC CCT ATT TTG GCT 1248
 Tyr Ile Asp Arg Tyr Gly Ile Ser Glu Glu Thr Ala Ala Ile Leu Ala
 405 410 415

AAT ATT AAT ATC TCT CAG CAA GCT GTT CCG AAT CAG CTT AGC CAG TTT 1296
 Asn Ile Asn Ile Ser Gln Gln Ala Val Gly Asn Gln Leu Ser Gln Phe
 420 425 430

GAG CAA CTA TTT AAT CAC CCG CCG CTC AAT GGT ATT CCG TAT GAA ATC 1344
 Glu Gln Leu Phe Asn His Pro Pro Leu Asn Gly Ile Arg Tyr Glu Ile
 435 440 445

AGT GAG GAC AAC TCC AAA CAT CTT CCT AAT CCT GAT CTG AAC CTT AAA 1392
 Ser Glu Asp Asn Ser Lys His Leu Pro Asn Pro Asp Leu Asn Leu Lys
 450 455 460

CCA GAC AGT ACC GGT GAT GAT CAA CCG AAG GCG GTT TTA AAA CCG GCG 1440
 Pro Asp Ser Thr Gly Asp Asp Gln Arg Lys Ala Val Leu Lys Arg Ala
 465 470 475 480

TTT CAG GTT AAC GCC AGT GAG TTG TAT CAG ATG TTA TTG ATC ACT GAT 1488
 Phe Gln Val Asn Ala Ser Glu Leu Tyr Gln Met Leu Leu Ile Thr Asp
 485 490 495

CCT AAA GAA GAC GGT GTT ATC AAA AAT AAC TTA GAG AAT TTG TCT GAT 1536
 Arg Lys Glu Asp Gly Val Ile Lys Asn Asn Leu Glu Asn Leu Ser Asp
 500 505 510

CTG TAT TTG GTT AGT TTG CTG GCC CAG ATT CAT AAC CTG ACT ATT GCT 1584
 Leu Tyr Leu Val Ser Leu Leu Ala Gln Ile His Asn Leu Thr Ile Ala
 515 520 525

GAA TTG AAC ATT TTG TTG GTG ATT TGT GCG TAT GCG GAC ACC AAC ATT 1632
 Glu Leu Asn Ile Leu Leu Val Ile Cys Gly Tyr Gly Asp Thr Asn Ile
 530 535 540

TAT CAG ATT ACC GAC GAT AAT TTA GCC AAA ATA GTG GAA ACA TTG TTG 1680
 Tyr Gln Ile Thr Asp Asp Asn Leu Ala Lys Ile Val Glu Thr Leu Leu
 545 550 555 560

TGG ATC ACT CAA TGG TTG AAG ACC CAA AAA TGG ACA GTT ACC GAC CTG 1728
 Trp Ile Thr Gln Trp Leu Lys Thr Gln Lys Trp Thr Val Thr Asp Leu
 565 570 575

TTT CTG ATG ACC ACC GCC ACT TAC AGC ACC ACT TTA ACC CCA GAA ATT 1776
 Phe Leu Met Thr Thr Ala Thr Tyr Ser Thr Thr Leu Thr Pro Glu Ile
 580 585 590

U A 8 2 4 8 5 C 2

ACC AAT CTC ACC GCT ACC TTG TCT TCA ACT TTG CAT GGC AAA CAG ACT 1824
 Ser Asn Leu Thr Ala Thr Leu Ser Ser Thr Leu His Gly Lys Glu Ser
 595 600 605

CTG ATT GCG GAA GAT CTG AAA ACA GCA ATG GCG CCT TGC TTC ACT TCG 1872
 Leu Ile Gly Glu Asp Leu Lys Arg Ala Met Ala Pro Cys Phe Thr Ser
 610 615 620

OCT TTG CAT TTG ACT TCT CAA GAA GTT GCG TAT GAC CTG CTG TTG TGG 1920
 Ala Leu His Leu Thr Ser Gln Glu Val Ala Tyr Asp Leu Leu Leu Trp
 625 630 635 640

ATA GAC CAG ATT CAA CCG GCA CAA ATA ACT GTT GAT GCG TTT TCG GAA 1968
 Ile Asp Gln Ile Gln Pro Ala Gln Ile Thr Val Asp Gly Phe Trp Glu
 645 650 655

GAA CTG CAA ACA ACA CCA ACC AGC TTG AAG CTG ATT ACC TTT GCT CAG 2016
 Glu Val Gln Thr Thr Pro Thr Ser Leu Lys Val Ile Thr Phe Ala Gln
 660 665 670

JTG CTG GCA CAA TTG AGC CTG ATC TAT CGT CGT ATT GCG TTA AGT GAA 2064
 Val Leu Ala Gln Leu Ser Leu Ile Tyr Arg Arg Ile Gly Leu Ser Glu
 675 680 685

ACC GAA CTG TCA CTG ATC GTG ACT CAA TCT TCT CTG CTA GTG GCA GGC 2112
 Thr Glu Leu Ser Leu Ile Val Thr Gln Ser Ser Leu Leu Val Ala Gly
 690 695 700

AAA AGC ATA CTG GAT CAC GGT CTG TTA ACC CTG ATG GCC TTG GAA GGT 2160
 Lys Ser Ile Leu Asp His Gly Leu Leu Thr Leu Met Ala Leu Glu Gly
 705 710 715 720

TTT CAT ACC TGG GTT AAT GGC TTG GCG CAA CAT GCC TCC TTG ATA TTG 2208
 Phe His Thr Trp Val Asn Gly Leu Gly Gln His Ala Ser Leu Ile Leu
 725 730 735

GCG GCG TTG AAA GAC GGA GCC TTG ACA GTT ACC GAT GTA GCA CAA GCT 2256
 Ala Ala Leu Lys Asp Gly Ala Leu Thr Val Thr Asp Val Ala Gln Ala
 740 745 750

ATG AAT AAG GAG GAA TCT CTC CTA CAA ATG GCA GCT AAT CAG GTG GAG 2304
 Met Asn Lys Glu Glu Ser Leu Leu Gln Met Ala Ala Asn Gln Val Glu
 755 760 765

AAG GAT CTA ACA AAA CTG ACC AGT TCG ACA CAG ATT GAC GCT ATT CTG 2352
 Lys Asp Leu Thr Lys Leu Thr Ser Trp Thr Gln Ile Asp Ala Ile Leu
 770 775 780

CAA TGG TTA CAG ATG TCT TCG GCC TTG GCG GTT TCT CCA CTG GAT CTG 2400
 Gln Trp Leu Gln Met Ser Ser Ala Leu Ala Val Ser Pro Leu Asp Leu
 785 790 795 800

GCA GCG ATG ATG GCC CTG AAA TAT GCG ATA GAT CAT AAC TAT GCT GCC 2448
 Ala Gly Met Met Ala Leu Lys Tyr Gly Ile Asp His Asn Tyr Ala Ala
 805 810 815

TGG CAA GCT GCG GCG GCT GCG CTG ATG GCT GAT CAT GCT AAT CAG GCA 2496
 Trp Gln Ala Ala Ala Ala Leu Met Ala Asp His Ala Asn Gln Ala
 820 825 830

CAG AAA AAA CTG GAT GAG ACG TTC AGT AAG GCA TTA TGT AAC TAT TAT 2544
 Gln Lys Lys Leu Asp Glu Thr Phe Ser Lys Ala Leu Cys Asn Tyr Tyr
 835 840 845

ATT AAT GCT GTT GTC GAT AGT GCT GCT GGA GTA CGT GAT GCT AAC GCT 2592
 Ile Asn Ala Val Val Asp Ser Ala Ala Gly Val Arg Asp Arg Asn Gly
 850 855 860

TTA TAT ACC TAT TTG CTG ATT GAT AAT CAG GTT TCT GCC GAT GTC ATC 2640
 Leu Tyr Thr Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Asp Val Ile

U. 82485 C2

-170-

U A 8 2 4 8 5 C 2

Thr	Asp	Asn	Asn	Ala	Pro	Val	Thr	Gly	Leu	Tyr	Ile	Phe	Ala	Asp	Met		
		1135						1130					1165				
TCA	TCA	GAC	AAT	ATG	ACC	AAT	CCA	CAA	GCA	ACT	AAC	TAT	TGG	AAT	AAC	1552	
Ser	Ser	Asp	Asn	Met	Thr	Asn	Ala	Gln	Ala	Thr	Asn	Tyr	Trp	Asn	Asn		
		1170				1175					1130						
AGT	TAT	CCG	CAA	TTT	GAT	ACT	GTG	ATG	GCA	GAT	CCG	GAT	ACC	GAC	AAT	1600	
Ser	Tyr	Pro	Gln	Phe	Asp	Thr	Val	Met	Ala	Asp	Pro	Asp	Ser	Asp	Asn		
		1185				1190				1195					1200		
AAA	AAA	GTC	ATA	ACC	AGA	ACA	GTT	AAT	AAC	CCT	TAT	CCG	GAG	GAT	TAT	1648	
Lys	Lys	Val	Ile	Thr	Arg	Arg	Val	Asn	Asn	Arg	Tyr	Ala	Glu	Asp	Tyr		
				1205					1210					1215			
GAA	ATT	CCT	TCC	TCT	GTG	ACA	AGT	AAC	AGT	AAT	TAT	TCT	TGG	GCT	GAT	1696	
Glu	Ile	Pro	Ser	Ser	Val	Thr	Ser	Asn	Ser	Asn	Tyr	Ser	Trp	Gly	Asp		
			1220					1225					1230				
CAC	AGT	TTA	ACC	ATG	CTT	TAT	GGT	GGT	AGT	GTT	CCT	AAT	ATT	ACT	TTT	1744	
His	Ser	Leu	Thr	Met	Leu	Tyr	Gly	Gly	Ser	Val	Pro	Asn	Ile	Thr	Phe		
		1235					1240					1245					
GAA	TCC	CCG	CCA	GAA	GAT	TTA	AGG	CTA	TCT	ACC	AAT	ATC	GCA	TTG	ACT	1792	
Glu	Ser	Ala	Ala	Glu	Asp	Leu	Arg	Leu	Ser	Thr	Asn	Met	Ala	Leu	Ser		
		1250				1255					1260						
ATT	ATT	CAT	AAT	GGA	TAT	GCG	GGA	ACC	CGC	CGT	ATA	CAA	TGT	AAT	CTT	1840	
Ile	Ile	His	Asn	Gly	Tyr	Ala	Gly	Thr	Arg	Arg	Ile	Gln	Cys	Asn	Leu		
		1265				1270				1275					1280		
ATC	AAA	CAA	TAC	GCT	TCA	TTA	GGT	GAT	AAA	TTT	ATA	ATT	TAT	GAT	TCA	1888	
Met	Lys	Gln	Tyr	Ala	Ser	Leu	Gly	Asp	Lys	Phe	Ile	Ile	Tyr	Asp	Ser		
			1285						1290					1295			
TCA	TTT	GAT	GAT	GCA	AAC	CCT	TTT	AAT	CTG	GTG	CCA	TTG	TTT	AAA	TTC	1936	
Ser	Phe	Asp	Asp	Ala	Asn	Arg	Phe	Asn	Leu	Val	Pro	Leu	Phe	Lys	Phe		
			1300					1305					1310				
GGA	AAA	GAC	GAG	AAC	TCA	GAT	GAT	AGT	ATT	TGT	ATA	TAT	AAT	GAA	AAC	1984	
Gly	Lys	Asp	Glu	Asn	Ser	Asp	Asp	Ser	Ile	Cys	Ile	Tyr	Asn	Glu	Asn		
		1315					1320					1325					
CCT	TCC	TCT	GAA	GAT	AAG	AAG	TGG	TAT	TTT	TCT	TCC	AAA	GAT	GAC	AAT	4032	
Pro	Ser	Ser	Glu	Asp	Lys	Lys	Trp	Tyr	Phe	Ser	Ser	Lys	Asp	Asp	Asn		
		1330				1335					1340						
AAA	ACA	CCG	GAT	TAT	AAT	GGT	GGA	ACT	CAA	TGT	ATA	GAT	GCT	GGA	ACC	4080	
Lys	Thr	Ala	Asp	Tyr	Asn	Gly	Gly	Thr	Gln	Cys	Ile	Asp	Ala	Gly	Thr		
		1345			1350					1355					1360		
AGT	AAC	AAA	GAT	TTT	TAT	TAT	AAT	CTC	CAG	GAG	ATT	GAA	GTA	ATT	AGT	4128	
Ser	Asn	Lys	Asp	Phe	Tyr	Tyr	Asn	Leu	Gln	Glu	Ile	Glu	Val	Ile	Ser		
			1365						1370					1375			
GTT	ACT	GGT	GGG	TAT	TGG	TGG	AGT	TAT	AAA	ATA	TCC	AAC	CCG	ATT	AAT	4176	
Val	Thr	Gly	Gly	Tyr	Trp	Ser	Ser	Tyr	Lys	Ile	Ser	Asn	Pro	Ile	Asn		
			1380					1385					1390				
ATC	AAT	ACG	GGC	ATT	GAT	AGT	GCT	AAA	GTA	AAA	GTC	ACC	GTA	AAA	GCG	4224	
Ile	Asn	Thr	Gly	Ile	Asp	Ser	Ala	Lys	Val	Lys	Val	Thr	Val	Lys	Ala		
		1395					1400					1405					
GGT	GGT	GAC	GAT	CAA	ATC	TTT	ACT	GCT	GAT	AAT	AGT	ACC	TAT	GTT	CCT	4272	
Gly	Gly	Asp	Asp	Gln	Ile	Phe	Thr	Ala	Asp	Asn	Ser	Thr	Tyr	Val	Pro		
		1410				1415					1420						
CAG	CAA	CCG	GCA	CCC	AGT	TTT	GAG	GAG	ATG	ATT	TAT	CAG	TTC	AAT	AAC	4320	
Gln	Gln	Pro	Ala	Pro	Ser	Phe	Glu	Glu	Met	Ile	Tyr	Gln	Phe	Asn	Asn		
		1425			1430					1435					1440		

U A 8 2 4 8 5 C 2

ITG ACA ATA GAT TGT AAG AAT TTA AAT TTC ATC GAC AAT CAG GCA CAT 4563
 Leu Thr Ile Asp Cys Lys Asn Leu Asn Phe Ile Asp Asn Gln Ala His
 1445 1450 1455
 ATT GAG ATT GAT TTC ACC GCT ACG GCA CAA GAT GGC CGA TTC TTG GGT 4416
 Ile Glu Ile Asp Phe Thr Ala Thr Ala Gln Asp Gly Arg Phe Leu Gly
 1460 1465 1470
 GCA GAA ACT TTT ATT ATC CCG GTA ACT AAA AAA GTT CTC GGT ACT GAG 4464
 Ala Glu Thr Phe Ile Ile Pro Val Thr Lys Lys Val Leu Gly Thr Glu
 1475 1480 1485
 AAC GTG ATT GCG TTA TAT AGC GAA AAT AAC CGT GTT CAA TAT ATG CAA 4512
 Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly Val Gln Tyr Met Gln
 1490 1495 1500
 ATT GGC GCA TAT CGT ACC CGT TTG AAT ACG TTA TTC GCT CAA CAG TTC 4560
 Ile Gly Ala Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Gln Gln Leu
 1505 1510 1515 1520
 GTT AGC CGT GCT AAT CGT GGC ATT GAT GCA GTG CTC AGT ATG GAA ACT 4608
 Val Ser Arg Ala Asn Arg Gly Ile Asp Ala Val Leu Ser Met Glu Thr
 1525 1530 1535
 CAG AAT ATT CAG GAA CCG CAA TTA GCA GCG GGC ACA TAT GTG CAG CTT 4656
 Gln Asn Ile Gln Glu Pro Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu
 1540 1545 1550
 GTG TTG GAT AAA TAT GAT GAG TCT ATT CAT GGC ACT AAT AAA AGC TTT 4704
 Val Leu Asp Lys Tyr Asp Glu Ser Ile His Gly Thr Asn Lys Ser Phe
 1555 1560 1565
 GCT ATT GAA TAT GTT GAT ATA TTT AAA GAG AAC GAT AGT TTT GTG ATT 4752
 Ala Ile Glu Tyr Val Asp Ile Phe Lys Glu Asn Asp Ser Phe Val Ile
 1570 1575 1580
 TAT CAA GCA GAA CTT AGC GAA ACA ACT CAA ACT GTT GTG AAA GTT TTC 4800
 Tyr Gln Gly Glu Leu Ser Glu Thr Ser Gln Thr Val Val Lys Val Phe
 1585 1590 1595 1600
 TTA TCC TAT TTT ATA GAG GCG ACT CGA AAT AAG AAC CAC TTA TCG GTA 4848
 Leu Ser Tyr Phe Ile Glu Ala Thr Gly Asn Lys Asn His Leu Trp Val
 1605 1610 1615
 CGT GCT AAA TAC CAA AAG GAA ACG ACT GAT AAG ATC TTC TTC GAC CGT 4896
 Arg Ala Lys Tyr Gln Lys Glu Thr Thr Asp Lys Ile Leu Phe Asp Arg
 1620 1625 1630
 ACT GAT GAG AAA GAT CCG CAC GGT TCG TTT CTC ACC GAC GAT CAC AAG 4944
 Thr Asp Glu Lys Asp Pro His Gly Trp Phe Leu Ser Asp Asp His Lys
 1635 1640 1645
 ACC TTT AGT GGT CTC TCT TCC GCA CAG GCA TTA AAG AAC CAC ACT GAA 4992
 Thr Phe Ser Gly Leu Ser Ser Ala Gln Ala Leu Lys Asn Asp Ser Glu
 1650 1655 1660
 CCG ATG GAT TTC TCT GCG GCC AAT GCT CTC TAT TTC TCG GAA CTC TTC 5040
 Pro Met Asp Phe Ser Gly Ala Asn Ala Leu Tyr Phe Trp Glu Leu Phe
 1665 1670 1675 1680
 TAT TAC ACG CCG ATG ATG ATG GCT CAT CCT TTG TTG CAG GAA CAG AAT 5088
 Tyr Tyr Thr Pro Met Met Met Ala His Arg Leu Leu Gln Glu Gln Asn
 1685 1690 1695
 TTT GAT GCG GCG AAC CAT TCG TTC CGT TAT GTC TCG AGT CCA TCC GGT 5136
 Phe Asp Ala Ala Asn His Trp Phe Arg Tyr Val Trp Ser Pro Ser Gly
 1700 1705 1710
 TAT ATC GTT GAT GGT AAA ATT GCT ATC TAC CAC TCG AAC GTG CGA CCG 5184
 Tyr Ile Val Asp Gly Lys Ile Ala Ile Tyr His Trp Asn Val Arg Pro
 1715 1720 1725

CTG GAA GAA GAC ACC AGT TGG AAT GCA CAA CAA CTG GAC TCC ACC GAT 5230
 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp
 1730 1735 1740
 CCA GAT GCT GTA GCC CAA GAT GAT CCG ATG CAC TAC AAG GTG GCT ACC 5280
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr
 1745 1750 1755 1760
 TTT ATG CCG ACC TTG GAT CTG CTA ATG GCC CGT GGT GAT GCT GCT TAC 5328
 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr
 1765 1770 1775
 CCC CAG TTA GAG CCG GAT ACC TTG GCT GAA GCT AAA ATG TGG TAT ACA 5376
 Arg Gln Leu Glu Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr
 1780 1785 1790
 CAG GCG CTT AAT CTG TTG GGT GAT GAG CCA CAA GTG ATG CTG AGT ACC 5424
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr
 1795 1800 1805
 ACT TGG GCT AAT CCA ACA TTG GGT AAT GCT GCT TCA AAA ACC ACA CAG 5472
 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln
 1810 1815 1820
 CAG GTT CGT CAG CAA CTG CTT ACC CAG TTG CCG CTC AAT AGC AGG GTA 5520
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val
 1825 1830 1835 1840
 AAA ACC CCG TTG 5532
 Lys Thr Pro Leu
 1844

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 53:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1844 амінокислот
- (B) ТИП: амінокислоти
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 53 (TсbA₁₁):

Ознаки	Від	До	Опис
Пептид	1	1844	TсbA ₁₁ Пептид
Fragment	1	11	(SEQ ID NO:1)
Fragment	978	990	(SEQ ID NO:23)
Fragment	1387	1401	(SEQ ID NO:22)
Fragment	1484	1505	(SEQ ID NO:24)
Фрагмент	1527	1552	(SEQ ID NO:21)

Phe Ile Gln Gly Tyr Ser Asp Leu Phe Gly Asn Arg Ala Asp Asn Tyr
 1 5 10 15
 Ala Ala Pro Gly Ser Val Ala Ser Met Phe Ser Pro Ala Ala Tyr Leu
 20 25 30
 Thr Glu Leu Tyr Arg Glu Ala Lys Asn Leu His Asp Ser Ser Ser Ile
 35 40 45
 Tyr Tyr Leu Asp Lys Arg Arg Pro Asp Leu Ala Ser Leu Met Leu Ser
 50 55 60
 Gln Lys Asn Met Asp Glu Glu Ile Ser Thr Leu Ala Leu Ser Asn Glu
 65 70 75 80
 Leu Cys Leu Ala Gly Ile Glu Thr Lys Thr Gly Lys Ser Gln Asp Glu

U A 8 2 4 8 5 C 2

85										90										95										
Val	Met	Asp	Met	Leu	Ser	Thr	Tyr	Arg	Leu	Ser	Gly	Glu	Thr	Pro	Tyr															
			100					105						110																
His	His	Ala	Tyr	Glu	Thr	Val	Arg	Glu	Ile	Val	His	Glu	Arg	Asp	Pro															
		115					120					125																		
Gly	Phe	Arg	His	Leu	Ser	Gln	Ala	Pro	Ile	Val	Ala	Ala	Lys	Leu	Asp															
	130					135					140																			
Pro	Val	Thr	Leu	Leu	Gly	Ile	Ser	Ser	His	Ile	Ser	Pro	Glu	Leu	Tyr															
	145				150					155					160															
Asn	Leu	Leu	Ile	Glu	Glu	Ile	Pro	Glu	Lys	Asp	Glu	Ala	Ala	Leu	Asp															
			165					170						175																
Thr	Leu	Tyr	Lys	Thr	Asn	Phe	Gly	Asp	Ile	Thr	Thr	Ala	Gln	Leu	Met															
			180					185					190																	
Ser	Pro	Ser	Tyr	Leu	Ala	Arg	Tyr	Tyr	Gly	Val	Ser	Pro	Glu	Asp	Ile															
		195					200					205																		
Ala	Tyr	Val	Thr	Thr	Ser	Leu	Ser	His	Val	Gly	Tyr	Ser	Ser	Asp	Ile															
	210					215					220																			
Leu	Val	Ile	Pro	Leu	Val	Asp	Gly	Val	Gly	Lys	Met	Glu	Val	Val	Arg															
	225				230					235					240															
Val	Thr	Arg	Thr	Pro	Ser	Asp	Asn	Tyr	Thr	Ser	Gln	Thr	Asn	Tyr	Ile															
			245					250						255																
Glu	Leu	Tyr	Pro	Gln	Gly	Gly	Asp	Asn	Tyr	Leu	Ile	Lys	Tyr	Asn	Leu															
			260					265					270																	
Ser	Asn	Ser	Phe	Gly	Leu	Asp	Asp	Phe	Tyr	Leu	Gln	Tyr	Lys	Asp	Gly															
		275					280					285																		
Ser	Ala	Asp	Trp	Thr	Glu	Ile	Ala	His	Asn	Pro	Tyr	Pro	Asp	Met	Val															
	290					295				300																				
Ile	Asn	Gln	Lys	Tyr	Glu	Ser	Gln	Ala	Thr	Ile	Lys	Arg	Ser	Asp	Ser															
	305				310					315					320															
Asp	Asn	Ile	Leu	Ser	Ile	Gly	Leu	Gln	Arg	Trp	His	Ser	Gly	Ser	Tyr															
			325					330						335																
Asn	Phe	Ala	Ala	Ala	Asn	Phe	Lys	Ile	Asp	Gln	Tyr	Ser	Pro	Lys	Ala															
		340						345					350																	
Phe	Leu	Leu	Lys	Met	Asn	Lys	Ala	Ile	Arg	Leu	Leu	Lys	Ala	Thr	Gly															
		355					360					365																		
Leu	Ser	Phe	Ala	Thr	Leu	Glu	Arg	Ile	Val	Asp	Ser	Val	Asn	Ser	Thr															
	370					375					380																			
Lys	Ser	Ile	Thr	Val	Glu	Val	Leu	Asn	Lys	Val	Tyr	Arg	Val	Lys	Phe															
	385				390					395					400															
Tyr	Ile	Asp	Arg	Tyr	Gly	Ile	Ser	Glu	Glu	Thr	Ala	Ala	Ile	Leu	Ala															
			405					410					415																	
Asn	Ile	Asn	Ile	Ser	Gln	Gln	Ala	Val	Gly	Asn	Gln	Leu	Ser	Gln	Phe															
		420						425					430																	
Glu	Gln	Leu	Phe	Asn	His	Pro	Pro	Leu	Asn	Gly	Ile	Arg	Tyr	Glu	Ile															
		435				440						445																		
Ser	Glu	Asp	Asn	Ser	Lys	His	Leu	Pro	Asn	Pro	Asp	Leu	Asn	Leu	Lys															
	450					455					460																			

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Pro Asp Ser Thr Gly Asp Asp Gln Arg Lys Ala Val Leu Lys Arg Ala
465 470 475 480

Phe Gln Val Asn Ala Ser Glu Leu Tyr Gln Met Leu Leu Ile Thr Asp
485 490 495

Arg Lys Glu Asp Gly Val Ile Lys Asn Asn Leu Glu Asn Leu Ser Asp
500 505 510

Leu Tyr Leu Val Ser Leu Leu Ala Gln Ile His Asn Leu Thr Ile Ala
515 520 525

Glu Leu Asn Ile Leu Leu Val Ile Cys Gly Tyr Gly Asp Thr Asn Ile
530 535 540

Tyr Gln Ile Thr Asp Asp Asn Leu Ala Lys Ile Val Glu Thr Leu Leu
545 550 555 560

Trp Ile Thr Gln Trp Leu Lys Thr Gln Lys Trp Thr Val Thr Asp Leu
565 570 575

Phe Leu Met Thr Thr Ala Thr Tyr Ser Thr Thr Leu Thr Pro Glu Ile
580 585 590

Ser Asn Leu Thr Ala Thr Leu Ser Ser Thr Leu His Gly Lys Glu Ser
595 600 605

Leu Ile Gly Glu Asp Leu Lys Arg Ala Met Ala Pro Cys Phe Thr Ser
610 615 620

Ala Leu His Leu Thr Ser Gln Glu Val Ala Tyr Asp Leu Leu Leu Trp
625 630 635 640

Ile Asp Gln Ile Gln Pro Ala Gln Ile Thr Val Asp Gly Phe Trp Glu
645 650 655

Glu Val Gln Thr Thr Pro Thr Ser Leu Lys Val Ile Thr Phe Ala Gln
660 665 670

Val Leu Ala Gln Leu Ser Leu Ile Tyr Arg Arg Ile Gly Leu Ser Glu
675 680 685

Thr Glu Leu Ser Leu Ile Val Thr Gln Ser Ser Leu Leu Val Ala Gly
690 695 700

Lys Ser Ile Leu Asp His Gly Leu Leu Thr Leu Met Ala Leu Glu Gly
705 710 715 720

Phe His Thr Trp Val Asn Gly Leu Gly Gln His Ala Ser Leu Ile Leu
725 730 735

Ala Ala Leu Lys Asp Gly Ala Leu Thr Val Thr Asp Val Ala Gln Ala
740 745 750

Met Asn Lys Glu Glu Ser Leu Leu Gln Met Ala Ala Asn Gln Val Glu
755 760 765

Lys Asp Leu Thr Lys Leu Thr Ser Trp Thr Gln Ile Asp Ala Ile Leu
770 775 780

Gln Trp Leu Gln Met Ser Ser Ala Leu Ala Val Ser Pro Leu Asp Leu
785 790 795 800

Ala Gly Met Met Ala Leu Lys Tyr Gly Ile Asp His Asn Tyr Ala Ala
805 810 815

Trp Gln Ala Ala Ala Ala Ala Leu Met Ala Asp His Ala Asn Gln Ala
820 825 830

Gln Lys Lys Leu Asp Glu Thr Phe Ser Lys Ala Leu Cys Asn Tyr Tyr
835 840 845

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

Ile Asn Ala Val Val Asp Ser Ala Ala Gly Val Arg Asp Arg Asn Gly
350 855 860

Leu Tyr Thr Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Asp Val Ile
865 870 875 880

Thr Ser Arg Ile Ala Glu Ala Ile Ala Gly Ile Gln Leu Tyr Val Asn
885 890 895

Arg Ala Leu Asn Arg Asp Glu Gly Gln Leu Ala Ser Asp Val Ser Thr
900 905 910

Arg Gln Phe Phe Thr Asp Trp Glu Arg Tyr Asn Lys Arg Tyr Ser Thr
915 920 925

Trp Ala Gly Val Ser Glu Leu Val Tyr Tyr Pro Glu Asn Tyr Val Asp
930 935 940

Pro Thr Gln Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu Gln
945 950 955 960

Ser Ile Asn Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala Phe
965 970 975

Lys Thr Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val Ile
980 985 990

Ser Ala Tyr His Asp Asn Val Asn Val Asp Gln Gly Leu Thr Tyr Phe
995 1000 1005

Ile Gly Ile Asp Gln Ala Ala Pro Gly Thr Tyr Tyr Trp Arg Ser Val
1010 1015 1020

Asp His Ser Lys Cys Glu Asn Gly Lys Phe Ala Ala Asn Ala Trp Gly
1025 1030 1035 1040

Glu Trp Asn Lys Ile Thr Cys Ala Val Asn Pro Trp Lys Asn Ile Ile
1045 1050 1055

Arg Pro Val Val Tyr Met Ser Arg Leu Tyr Leu Leu Trp Leu Glu Gln
1060 1065 1070

Gln Ser Lys Lys Ser Asp Asp Gly Lys Thr Thr Ile Tyr Gln Tyr Asn
1075 1080 1085

Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Ser Trp Asn Thr Pro Phe
1090 1095 1100

Thr Phe Asp Val Thr Glu Lys Val Lys Asn Tyr Thr Ser Ser Thr Asp
1105 1110 1115 1120

Ala Ala Glu Ser Leu Gly Leu Tyr Cys Thr Gly Tyr Gln Gly Glu Asp
1125 1130 1135

Thr Leu Leu Val Met Phe Tyr Ser Met Gln Ser Ser Tyr Ser Ser Tyr
1140 1145 1150

Thr Asp Asn Asn Ala Pro Val Thr Gly Leu Tyr Ile Phe Ala Asp Met
1155 1160 1165

Ser Ser Asp Asn Met Thr Asn Ala Gln Ala Thr Asn Tyr Trp Asn Asn
1170 1175 1180

Ser Tyr Pro Gln Phe Asp Thr Val Met Ala Asp Pro Asp Ser Asp Asn
1185 1190 1195 1200

Lys Lys Val Ile Thr Arg Arg Val Asn Asn Arg Tyr Ala Glu Asp Tyr
1205 1210 1215

Glu Ile Pro Ser Ser Val Thr Ser Asn Ser Asn Tyr Ser Trp Gly Asp

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

1220	1225	1230
His Ser Leu Thr Met Leu Tyr Gly Gly Ser Val Pro Asn Ile Thr Phe 1235 1240 1245		
Glu Ser Ala Ala Glu Asp Leu Arg Leu Ser Thr Asn Met Ala Leu Ser 1250 1255 1260		
Ile Ile His Asn Gly Tyr Ala Gly Thr Arg Arg Ile Gln Cys Asn Leu 1265 1270 1275 1280		
Met Lys Gln Tyr Ala Ser Leu Gly Asp Lys Phe Ile Ile Tyr Asp Ser 1285 1290 1295		
Ser Phe Asp Asp Ala Asn Arg Phe Asn Leu Val Pro Leu Phe Lys Phe 1300 1305 1310		
Gly Lys Asp Glu Asn Ser Asp Asp Ser Ile Cys Ile Tyr Asn Glu Asn 1315 1320 1325		
Pro Ser Ser Glu Asp Lys Lys Trp Tyr Phe Ser Ser Lys Asp Asp Asn 1330 1335 1340		
Lys Thr Ala Asp Tyr Asn Gly Gly Thr Gln Cys Ile Asp Ala Gly Thr 1345 1350 1355 1360		
Ser Asn Lys Asp Phe Tyr Tyr Asn Leu Gln Glu Ile Glu Val Ile Ser 1365 1370 1375		
Val Thr Gly Gly Tyr Trp Ser Ser Tyr Lys Ile Ser Asn Pro Ile Asn 1380 1385 1390		
Ile Asn Thr Gly Ile Asp Ser Ala Lys Val Lys Val Thr Val Lys Ala 1395 1400 1405		
Gly Gly Asp Asp Gln Ile Phe Thr Ala Asp Asn Ser Thr Tyr Val Pro 1410 1415 1420		
Gln Gln Pro Ala Pro Ser Phe Glu Glu Met Ile Tyr Gln Phe Asn Asn 1425 1430 1435 1440		
Leu Thr Ile Asp Cys Lys Asn Leu Asn Phe Ile Asp Asn Gln Ala His 1445 1450 1455		
Ile Glu Ile Asp Phe Thr Ala Thr Ala Gln Asp Gly Arg Phe Leu Gly 1460 1465 1470		
Ala Glu Thr Phe Ile Ile Pro Val Thr Lys Lys Val Leu Gly Thr Glu 1475 1480 1485		
Asn Val Ile Ala Leu Tyr Ser Glu Asn Asn Gly Val Gln Tyr Met Gln 1490 1495 1500		
Ile Gly Ala Tyr Arg Thr Arg Leu Asn Thr Leu Phe Ala Gln Gln Leu 1505 1510 1515 1520		
Val Ser Arg Ala Asn Arg Gly Ile Asp Ala Val Leu Ser Met Glu Thr 1525 1530 1535		
Gln Asn Ile Gln Glu Pro Gln Leu Gly Ala Gly Thr Tyr Val Gln Leu 1540 1545 1550		
Val Leu Asp Lys Tyr Asp Glu Ser Ile His Gly Thr Asn Lys Ser Phe 1555 1560 1565		
Ala Ile Glu Tyr Val Asp Ile Phe Lys Glu Asn Asp Ser Phe Val Ile 1570 1575 1580		
Tyr Gln Gly Glu Leu Ser Glu Thr Ser Gln Thr Val Val Lys Val Phe 1585 1590 1595 1600		

U A 8 2 4 8 5 C 2

Leu Ser Tyr Phe Ile Glu Ala Thr Gly Asn Lys Asn His Leu Trp Val
 1605 1610 1615
 Arg Ala Lys Tyr Gln Lys Glu Thr Thr Asp Lys Ile Leu Phe Asp Arg
 1620 1625 1630
 Thr Asp Glu Lys Asp Pro His Gly Trp Phe Leu Ser Asp Asp His Lys
 1635 1640 1645
 Thr Phe Ser Gly Leu Ser Ser Ala Gln Ala Leu Lys Asn Asp Ser Glu
 1650 1655 1660
 Pro Met Asp Phe Ser Gly Ala Asn Ala Leu Tyr Phe Trp Glu Leu Phe
 1665 1670 1675 1680
 Tyr Tyr Thr Pro Met Met Met Ala His Arg Leu Leu Gln Glu Gln Asn
 1685 1690 1695
 Phe Asp Ala Ala Asn His Trp Phe Arg Tyr Val Trp Ser Pro Ser Gly
 1700 1705 1710
 Tyr Ile Val Asp Gly Lys Ile Ala Ile Tyr His Trp Asn Val Arg Pro
 1715 1720 1725
 Leu Glu Glu Asp Thr Ser Trp Asn Ala Gln Gln Leu Asp Ser Thr Asp
 1730 1735 1740
 Pro Asp Ala Val Ala Gln Asp Asp Pro Met His Tyr Lys Val Ala Thr
 1745 1750 1755 1760
 Phe Met Ala Thr Leu Asp Leu Leu Met Ala Arg Gly Asp Ala Ala Tyr
 1765 1770 1775
 Arg Gln Leu Glu Arg Asp Thr Leu Ala Glu Ala Lys Met Trp Tyr Thr
 1780 1785 1790
 Gln Ala Leu Asn Leu Leu Gly Asp Glu Pro Gln Val Met Leu Ser Thr
 1795 1800 1805
 Thr Trp Ala Asn Pro Thr Leu Gly Asn Ala Ala Ser Lys Thr Thr Gln
 1810 1815 1820
 Gln Val Arg Gln Gln Val Leu Thr Gln Leu Arg Leu Asn Ser Arg Val
 1825 1830 1835 1840
 Lys Thr Pro Leu
 1844

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 54:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1722 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 54 (кодуючий район TcbA_{III}):

CTA CGA ACA GCC AAT TCC CTC ACC GGT TTA TTC CTG CCG CAG GAA AAT 43
 Leu Gly Thr Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn
 1 5 10 15
 AGC AAG CTC AAA GGC TAC TCG CCG ACA CTC GCG CAG GGT ATG TTT AAT 96
 Ser Lys Leu Lys Gly Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn
 20 25 30

TTA CGT CAT AAT CTG TCG ATT GAC GCG CAG CCG CTC TCC TTG CCG CTG 144
 Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu
 35 40 45

TAT GCT AAA CCG GCT GAT CCA AAA GCT TTA CTC AGT GCG GCG GTT TCA 152
 Tyr Ala Lys Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser
 50 55 60

GCT TCT CAA GCG GGA GCG GAC TTG CCG AAG GCG CCG CTC ACT ATT CAC 240
 Ala Ser Gln Gly Gly Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His
 65 70 75 80

CGC TTC CCT CAA ATG CTA GAA GCG GCA CCG GCG TTG GTT AAC CAG CTT 288
 Arg Phe Pro Gln Met Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu
 85 90 95

ATA CAG TTC GGT AGT TCA CTA TTG GCG TAC AGT GAG CGT CAG GAT GCG 336
 Ile Gln Phe Gly Ser Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala
 100 105 110

GAA GCT ATG AGT CAA CTA CTG CAA ACC CAA GCG AGC GAG TTA ATA CTC 384
 Glu Ala Met Ser Gln Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu
 115 120 125

ACC AGT ATT CGT ATG CAG GAT AAC CAA TTG GCA GAG CTC GAT TCG GAA 432
 Thr Ser Ile Arg Met Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu
 130 135 140

AAA ACC GCG TTG CAA CTC TCT TTA GCT GGA GTG CAA CAA CCG TTT GAC 480
 Lys Thr Ala Leu Gln Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp
 145 150 155 160

AGC TAT AGC CAA CTG TAT GAG GAG AAC ATC AAC GCA GGT GAG CAG CGA 528
 Ser Tyr Ser Gln Leu Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg
 165 170 175

GCG CTG GCG TTA CCG TCA CAA TCT GCT ATT GAG TCT CAG GGA GCG CAG 576
 Ala Leu Ala Leu Arg Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln
 180 185 190

ATT TCC CGT ATC GCA GCG GCG GGT GTT GAT ATC CCA CCA AAT ATC TTC 624
 Ile Ser Arg Met Ala Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe
 195 200 205

GCG CTG GCT GAT GCG GCG ATG CAT TAT GGT GCT ATT GCC TAT GCC ATC 572
 Gly Leu Ala Asp Gly Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile
 210 215 220

GCT GAC GGT ATT GAG TTG AGT GCT TCT GCC AAG ATG GTT GAT GCG GAG 720
 Ala Asp Gly Ile Glu Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu
 225 230 235 240

AAA GTT GCT CAG TCG GAA ATA TAT CCG CGT CCG CGT CAA GAA TGG AAA 768
 Lys Val Ala Gln Ser Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys
 245 250 255

ATT CAG CGT GAC AAC GCA CAA GCG GAG ATT AAC CAG TTA AAC GCG CAA 816
 Ile Gln Arg Asp Asn Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln
 260 265 270

CTG GAA TCA CTG TCT ATT CCG CGT GAA GCG GCT CAA ATC CAA AAA GAG 864
 Leu Glu Ser Leu Ser Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu
 275 280 285

TAC CTG AAA ACC CAG CAA GCT CAG GCG CAG GCA CAA CTT ACT TTC TTA 912
 Tyr Leu Lys Thr Gln Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu
 290 295 300

AGA AGC AAA TTC AGT AAT CAA GCG TTA TAT AGT TCG TTA CCA GCG CGT 960
 Arg Ser Lys Phe Ser Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg
 305 310 315 320

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

TTG TCA GGT ATT TAT TTC CAG TTC TAT GAC TCG GCC GTA TCA CGT TCC	1008
Leu Ser Gly Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys	
325 330 335	
CTG ATG GCA GAG CAA TCC TAT CAA TCG GAA GCT AAT GAT AAT TCC ATT	1056
Leu Met Ala Glu Gln Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile	
340 345 350	
AGC TTT GTC AAA CCG GGT GCA TGG CAA GCA ACT TAC GCC GGC TTA TTG	1104
Ser Phe Val Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu	
355 360 365	
TGT GCA GAA GCT TTG ATA CAA AAT CTG GCA CAA ATG GAA GAG GCA TAT	1152
Cys Gly Glu Ala Leu Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr	
370 375 380	
CTG AAA TGG GAA TCT CCG GCT TTC GAA GTA GAA CCG ACC GTT TCA TTG	1200
Leu Lys Trp Glu Ser Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu	
385 390 395 400	
GCA CTC GTT TAT GAT TCA CTG GAA GGT AAT GAT CGT TTT AAT TTA GCG	1248
Ala Val Val Tyr Asp Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala	
405 410 415	
GAA CAA ATA CCT GCA TTA TTG GAT AAG GGG GAG GCA ACA GCA GGA ACT	1296
Glu Gln Ile Pro Ala Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr	
420 425 430	
AAA GAA AAT GCG TTA TCA TTG GCT AAT GCT ATC CTG TCA GCT TCG GTC	1344
Lys Glu Asn Gly Leu Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val	
435 440 445	
AAA TTG TCC GAC TTG AAA CTG GCA ACC GAT TAT CCA GAC AGT ATC GTT	1392
Lys Leu Ser Asp Leu Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val	
450 455 460	
GGT AGC AAC AAG GTT CGT CGT ATT AAG CAA ATC AGT GTT TCG CTA CCT	1440
Gly Ser Asn Lys Val Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro	
465 470 475 480	
GCA TTG GTT GCG CCT TAT CAG GAT GTT CAG GCT ATG CTC AGC TAT GGT	1488
Ala Leu Val Gly Pro Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly	
485 490 495	
GGC AGT ACT CAA TTG CCG AAA GGT TGT TCA GCG TTG GCT GTG TCT CAT	1536
Gly Ser Thr Gln Leu Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His	
500 505 510	
GGT ACC AAT GAT AGT GGT CAG TTC CAG TTG GAT TTC AAT GAC GGC AAA	1584
Gly Thr Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys	
515 520 525	
TAC CTG CCA TTT GAA GCT ATT GCT CTT GAT GAT CAG GGT ACA CTG AAT	1632
Tyr Leu Pro Phe Glu Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn	
530 535 540	
CTT CAA TTT CCG AAT GCT ACC GAC AAG CAG AAA GCA ATA TTG CAA ACT	1680
Leu Gln Phe Pro Asn Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr	
545 550 555 560	
ATG AGC GAT ATT ATT TTG CAT ATT CGT TAT ACC ATC CGT TAA	1722
Met Ser Asp Ile Ile Leu His Ile Arg Tyr Thr Ile Arg ...	
565 570 573	

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 55:

- (1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:
(А) ДОВЖИНА: 573 амінокислоти

(B) ТИП: амінокислота
(C) КІЛЬКІСТЬ ЛАНЦЮГІВ: один
(D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 55 (TcbA₁₁₁):

```

Leu Gly Thr Ala Asn Ser Leu Thr Ala Leu Phe Leu Pro Gln Glu Asn
 1           5           10           15
Ser Lys Leu Lys Gly Tyr Trp Arg Thr Leu Ala Gln Arg Met Phe Asn
 20           25           30
Leu Arg His Asn Leu Ser Ile Asp Gly Gln Pro Leu Ser Leu Pro Leu
 35           40           45
Tyr Ala Lys Pro Ala Asp Pro Lys Ala Leu Leu Ser Ala Ala Val Ser
 50           55           60
Ala Ser Gln Gly Gly Ala Asp Leu Pro Lys Ala Pro Leu Thr Ile His
 65           70           75           80
Arg Phe Pro Gln Met Leu Glu Gly Ala Arg Gly Leu Val Asn Gln Leu
 85           90           95
Ile Gln Phe Gly Ser Ser Leu Leu Gly Tyr Ser Glu Arg Gln Asp Ala
 100          105          110
Glu Ala Met Ser Gln Leu Leu Gln Thr Gln Ala Ser Glu Leu Ile Leu
 115          120          125
Thr Ser Ile Arg Met Gln Asp Asn Gln Leu Ala Glu Leu Asp Ser Glu
 130          135          140
Lys Thr Ala Leu Gln Val Ser Leu Ala Gly Val Gln Gln Arg Phe Asp
 145          150          155          160
Ser Tyr Ser Gln Leu Tyr Glu Glu Asn Ile Asn Ala Gly Glu Gln Arg
 165          170          175
Ala Leu Ala Leu Arg Ser Glu Ser Ala Ile Glu Ser Gln Gly Ala Gln
 180          185          190
Ile Ser Arg Met Ala Gly Ala Gly Val Asp Met Ala Pro Asn Ile Phe
 195          200          205
Gly Leu Ala Asp Gly Gly Met His Tyr Gly Ala Ile Ala Tyr Ala Ile
 210          215          220
Ala Asp Gly Ile Glu Leu Ser Ala Ser Ala Lys Met Val Asp Ala Glu
 225          230          235          240
Lys Val Ala Gln Ser Glu Ile Tyr Arg Arg Arg Arg Gln Glu Trp Lys
 245          250          255
Ile Gln Arg Asp Asn Ala Gln Ala Glu Ile Asn Gln Leu Asn Ala Gln
 260          265          270
Leu Glu Ser Leu Ser Ile Arg Arg Glu Ala Ala Glu Met Gln Lys Glu
 275          280          285
Tyr Leu Lys Thr Gln Gln Ala Gln Ala Gln Ala Gln Leu Thr Phe Leu
 290          295          300
Arg Ser Lys Phe Ser Asn Gln Ala Leu Tyr Ser Trp Leu Arg Gly Arg
 305          310          315          320
Leu Ser Gly Ile Tyr Phe Gln Phe Tyr Asp Leu Ala Val Ser Arg Cys
 325          330          335

```

Leu Met Ala Glu Gln Ser Tyr Gln Trp Glu Ala Asn Asp Asn Ser Ile
 340 345 350
 Ser Phe Val Lys Pro Gly Ala Trp Gln Gly Thr Tyr Ala Gly Leu Leu
 355 360 365
 Cys Gly Glu Ala Leu Ile Gln Asn Leu Ala Gln Met Glu Glu Ala Tyr
 370 375 380
 Leu Lys Trp Glu Ser Arg Ala Leu Glu Val Glu Arg Thr Val Ser Leu
 385 390 395 400
 Ala Val Val Tyr Asp Ser Leu Glu Gly Asn Asp Arg Phe Asn Leu Ala
 405 410 415
 Glu Gln Ile Pro Ala Leu Leu Asp Lys Gly Glu Gly Thr Ala Gly Thr
 420 425 430
 Lys Glu Asn Gly Leu Ser Leu Ala Asn Ala Ile Leu Ser Ala Ser Val
 435 440 445
 Lys Leu Ser Asp Leu Lys Leu Gly Thr Asp Tyr Pro Asp Ser Ile Val
 450 455 460
 Gly Ser Asn Lys Val Arg Arg Ile Lys Gln Ile Ser Val Ser Leu Pro
 465 470 475 480
 Ala Leu Val Gly Pro Tyr Gln Asp Val Gln Ala Met Leu Ser Tyr Gly
 485 490 495
 Gly Ser Thr Gln Leu Pro Lys Gly Cys Ser Ala Leu Ala Val Ser His
 500 505 510
 Gly Thr Asn Asp Ser Gly Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys
 515 520 525
 Tyr Leu Pro Phe Glu Gly Ile Ala Leu Asp Asp Gln Gly Thr Leu Asn
 530 535 540
 Leu Gln Phe Pro Asn Ala Thr Asp Lys Gln Lys Ala Ile Leu Gln Thr
 545 550 555 560
 Met Ser Asp Ile Ile Leu His Ile Arg Tyr Thr Ile Arg ...
 565 570 573

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 56:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 2898 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 56 (tcccA):

1	ATG AAT CAA CTC CCC AGT CCC CTC ATT TCC CGC ACC GAA GAG ATC CAC	48
1	Met Asn Gln Leu Ala Ser Pro Leu Ile Ser Arg Thr Glu Glu Ile His	16
49	AAC TTA CCC GGT AAA TTG ACC GAT CTT GGT TAT ACC TCA GTC TTT GAT	96
17	Asn Leu Pro Gly Lys Leu Thr Asp Leu Gly Tyr Thr Ser Val Phe Asp	32
97	GTC GTA CGT ATG CCC CGT GAG CGT TTT ATT CGT GAG CAT CGT GCT GAT	144

133	Val Val Arg Met Pro Arg Glu Arg Phe Ile Arg Glu His Arg Ala Asp	43
145	CTC GGG CGC AGT GCT GAA AAA ATC TAT CAC CTG GCA GTC GGC TAT GCT	192
49	Leu Gly Arg Ser Ala Glu Lys Met Tyr Asp Leu Ala Val Gly Tyr Ala	64
193	CAT CAG CTG TTA CAC CAT TTT CCC CGT AAT TCT CTT AGT GAA GCT GTT	240
65	His Gln Val Leu His His Phe Arg Arg Asn Ser Leu Ser Glu Ala Val	30
241	CAG TTT GGC TTG AGA AGT CCC TTC TCC GTA TCA GGC CCG GAT TAC GCC	288
81	Gln Phe Gly Leu Arg Ser Pro Phe Ser Val Ser Gly Pro Asp Tyr Ala	96
289	AAT CAG TTT CTT GAT GCA AAC ACG GGT TCG AAA GAT AAA GCA CCA AGT	336
97	Asn Gln Phe Leu Asp Ala Asn Thr Gly Trp Lys Asp Lys Ala Pro Ser	112
337	GGA TCA CCG GAA GCC AAT GAT GCG CCG GTA GCC TAT CTG ACT CAT ATT	384
113	Gly Ser Pro Glu Ala Asn Asp Ala Pro Val Ala Tyr Leu Thr His Ile	128
385	TAT CAA TTG GCC CTT GAA CAG GAA AAG AAT GGC GCC ACT ACC ATT ATC	432
129	Tyr Gln Leu Ala Leu Glu Gln Glu Lys Asn Gly Ala Thr Thr Ile Met	144
433	AAT ACG CTG GCG GAG CGT CGC CCC GAT CTG GGT GCT TTG TTA ATT AAT	480
145	Asn Thr Leu Ala Glu Arg Arg Pro Asp Leu Gly Ala Leu Leu Ile Asn	160
481	GAT AAA GCA ATC AAT GAG GTG ATA CCG CAA TTG CAG TTG GTC AAT GAA	528
161	Asp Lys Ala Ile Asn Glu Val Ile Pro Gln Leu Gln Leu Val Asn Glu	176
529	ATT CTG TCC AAA GCT ATT CAG AAG AAA CTG AGT TTG ACT GAT CTG GAA	576
177	Ile Leu Ser Lys Ala Ile Gln Lys Lys Leu Ser Leu Thr Asp Leu Glu	192
577	GCG GTA AAC GCC AGA CTT TCC ACT ACC CGT TAC CCG AAT AAT CTG CCG	624
193	Ala Val Asn Ala Arg Leu Ser Thr Thr Arg Tyr Pro Asn Asn Leu Pro	208
625	TAT CAT TAT GGT CAT CAG CAG ATT CAG ACA GCT CAA TCG GTA TTG GGT	672
209	Tyr His Tyr Gly His Gln Gln Ile Gln Thr Ala Gln Ser Val Leu Gly	224
673	ACT ACG TTG CAA GAT ATC ACT TTG CCA CAG ACG CTG GAT CTG CCG CAA	720
225	Thr Thr Leu Gln Asp Ile Thr Leu Pro Gln Thr Leu Asp Leu Pro Gln	240
721	AAC TTC TGG GCA ACA GCA AAA GGA AAA CTG AGC GAT ACG ACT GCC AGT	768
241	Asn Phe Trp Ala Thr Ala Lys Gly Lys Leu Ser Asp Thr Thr Ala Ser	256
769	GCT TTG ACC CGA CTG CAA ATC ATG GCG AGT CAG TTT TCG CCA GAG CAG	816
257	Ala Leu Thr Arg Leu Gln Ile Met Ala Ser Gln Phe Ser Pro Glu Gln	272
817	CAG AAA ATC ATT ACC GAG ACT GTC GGT CAG GAT TTC TAT CAG CTT AAC	864
273	Gln Lys Ile Ile Thr Glu Thr Val Gly Gln Asp Phe Tyr Gln Leu Asn	288
865	TAT GGT GAC AGT TCG CTT ACT GTG AAT AGT TTC AGC GAC ATC ACC ATA	912
289	Tyr Gly Asp Ser Ser Leu Thr Val Asn Ser Phe Ser Asp Met Thr Ile	304

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

313	ATG ACT GAT CGA ACA ACT TTG ACT GTA CCC CAG GTA GAA GTG ATG TTG	360
363	Met Thr Asp Arg Thr Ser Leu Thr Val Pro Gln Val Glu Leu Met Leu	363
961	TGT TCA ACT GTC CGA GGT TCT ACC GTT GTT AAG TCT GAT AAT GTG AGT	1003
321	Cys Ser Thr Val Gly Gly Ser Thr Val Val Lys Ser Asp Asn Val Ser	336
1009	TCT GGT GAC ACC ACA GCG ACC CGA TTT GCG TAT GCG GCC CGC TTT ATT	1056
337	Ser Gly Asp Thr Thr Ala Thr Pro Phe Ala Tyr Gly Ala Arg Phe Ile	352
1057	CAT GCC GGT AAG CCG GAG GCG ATT ACC CTG AGT CGC AGT GGT GCG GAG	1104
353	His Ala Gly Lys Pro Glu Ala Ile Thr Leu Ser Arg Ser Gly Ala Glu	368
1105	GCG CAT TTT GCT CTG ACC GTT AAC AAT CTC ACA GAT GAC AAG TTG GAC	1152
369	Ala His Phe Ala Leu Thr Val Asn Asn Leu Thr Asp Asp Lys Leu Asp	384
1153	CGT ATT AAC CGC ACA GTG CCG CTG CAA AAA TGG CTG AAT CTC CCT TAT	1200
385	Arg Ile Asn Arg Thr Val Arg Leu Gln Lys Trp Leu Asn Leu Pro Tyr	400
1201	GAG GAT ATT GAC CTG TTA GTG ACT TCT GCT ATG GAT GCG GAA ACA GGA	1248
401	Glu Asp Ile Asp Leu Leu Val Thr Ser Ala Met Asp Ala Glu Thr Gly	416
1249	AAT ACC GCG CTG TCG ATG AAC GAC AAT ACC CTG GGT ATG TTG GGA GTG	1296
417	Asn Thr Ala Leu Ser Met Asn Asp Asn Thr Leu Arg Met Leu Gly Val	432
1297	TTC AAA CAT TAT CAG GCG AAG TAT GGT GTT ACC GGT AAA CAA TTT GCT	1344
433	Phe Lys His Tyr Gln Ala Lys Tyr Gly Val Ser Ala Lys Gln Phe Ala	448
1345	GCG TGG CTG CCG GTA GTG GCC CCG TTT GCC ATT ACA CCG GCA ACC CCG	1392
449	Gly Trp Leu Arg Val Val Ala Pro Phe Ala Ile Thr Pro Ala Thr Pro	464
1393	TTT TTA GAC CAA GTG TTT AAC TCC GTC GCG ACC TTT GAT ACA CCG TTT	1440
465	Phe Leu Asp Gln Val Phe Asn Ser Val Gly Thr Phe Asp Thr Pro Phe	480
1441	GTG ATA GAT AAT CAG GAT TTT GTC TAT ACA TTG ACC ACC GCG GCG GAT	1488
481	Val Ile Asp Asn Gln Asp Phe Val Tyr Thr Leu Thr Thr Gly Gly Asp	496
1489	GCG GCG CGT GTT AAG CAT ATC ACC ACC GCA CTG GCG CTC AAT CAT CGT	1536
497	Gly Ala Arg Val Lys His Ile Ser Thr Ala Leu Gly Leu Asn His Arg	512
1537	CAG TTC CTG TTA TTG GCG GAT AAT ATT GCC CGT CAA CAG GCG AAT GTG	1584
513	Gln Phe Leu Leu Leu Ala Asp Asn Ile Ala Arg Gln Gln Gly Asn Val	529
1585	ACC CAA AGC ACA CTC AAC TGT AAT CTG TTT GTG CTG TCA GGT TTC TAC	1632
529	Thr Gln Ser Thr Leu Asn Cys Asn Leu Phe Val Val Ser Ala Phe Tyr	544
1633	CGT CTG GGT AAT TTG GCG GCG ACA TTG GCG ATA AAT CCA GAG TCT TTC	1680
545	Arg Leu Ala Asn Leu Ala Arg Thr Leu Gly Ile Asn Pro Glu Ser Phe	560
1681	TGT GCG TTG GTT GAT CGA TTA GAT GCA GGT ACA GCG ATC GTC TGG CAG	1728

531	Cys Ala Leu Val Asp Arg Leu Asp Ala Gly Thr Gly Ile Val Trp Gln	575
1729	CAA TTG GCA GCG AAA GCG ACA ATC ACG GTA CCA CAA AAA GAT TCC GCG	1773
577	Gln Leu Ala Gly Lys Pro Thr Ile Thr Val Pro Gln Lys Asp Ser Pro	531
1777	CTG GCG GCG GAT ATT CTG AGT TTG CTG CAA GCG CTA AGT GCG ATT GCT	
1924		
593	Leu Ala Ala Asp Ile Leu Ser Leu Leu Gln Ala Leu Ser Ala Ile Ala	
608		
1825	CAA TGG CAA CAA CAG CAC GAT TTA GAA TTT TCA GCA CTG CTT TTG CTG	1872
609	Gln Trp Gln Gln Gln His Asp Leu Glu Phe Ser Ala Leu Leu Leu Leu	624
1873	TTG AGT GAC AAC CCT ATT TCT ACC TCG CAG GGC ACT GAC GAT CAA TTG	1920
625	Leu Ser Asp Asn Pro Ile Ser Thr Ser Gln Gly Thr Asp Asp Gln Leu	640
1921	AAC TTT ATC CGT CAA GTG TGG CAG AAC CTA GGC AGT ACG TTT CTG GGT	1958
641	Asn Phe Ile Arg Gln Val Trp Gln Asn Leu Gly Ser Thr Phe Val Gly	656
1969	GCA ACA TTG TTG TCC CCG AGT GGG GCA CCA TTA GTC GAT ACC AAC GGC	2016
657	Ala Thr Leu Leu Ser Arg Ser Gly Ala Pro Leu Val Asp Thr Asn Gly	672
2017	CAC GCT ATT GAC TGG TTT GCT CTG CTC TCA GCA GGT AAT AGT CCG CTT	2064
673	His Ala Ile Asp Trp Phe Ala Leu Leu Ser Ala Gly Asn Ser Pro Leu	688
2065	ATC GAT AAG GTT GGT CTG GTG ACT GAT GCT GGC ATA CAA AGT GTT ATA	2112
689	Ile Asp Lys Val Gly Leu Val Thr Asp Ala Gly Ile Gln Ser Val Ile	704
2113	GCA ACG CTG GTC AAT ACA CAA AGC TTA TCT GAT CAA GAT AAG AAG CTG	2160
705	Ala Thr Val Val Asn Thr Gln Ser Leu Ser Asp Glu Asp Lys Lys Leu	720
2161	GCA ATC ACT ACT CTG ACT AAT ACG TTG AAT CAG GTA CAG AAA ACT CAA	2208
721	Ala Ile Thr Thr Leu Thr Asn Thr Leu Asn Gln Val Gln Lys Thr Gln	736
2209	CAG GGC GTG GCC GTC AGT CTG TTG GCG CAG ACT CTG AAC GTC ACT CAG	2256
737	Gln Gly Val Ala Val Ser Leu Leu Ala Gln Thr Leu Asn Val Ser Gln	752
2257	TCA CTG CCT GCG TTA TTG TTG GCG TGG AGT GGA CAA ACA ACC TAC CAG	2304
753	Ser Leu Pro Ala Leu Leu Leu Arg Trp Ser Gly Gln Thr Thr Tyr Gln	769
2305	TGG TTG AGT GCG ACT TGG GCA TTG AAG GAT GCC GTT AAG ACT GCC GCC	2352
769	Trp Leu Ser Ala Thr Trp Ala Leu Lys Asp Ala Val Lys Thr Ala Ala	784
2353	GAT ATT CCC GCT GAC TAT CTG CGT CAA TTA CGT GAA GTG GTA CGC CGC	2400
785	Asp Ile Pro Ala Asp Tyr Leu Arg Gln Leu Arg Glu Val Val Arg Arg	800
2401	TCC TTG TTG ACC CAA CAA TTC ACG CTG AGT CCT GCA ATG GTG CAA ACC	2448
801	Ser Leu Leu Thr Gln Gln Phe Thr Leu Ser Pro Ala Met Val Gln Thr	816

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

2443	TTG CTG GAC TAT CCA GCC TAT TTT GGC GGT TCC GCA GAA ACA GTG ACC	2494
317	Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr	330
2497	GAT ATC AGT TTG TCG ATG CTT TAT ACC CTG AGC TGT TAT ACC GAT TTA	2544
333	Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu	348
2545	TTG CTC CAA ATG GGT CAA GCT GGT GGT ACC GAA GAT GAT GTA CTG GCC	2592
349	Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala	364
2593	TAC TTA CGC ACA GCT AAT GCT ACC ACA CCG TTG AGC CAA TCT GAT GCT	2640
365	Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala	380
2641	GCA CAG ACC TTG GCA ACC CTA TTG GGT TCG GAG GTT AAC GAG TTG CAA	2688
381	Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln	396
2689	GCC GCT TCG TCG GTA TTG GGC GGG ATT CCC AAA ACC ACA CCG CAA CTG	2736
397	Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu	512
2737	GAT CGG CTT CTG CGT TTG CAA CAG GCA CAG AAC CAA ACT GGT CTT GGC	2784
513	Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly	528
2785	GTT ACA CAG CAA CAG CAA GGC TAT CTC CTG AGT CGT GAC ACT GAT TAT	2832
529	Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr	544
2833	ACC CTT TCG CAA ACC ACC GGT CAG GCG CTG GTG GCT GGC GTA TCC CAT	2880
545	Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His	560
2881	CTC AAG GGC AGT AAC TGA	2898
561	Val Lys Gly Ser Asn End	565

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 57:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 965 амінокислоти
- (B) ТИП: амінокислота
- (C) ТОПОЛОГІЯ: лінійна

(1i) ТИП МОЛЕКУЛИ: білок

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 57 (TссА пептид):

Ознаки	Від	До	Опис
	1	10	SEQ ID № 8
1	Met	Asn	Gln Leu Ala Ser Pro Leu Ile Ser Arg Thr Glu Glu Ile His 16
17	Asn	Leu	Pro Gly Lys Leu Thr Asp Leu Gly Tyr Thr Ser Val Phe Asp 32
33	Val	Val	Arg Met Pro Arg Glu Arg Phe Ile Arg Glu His Arg Ala Asp 48
49	Leu	Gly	Arg Ser Ala Glu Lys Met Tyr Asp Leu Ala Val Gly Tyr Ala 64
65	His	Gln	Val Leu His His Phe Arg Arg Asn Ser Leu Ser Glu Ala Val 80
81	Gln	Phe	Gly Leu Arg Ser Pro Phe Ser Val Ser Gly Pro Asp Tyr Ala 96

U A 8 2 4 8 5 C 2

97	Asn	Gln	Phe	Leu	Asp	Ala	Asn	Thr	Gly	Trp	Lys	Asp	Lys	Ala	Pro	Ser	111
111	Gly	Ser	Pro	Glu	Ala	Asn	Asp	Ala	Pro	Val	Ala	Tyr	Leu	Thr	His	Ile	128
129	Tyr	Gln	Leu	Ala	Leu	Glu	Gln	Glu	Lys	Asn	Gly	Ala	Thr	Thr	Ile	Met	144
145	Asn	Thr	Leu	Ala	Glu	Arg	Arg	Pro	Asp	Leu	Gly	Ala	Leu	Leu	Ile	Asn	160
161	Asp	Lys	Ala	Ile	Asn	Glu	Val	Ile	Pro	Gln	Leu	Gln	Leu	Val	Asn	Glu	176
177	Ile	Leu	Ser	Lys	Ala	Ile	Gln	Lys	Lys	Leu	Ser	Leu	Thr	Asp	Leu	Glu	192
193	Ala	Val	Asn	Ala	Arg	Leu	Ser	Thr	Thr	Arg	Tyr	Pro	Asn	Asn	Leu	Pro	208
209	Tyr	His	Tyr	Gly	His	Gln	Gln	Ile	Gln	Thr	Ala	Gln	Ser	Val	Leu	Gly	224
225	Thr	Thr	Leu	Gln	Asp	Ile	Thr	Leu	Pro	Gln	Thr	Leu	Asp	Leu	Pro	Gln	240
241	Asn	Phe	Trp	Ala	Thr	Ala	Lys	Gly	Lys	Leu	Ser	Asp	Thr	Thr	Ala	Ser	256
257	Ala	Leu	Thr	Arg	Leu	Gln	Ile	Met	Ala	Ser	Gln	Phe	Ser	Pro	Glu	Gln	272
273	Gln	Lys	Ile	Ile	Thr	Glu	Thr	Val	Gly	Gln	Asp	Phe	Tyr	Gln	Leu	Asn	288
289	Tyr	Gly	Asp	Ser	Ser	Leu	Thr	Val	Asn	Ser	Phe	Ser	Asp	Met	Thr	Ile	304
305	Met	Thr	Asp	Arg	Thr	Ser	Leu	Thr	Val	Pro	Gln	Val	Glu	Leu	Met	Leu	320
321	Cys	Ser	Thr	Val	Gly	Gly	Ser	Thr	Val	Val	Lys	Ser	Asp	Asn	Val	Ser	336
337	Ser	Gly	Asp	Thr	Thr	Ala	Thr	Pro	Phe	Ala	Tyr	Gly	Ala	Arg	Phe	Ile	352
353	His	Ala	Gly	Lys	Pro	Glu	Ala	Ile	Thr	Leu	Ser	Arg	Ser	Gly	Ala	Glu	368
369	Ala	His	Phe	Ala	Leu	Thr	Val	Asn	Asn	Leu	Thr	Asp	Asp	Lys	Leu	Asp	384
385	Arg	Ile	Asn	Arg	Thr	Val	Arg	Leu	Gln	Lys	Trp	Leu	Asn	Leu	Pro	Tyr	400
401	Glu	Asp	Ile	Asp	Leu	Leu	Val	Thr	Ser	Ala	Met	Asp	Ala	Glu	Thr	Gly	416
417	Asn	Thr	Ala	Leu	Ser	Met	Asn	Asp	Asn	Thr	Leu	Arg	Met	Leu	Gly	Val	432
433	Phe	Lys	His	Tyr	Gln	Ala	Lys	Tyr	Gly	Val	Ser	Ala	Lys	Gln	Phe	Ala	448
449	Gly	Trp	Leu	Arg	Val	Val	Ala	Pro	Phe	Ala	Ile	Thr	Pro	Ala	Thr	Pro	464
465	Phe	Leu	Asp	Gln	Val	Phe	Asn	Ser	Val	Gly	Thr	Phe	Asp	Thr	Pro	Phe	480
481	Val	Ile	Asp	Asn	Gln	Asp	Phe	Val	Tyr	Thr	Leu	Thr	Thr	Gly	Gly	Asp	496
497	Gly	Ala	Arg	Val	Lys	His	Ile	Ser	Thr	Ala	Leu	Gly	Leu	Asn	His	Arg	512
513	Gln	Phe	Leu	Leu	Leu	Ala	Asp	Asn	Ile	Ala	Arg	Gln	Gln	Gly	Asn	Val	528
529	Thr	Gln	Ser	Thr	Leu	Asn	Cys	Asn	Leu	Phe	Val	Val	Ser	Ala	Phe	Tyr	544
545	Arg	Leu	Ala	Asn	Leu	Ala	Arg	Thr	Leu	Gly	Ile	Asn	Pro	Glu	Ser	Phe	560
561	Cys	Ala	Leu	Val	Asp	Arg	Leu	Asp	Ala	Gly	Thr	Gly	Ile	Val	Trp	Gln	576
577	Gln	Leu	Ala	Gly	Lys	Pro	Thr	Ile	Thr	Val	Pro	Gln	Lys	Asp	Ser	Pro	592
593	Leu	Ala	Ala	Asp	Ile	Leu	Ser	Leu	Leu	Gln	Ala	Leu	Ser	Ala	Ile	Ala	608
609	Gln	Trp	Gln	Gln	Gln	His	Asp	Leu	Glu	Phe	Ser	Ala	Leu	Leu	Leu	Leu	624

U A 8 2 4 8 5 C 2

625	Leu Ser Asp Asn Pro Ile Ser Thr Ser Gln Gly Thr Asp Asp Gln Leu	640
641	Asn Phe Ile Arg Gln Val Trp Gln Asn Leu Gly Ser Thr Phe Val Gly	655
657	Ala Thr Leu Leu Ser Arg Ser Gly Ala Pro Leu Val Asp Thr Asn Gly	672
673	His Ala Ile Asp Trp Phe Ala Leu Leu Ser Ala Gly Asn Ser Pro Leu	688
689	Ile Asp Lys Val Gly Leu Val Thr Asp Ala Gly Ile Gln Ser Val Ile	704
705	Ala Thr Val Val Asn Thr Gln Ser Leu Ser Asp Glu Asp Lys Lys Leu	720
721	Ala Ile Thr Thr Leu Thr Asn Thr Leu Asn Gln Val Gln Lys Thr Gln	736
737	Gln Gly Val Ala Val Ser Leu Leu Ala Gln Thr Leu Asn Val Ser Gln	752
753	Ser Leu Pro Ala Leu Leu Leu Arg Trp Ser Gly Gln Thr Thr Tyr Gln	768
769	Trp Leu Ser Ala Thr Trp Ala Leu Lys Asp Ala Val Lys Thr Ala Ala	784
785	Asp Ile Pro Ala Asp Tyr Leu Arg Gln Leu Arg Glu Val Val Arg Arg	800
801	Ser Leu Leu Thr Gln Gln Phe Thr Leu Ser Pro Ala Met Val Gln Thr	816
817	Leu Leu Asp Tyr Pro Ala Tyr Phe Gly Ala Ser Ala Glu Thr Val Thr	832
833	Asp Ile Ser Leu Trp Met Leu Tyr Thr Leu Ser Cys Tyr Ser Asp Leu	848
849	Leu Leu Gln Met Gly Glu Ala Gly Gly Thr Glu Asp Asp Val Leu Ala	864
865	Tyr Leu Arg Thr Ala Asn Ala Thr Thr Pro Leu Ser Gln Ser Asp Ala	880
881	Ala Gln Thr Leu Ala Thr Leu Leu Gly Trp Glu Val Asn Glu Leu Gln	896
897	Ala Ala Trp Ser Val Leu Gly Gly Ile Ala Lys Thr Thr Pro Gln Leu	912
913	Asp Ala Leu Leu Arg Leu Gln Gln Ala Gln Asn Gln Thr Gly Leu Gly	928
929	Val Thr Gln Gln Gln Gln Gly Tyr Leu Leu Ser Arg Asp Ser Asp Tyr	944
945	Thr Leu Trp Gln Ser Thr Gly Gln Ala Leu Val Ala Gly Val Ser His	960
961	Val Lys Gly Ser Asn	965

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 58:

(i) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 4698 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: ДНК (геномна)

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 58 (tccB):

1	ATG TTA TCG ACA ATG GAA AAA CAA CTG AAT GAA TCC CAG CGT GAT GCG	43
1	Met Leu Ser Thr Met Glu Lys Gln Leu Asn Glu Ser Gln Arg Asp Ala	15
49	TTG GTG ACT GGC TAT ATG AAT TTT GTG GCG CCG ACG TTG AAA GCG CTC	96
17	Leu Val Thr Gly Tyr Met Asn Phe Val Ala Pro Thr Leu Lys Gly Val	32

97	AGT GGT CAG CCG GTC ACC GTC GAA GAT TTA TAC GAA TAT TTC CTG ATT	144
33	Ser Gly Gln Pro Val Thr Val Glu Asp Leu Tyr Glu Tyr Leu Leu Ile	48
145	GAC CCG GAA GTC GCT GAT CAG GTT GAG ACC AGT CCG GTA GCA CAA GCG	192
49	Asp Pro Glu Val Ala Asp Glu Val Glu Thr Ser Arg Val Ala Gln Ala	64
193	ATT GCC AGC ATA CAG CAA TAT ATG ACT CGT CTG GTC AAC GGC TCT GAA	240
65	Ile Ala Ser Ile Gln Gln Tyr Met Thr Arg Leu Val Asn Gly Ser Glu	80
241	CCG CCG CGT CAG CCG ATG CAG CCT TCT ACA GCT AAC GAA TGG COT GAT	288
81	Pro Gly Arg Gln Ala Met Glu Pro Ser Thr Ala Asn Glu Trp Arg Asp	96
289	AAT GAT AAC CAA TAT GCT ATC TGG GCT GCG GCG GCT GAG GTT CGA AAT	336
97	Asn Asp Asn Gln Tyr Ala Ile Trp Ala Ala Gly Ala Glu Val Arg Asn	112
337	TAC GCT GAA AAC TAT ATT TCA CCC ATC ACC CCG CAG GAA AAA AGC CAT	384
113	Tyr Ala Glu Asn Tyr Ile Ser Pro Ile Thr Arg Gln Glu Lys Ser His	128
385	TAT TTC TCG CAG CTC GAG ACC ACT TTA AAT CAG AAT CGA CTC GAT CCG	432
129	Tyr Phe Ser Glu Leu Glu Thr Thr Leu Asn Gln Asn Arg Leu Asp Pro	144
433	GAT CGT GTG CAG GAT GCT GTT TTC GCG TAT CTC AAT CAG TTT GAG GCA	480
145	Asp Arg Val Gln Asp Ala Val Leu Ala Tyr Leu Asn Glu Phe Glu Ala	160
481	GTG AGT AAT CTA TAT GTG CTC AGT GGT TAT ATT AAT CAG GAT AAA TTT	528
161	Val Ser Asn Leu Tyr Val Leu Ser Gly Tyr Ile Asn Gln Asp Lys Phe	176
529	GAC CAA GCT ATC TAC TAC TTT ATT GGT CCG ACT ACC ACT AAA CCG TAT	576
177	Asp Gln Ala Ile Tyr Tyr Phe Ile Gly Arg Thr Thr Thr Lys Pro Tyr	192
577	CCG TAC TAC TCG CGT CAG ATG GAT TTG AGT AAG AAC CGT CAA GAT CCG	624
193	Arg Tyr Tyr Trp Arg Gln Met Asp Leu Ser Lys Asn Arg Gln Asp Pro	208
625	GCA GCG AAT CCG GTC ACC CCA AAT TGC TGC AAT GAT TGG CAG GAA ATC	672
209	Ala Gly Asn Pro Val Thr Pro Asn Cys Trp Asn Asp Trp Gln Glu Ile	224
673	ACT TTG CCG CTG TCT GGT GAT ACC GTG CTG GAG CAT ACA GTT CCG CCG	720
225	Thr Leu Pro Leu Ser Gly Asp Thr Val Leu Glu His Thr Val Arg Pro	240
721	GTA TTT TAT AAT GAT CGA CTA TAT GTG GCT TGG GTT CAG COT GAC CCG	768
241	Val Phe Tyr Asn Asp Arg Leu Tyr Val Ala Trp Val Glu Arg Asp Pro	256
769	GCA GTA CAG AAG GAT GCT GAC GGT AAA AAC ATC GGT AAA ACC CAT GCC	816
257	Ala Val Gln Lys Asp Ala Asp Gly Lys Asn Ile Gly Lys Thr His Ala	272
817	TAC AAC ATA AAG TTT GGT TAT AAA CGT TAT GAT GAT ACT TGG ACA GCG	864
273	Tyr Asn Ile Lys Phe Gly Tyr Lys Arg Tyr Asp Asp Thr Trp Thr Ala	288
865	CCG AAT ACC ACC ACC TTA ATG ACA CAA CAA GCA GCG GAA AGT TCA GAA	912
289	Pro Asn Thr Thr Thr Leu Met Thr Gln Gln Ala Gly Glu Ser Ser Glu	304

913	ACA CAG CGA TCC AGC CTC CTC ATT CAT GAA TCT AGC ACC ACA TTG CCC	963
105	Thr Gln Arg Ser Ser Leu Leu Ile Asp Glu Ser Ser Thr Thr Leu Arg	120
961	CAA GTT AAT CTC TTG GCT ACC ACC GAT TTT ACT ATC GAT CCG ACC GAG	1008
121	Gln Val Asn Leu Leu Ala Thr Thr Asp Phe Ser Ile Asp Pro Thr Glu	1136
1009	GAA ACC GAC AGT AAC CCG TAT GGC CCG CTA ATC TTG GGG CTC TTT GTC	1056
137	Glu Thr Asp Ser Asn Pro Tyr Gly Arg Leu Met Leu Gly Val Phe Val	1152
1057	CGT CAA TTT GAA GGT CAT GCG GCC AAT AGA AAA AAT AAA CCC GTT GTT	1104
153	Arg Gln Phe Glu Gly Asp Gly Ala Asn Arg Lys Asn Lys Pro Val Val	1168
1105	TAT GGT TAT CTC TAT TGT GAC TCA GGT TTC AAT CGT CAT GTT CTC AGG	1152
169	Tyr Gly Tyr Leu Tyr Cys Asp Ser Ala Phe Asn Arg His Val Leu Arg	1184
1153	CCG TTA AGT AAG AAC TTT TTG TTC AGT ACT TAC CGT GAT GAA ACC GAT	1200
185	Pro Leu Ser Lys Asn Phe Leu Phe Ser Thr Tyr Arg Asp Glu Thr Asp	1200
1201	GGT CAA AAC AGC TTG CAA TTT GCG CTA TAC GAT AAA AAG TAT CTA ATT	1248
401	Gly Gln Asn Ser Leu Gln Phe Ala Val Tyr Asp Lys Lys Tyr Val Ile	1216
1249	ACT AAG GTT GTT ACA GGT GCA ACC GAA GAT CCC GAA AAT ACA GGA TGG	1296
417	Thr Lys Val Val Thr Gly Ala Thr Glu Asp Pro Glu Asn Thr Gly Trp	1312
1297	GTA AGT AAA GTT GAT GAC TTG AAA CAA GGC ACT ACT GCG GCC TAT CTC	1344
433	Val Ser Lys Val Asp Asp Leu Lys Gln Gly Thr Thr Gly Ala Tyr Val	1348
1345	TAT ATC GAT CAA GAT GCG CTC ACC CTT CAT ATA CAA ACC ACA ACT AAT	1392
449	Tyr Ile Asp Gln Asp Gly Leu Thr Leu His Ile Gln Thr Thr Thr Asn	1404
1393	GCG GAT TTT ATT AAC CGT CAT ACC TTT GGA TAT AAC GAT CTT CTA TAT	1440
465	Gly Asp Phe Ile Asn Arg His Thr Phe Gly Tyr Asn Asp Leu Val Tyr	1440
1441	GAT TCT AAG TCT GGT TAT GGT TTC ACC TGG TCA GGA AAT GAA GGT TTT	1488
481	Asp Ser Lys Ser Gly Tyr Gly Phe Thr Trp Ser Gly Asn Glu Gly Phe	1496
1489	TAT CTC GAT TAC CAT GAT GGA AAT TAT TAC ACC TTT CAT AAT GCA ATA	1536
497	Tyr Leu Asp Tyr His Asp Gly Asn Tyr Tyr Thr Phe His Asn Ala Ile	1512
1537	ATC AAC TAC TAT CCG TCT GGA TAT GGT GGT GGA TCT GTT CCT AAT GGA	1584
513	Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Gly Ser Val Pro Asn Gly	1512
1585	ACG TGG GCG TTA GAG CAA AGC ATT AAT CAG GGA TGG GCT ATT GCT CCC	1632
529	Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro	1614
1633	CTG CTT GAT ACT CTC CAT ACT GTT ACT CTC AAG GCG AGT TAT ATC GCT	1680
545	Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala	1660

1431	TCG GAA GCG GAA ACA CCT ACC GGT TAT AAT GTS TAT ATT GCA GAT GGT	1728
561	Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly	576
1729	ACC CTG TTG CTA CAT TCG TTT GAT AAA ATA AAT TTT GGT ATT GGT CTT	1776
577	Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu	592
1777	AAT AAG CTT GAG TCT GTA TTT ACC TCG CCA GAT TCG CCA ACA CTA ACC	1824
593	Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr	608
1825	ACT ATC AAA AAT TTC AGT AAA ATC GCC GAT AAC CCG AAA TTC TAT CAG	1872
609	Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln	624
1873	GAA ATC AAT GCT GAG ACG GCG GAT GGA CCG AAC CTG TTT AAA CGT TAC	1920
625	Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr	640
1921	AGT ACT CAA ACT TTC CCA CTT ACC AGC GGT CCG ACT TAT TCT ACA ACT	1968
641	Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr	656
1969	TAT ACT TTC TCT GAG CCG GAT TTC TCC ACT GAT CCG GAC AAA AAC TAC	2016
657	Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr	672
2017	CTA CAG GTT TGT TTG AAT GTC CTG TCG GAT CAT TAT GAC CCG CCG TCA	2064
673	Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser	688
2065	GCG AAA AAA GCG GCT TAT TCT TCG GTC AGT AAG TCG TTT AAC GTC TAT	2112
689	Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr	704
2113	GTT GCG TTG CAA GAT AGC AAA GCT CCG GAT GCC ATT CCT CCA TTA GTT	2160
705	Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val	720
2161	TCC CGT TAC GAT AGT AAA COT GGT CTG GTG CAA TAT CTG GAC TTC TCG	2208
721	Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp	736
2209	ACC TCA TCA TTA CCC GCG AAA ACC GGT CTT AAC ACC ACC TTT GTG CGT	2256
737	Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg	752
2257	ACT TTG ATT GAG AAG GCT AAT CTG GCG CTG GAT AGT TTG CTG GAT TAC	2304
753	Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr	768
2305	ACC TTG CAG GCA GAT CCT TCT CTG GAA GCA GAT TTA CTG ACT GAC GCG	2352
769	Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly	784
2353	AAA AGC GAA CCA ATG GAC TTT AAT GGT TCA AAC GGT CTC TAT TTC TCG	2400
785	Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp	800
2401	GAA TTG TTC TTT CAC CTG CCG TTT TTG GTT GCT ACA CCG TTT GCG AAC	2448
801	Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn	816
2449	GAA CAG CAA TTT TCG CCG GCA CAA AAG AGT TTG CAT TAC ATC TTT GAC	2496
817	Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp	832

U A 8 2 4 8 5 C 2

2497 333	CCG GCG ATG AAA AAC AAG CCA CAC AAT CCG CCG GCT TAT TCG AAT GTG Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val	2544 348
2545 349	GCT CCG TTG GTT GAA GGA AAC AGC GAT TTG TCA GGT CAT TTG GAC GAT Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp	2592 364
2593 365	TCT ATA CAC CCA GAT ACT CAA GCT TAT GCT CAT CCG CTC ATA TAC CAG Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln	2640 380
2641 381	AAA GCG CTC TTT ATT GCG TAT GTC AGT AAC CTC ATT GCT CAG GCA GAT Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp	2688 398
2689 397	ATG TCG TAT CCG CAA TTG ACT GGT GAC GGT CTC ACT CAG GCC GGT CTC Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val	2736 412
2737 413	TAT TAC AAT CTC GCG GCT GAA TTG CTA GCG GCT GCT CCG GAT CTA TCG Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser	2784 428
2785 429	CTG ACT AGC ATT TCG ACC CCG CAA ACC CTC GAT ACC TTA GCA GCC GCG Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly	2832 444
2833 445	CAA AAA GCG GTT TTA GGT GAT TTT GAG CAC CAG TTG GGT AAT AGT GAT Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp	2880 460
2881 461	ACC GCT TTA CCC GCA TTG CCG GCG CCG AAT CTC ACC TAC TTG AAA CTC Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu	2928 476
2929 477	GCA GAT AAT GCG TAC TTT AAT GAA CCG CTC AAT GTT CTC ATG TTG TCT Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser	2976 492
2977 493	CAC TCG GAT ACC TTG GAT CCA CCG TTA TAC AAT CTC GGT CAT AAC CTC His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu	3024 508
3025 1009	ACC GTT GAT GGC AAG CCG CTT TCG CTC CCG CTC TAT GGT GCG CCT GTT Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val	3072 1024
3073 1025	GAT CCG GTA GCG TTG TTG GGT CAG GGT GGT CAG TCC GGC ACC TTG ACC Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr	3120 1040
3121 1041	AAT GCG GTC AGT GCG CCG ATG TTG ACC CTC CCG CCA TAC GGT TTC AGC Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser	3168 1056
3169 1057	GCT ATG TTG CCG CGA GCT TAC ACC GCG CTC GGT ACC TTG ACC ACT TTT Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe	3216 1072
3217 1073	GCT CAG AAC CTC CTT AGT TTG TTG GAA GGT ACC GAA CGA GCC TGT GAA Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln	3264 1088

U A 8 2 4 8 5 C 2

1085	GAA GAG TTC GCG CAA CAG CAA CTC TTG GAT ATG TCC AGC TAT GCG ATC	1111
1089	Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile	1105
1103	ACG TTG CAA CAA CAG GCG CTC GAT GGA TTG GCG GCA GAT CGT CTC GCG	1169
1105	Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala	1129
1121	CTG CTA GCT ACT CAG GCT ACC GCA CAA CAG CGT CAT GAC CAT TAT TAC	1133
1121	Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr	1135
1409	ACT CTG TAT CAG AAC AAC ATC TCC AGT GCG CAA CAA CTC GTG ATG GAC	1456
1137	Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp	1153
1457	ACC CAA ACC TCA GCA CAA TCC CTC ATT TCT TCT TCC ACT GGT GTA CAA	1504
1153	Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln	1168
1505	ACT GCG AGT GCG GCA CTC AAA GTG ATC CCG AAT ATC TTT GGT TTG GCT	1552
1169	Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala	1184
1553	GAT GCG GCG TCG CCG TAT GAA GGA GTA ACG GAA GCG ATT GCG ATC GCG	1600
1185	Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly	1200
1601	TTA ATG GCT GCC GGA CAA GCC ACC ACC CTC GTG GCG GAG CGT CTC GCA	1648
1201	Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala	1216
1649	ACC ACC CAG AAT TAC CCG CCG CCG CGT GAA GAG TCG CAA ATC CAA TAC	1696
1217	Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr	1232
1697	CAG CAG GCA CAG TCT GAG GTC GAC GCA TTA CAG AAA CAG TTG GAT GCG	1744
1233	Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala	1248
1745	CTG GCA GTC CCG CAG AAA GCA GCT CAA ACT TCC CTC CAA CAG GCG AAG	1792
1249	Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys	1264
1793	GCA CAG CAG GTA CAA ATT CCG ACC ATG CTC ACT TAC TTA ACT ACT CGT	1840
1265	Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg	1280
1841	TTG ACC CAG GCG ACT CTC TAC CAG TCG CTC AGT GGT CAA TTA TCC GCG	1888
1281	Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala	1296
1889	TTG TAT TAT CAA GCG TAT GAT GCG CTC GTT GCT CTC TCC CTC TCC GCG	1936
1297	Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala	1312
1937	CAA GCT TCC TCG CAG TAT GAA TTG GGT GAT TAC GCT ACC ACT TTT ATC	1984
1313	Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile	1328
1985	CAG ACC GGT ACC TCG AAC GAC CAT TAC CGT GGT TTG CAA GTG GCG GAG	2032
1329	Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu	1344
2033	ACA CTC CAA CTC AAT TTG CAT CAG ATG GAA GCG GCG TAT TTA GTT CGT	2080
1345	Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg	1360

4081	CAC GAA CCC GGT CTT AAT GTG ATC GGT ACT GTG TCG CTC AAA AGC CTA	4123
1361	His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu	1376
4129	TTG GGT GAT GAT GGT TTT GGT AAG TTA AIA ACC GAA GGC AAA CTC GAC	4176
1377	Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp	1392
4177	TTT CCA TTA ACC GAA AAG CTC TTT GAC AAC GAC TAT CCG GGC CAC TAT	4224
1393	Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr	1406
4225	TTG CGC CAG ATT AAA ACT GTG TCA GTG ACC TTG CCG ACC TTA GTC GCG	4272
1409	Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly	1424
4273	CCG TAT CAA AAC CTC AAG GCA ACG CTC ACT CAG ACC AGC AGC AGT ATA	4320
1425	Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile	1440
4321	TTG TTA GCA GCA GAT ATC AAT GGT GTT AAA CGT CTC AAT GAT CCG ACA	4368
1441	Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr	1456
4369	GGT AAA GAG GGT GAT GCG ACC CAT ATT GTC ACC AAT CTC CGT GCC AGC	4416
1457	Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser	1472
4417	CAG CAG GTG GCG CTC TCT TCT GGC ATT AAT GAT GCC GGT ACC TTT CAG	4464
1473	Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu	1488
4465	TTG CGT TTG GAA GAT CAG CGC TAT CTA TCA TTT CAG GCG ACT GGA GCT	4512
1489	Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala	1504
4513	GTT TCC AAA TCG ACT CTT AAC TTC CCG CGT TCT GTG GAT GAG CAT ATT	4560
1505	Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile	1520
4561	GAC GAT AAG ACA TTG AAA GCG GAT CAG ATG CAG GCG GCA CTC TTG GCG	4608
1521	Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala	1536
4609	AAT ATG GAT GAT GTG CTC GTG CAG GTG CAT TAT ACC GCG TCG CAC GCG	4656
1537	Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly	1552
4657	GCG GCC AGT TTC GCA AAC CAG GTC AAG AAA ACA CTC TCT TAA	4698
1553	Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser End	1566

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 59:

(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 1665 амінокислот
- (B) ТИП: амінокислота
- (C) ТОПОЛОГІЯ: лінійна

(ii) ТИП МОЛЕКУЛИ: білок

(v) ТИП ФРАГМЕНТУ: N-кінцевий

(xi) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 59 (TccB пептид):

Ознаки	Від	До	Опис
	1	11	SEQ ID № 7

U A 8 2 4 8 5 C 2

1	Met	Leu	Ser	Thr	Met	Glu	Lys	Gln	Leu	Asn	Glu	Ser	Gln	Arg	Asp	Ala	15
17	Leu	Val	Thr	Gly	Tyr	Met	Asn	Phe	Val	Ala	Pro	Thr	Leu	Lys	Gly	Val	32
33	Ser	Gly	Gln	Pro	Val	Thr	Val	Glu	Asp	Leu	Tyr	Glu	Tyr	Leu	Leu	Ile	43
49	Asp	Pro	Glu	Val	Ala	Asp	Glu	Val	Glu	Thr	Ser	Arg	Val	Ala	Gln	Ala	54
65	Ile	Ala	Ser	Ile	Gln	Gln	Tyr	Met	Thr	Arg	Leu	Val	Asn	Gly	Ser	Glu	80
81	Pro	Gly	Arg	Gln	Ala	Met	Glu	Pro	Ser	Thr	Ala	Asn	Glu	Trp	Arg	Asp	96
97	Asn	Asp	Asn	Gln	Tyr	Ala	Ile	Trp	Ala	Ala	Gly	Ala	Glu	Val	Arg	Asn	112
113	Tyr	Ala	Glu	Asn	Tyr	Ile	Ser	Pro	Ile	Thr	Arg	Gln	Glu	Lys	Ser	His	123
129	Tyr	Phe	Ser	Glu	Leu	Glu	Thr	Thr	Leu	Asn	Gln	Asn	Arg	Leu	Asp	Pro	144
145	Asp	Arg	Val	Gln	Asp	Ala	Val	Leu	Ala	Tyr	Leu	Asn	Glu	Phe	Glu	Ala	160
161	Val	Ser	Asn	Leu	Tyr	Val	Leu	Ser	Gly	Tyr	Ile	Asn	Gln	Asp	Lys	Phe	176
177	Asp	Gln	Ala	Ile	Tyr	Tyr	Phe	Ile	Gly	Arg	Thr	Thr	Thr	Lys	Pro	Tyr	192
193	Arg	Tyr	Tyr	Trp	Arg	Gln	Met	Asp	Leu	Ser	Lys	Asn	Arg	Gln	Asp	Pro	208
209	Ala	Gly	Asn	Pro	Val	Thr	Pro	Asn	Cys	Trp	Asn	Asp	Trp	Gln	Glu	Ile	224
225	Thr	Leu	Pro	Leu	Ser	Gly	Asp	Thr	Val	Leu	Glu	His	Thr	Val	Arg	Pro	240
241	Val	Phe	Tyr	Asn	Asp	Arg	Leu	Tyr	Val	Ala	Trp	Val	Glu	Arg	Asp	Pro	256
257	Ala	Val	Gln	Lys	Asp	Ala	Asp	Gly	Lys	Asn	Ile	Gly	Lys	Thr	His	Ala	272
273	Tyr	Asn	Ile	Lys	Phe	Gly	Tyr	Lys	Arg	Tyr	Asp	Asp	Thr	Trp	Thr	Ala	288
289	Pro	Asn	Thr	Thr	Thr	Leu	Met	Thr	Gln	Gln	Ala	Gly	Glu	Ser	Ser	Glu	304
305	Thr	Gln	Arg	Ser	Ser	Leu	Leu	Ile	Asp	Glu	Ser	Ser	Thr	Thr	Leu	Arg	320
321	Gln	Val	Asn	Leu	Leu	Ala	Thr	Thr	Asp	Phe	Ser	Ile	Asp	Pro	Thr	Glu	336
337	Glu	Thr	Asp	Ser	Asn	Pro	Tyr	Gly	Arg	Leu	Met	Leu	Gly	Val	Phe	Val	352
353	Arg	Gln	Phe	Glu	Gly	Asp	Gly	Ala	Asn	Arg	Lys	Asn	Lys	Pro	Val	Val	368
369	Tyr	Gly	Tyr	Leu	Tyr	Cys	Asp	Ser	Ala	Phe	Asn	Arg	His	Val	Leu	Arg	384
385	Pro	Leu	Ser	Lys	Asn	Phe	Leu	Phe	Ser	Thr	Tyr	Arg	Asp	Glu	Thr	Asp	400
401	Gly	Gln	Asn	Ser	Leu	Gln	Phe	Ala	Val	Tyr	Asp	Lys	Lys	Tyr	Val	Ile	416
417	Thr	Lys	Val	Val	Thr	Gly	Ala	Thr	Glu	Asp	Pro	Glu	Asn	Thr	Gly	Trp	432
433	Val	Ser	Lys	Val	Asp	Asp	Leu	Lys	Gln	Gly	Thr	Thr	Gly	Ala	Tyr	Val	448
449	Tyr	Ile	Asp	Gln	Asp	Gly	Leu	Thr	Leu	His	Ile	Gln	Thr	Thr	Thr	Asn	464
465	Gly	Asp	Phe	Ile	Asn	Arg	His	Thr	Phe	Gly	Tyr	Asn	Asp	Leu	Val	Tyr	480
481	Asp	Ser	Lys	Ser	Gly	Tyr	Gly	Phe	Thr	Trp	Ser	Gly	Asn	Glu	Gly	Phe	496
497	Tyr	Leu	Asp	Tyr	His	Asp	Gly	Asn	Tyr	Tyr	Thr	Phe	His	Asn	Ala	Ile	512

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

613	Ile Asn Tyr Tyr Pro Ser Gly Tyr Gly Gly Gly Ser Val Pro Asn Gly	623
629	Thr Trp Ala Leu Glu Gln Arg Ile Asn Glu Gly Trp Ala Ile Ala Pro	644
645	Leu Leu Asp Thr Leu His Thr Val Thr Val Lys Gly Ser Tyr Ile Ala	660
661	Trp Glu Gly Glu Thr Pro Thr Gly Tyr Asn Leu Tyr Ile Pro Asp Gly	676
677	Thr Val Leu Leu Asp Trp Phe Asp Lys Ile Asn Phe Ala Ile Gly Leu	692
693	Asn Lys Leu Glu Ser Val Phe Thr Ser Pro Asp Trp Pro Thr Leu Thr	608
609	Thr Ile Lys Asn Phe Ser Lys Ile Ala Asp Asn Arg Lys Phe Tyr Gln	624
625	Glu Ile Asn Ala Glu Thr Ala Asp Gly Arg Asn Leu Phe Lys Arg Tyr	640
641	Ser Thr Gln Thr Phe Gly Leu Thr Ser Gly Ala Thr Tyr Ser Thr Thr	656
657	Tyr Thr Leu Ser Glu Ala Asp Phe Ser Thr Asp Pro Asp Lys Asn Tyr	672
673	Leu Gln Val Cys Leu Asn Val Val Trp Asp His Tyr Asp Arg Pro Ser	688
689	Gly Lys Lys Gly Ala Tyr Ser Trp Val Ser Lys Trp Phe Asn Val Tyr	704
705	Val Ala Leu Gln Asp Ser Lys Ala Pro Asp Ala Ile Pro Arg Leu Val	720
721	Ser Arg Tyr Asp Ser Lys Arg Gly Leu Val Gln Tyr Leu Asp Phe Trp	736
737	Thr Ser Ser Leu Pro Ala Lys Thr Arg Leu Asn Thr Thr Phe Val Arg	752
753	Thr Leu Ile Glu Lys Ala Asn Leu Gly Leu Asp Ser Leu Leu Asp Tyr	768
769	Thr Leu Gln Ala Asp Pro Ser Leu Glu Ala Asp Leu Val Thr Asp Gly	784
795	Lys Ser Glu Pro Met Asp Phe Asn Gly Ser Asn Gly Leu Tyr Phe Trp	800
801	Glu Leu Phe Phe His Leu Pro Phe Leu Val Ala Thr Arg Phe Ala Asn	816
817	Glu Gln Gln Phe Ser Pro Ala Gln Lys Ser Leu His Tyr Ile Phe Asp	832
833	Pro Ala Met Lys Asn Lys Pro His Asn Ala Pro Ala Tyr Trp Asn Val	848
849	Arg Pro Leu Val Glu Gly Asn Ser Asp Leu Ser Arg His Leu Asp Asp	864
865	Ser Ile Asp Pro Asp Thr Gln Ala Tyr Ala His Pro Val Ile Tyr Gln	880
881	Lys Ala Val Phe Ile Ala Tyr Val Ser Asn Leu Ile Ala Gln Gly Asp	896
897	Met Trp Tyr Arg Gln Leu Thr Arg Asp Gly Leu Thr Gln Ala Arg Val	912
913	Tyr Tyr Asn Leu Ala Ala Glu Leu Leu Gly Pro Arg Pro Asp Val Ser	928
929	Leu Ser Ser Ile Trp Thr Pro Gln Thr Leu Asp Thr Leu Ala Ala Gly	944
945	Gln Lys Ala Val Leu Arg Asp Phe Glu His Gln Leu Ala Asn Ser Asp	960
961	Thr Ala Leu Pro Ala Leu Pro Gly Arg Asn Val Ser Tyr Leu Lys Leu	976
977	Ala Asp Asn Gly Tyr Phe Asn Glu Pro Leu Asn Val Leu Met Leu Ser	992
993	His Trp Asp Thr Leu Asp Ala Arg Leu Tyr Asn Leu Arg His Asn Leu	1009
1009	Thr Val Asp Gly Lys Pro Leu Ser Leu Pro Leu Tyr Ala Ala Pro Val	1024
1025	Asp Pro Val Ala Leu Leu Ala Gln Arg Ala Gln Ser Gly Thr Leu Thr	1040

U A 8 2 4 8 5 C 2

U A 8 2 4 8 5 C 2

1041 Asn Gly Val Ser Gly Ala Met Leu Thr Val Pro Pro Tyr Arg Phe Ser 1056
 1057 Ala Met Leu Pro Arg Ala Tyr Ser Ala Val Gly Thr Leu Thr Ser Phe 1071
 1073 Gly Gln Asn Leu Leu Ser Leu Leu Glu Arg Ser Glu Arg Ala Cys Gln 1088
 1089 Glu Glu Leu Ala Gln Gln Gln Leu Leu Asp Met Ser Ser Tyr Ala Ile 1104
 1105 Thr Leu Gln Gln Gln Ala Leu Asp Gly Leu Ala Ala Asp Arg Leu Ala 1120
 1121 Leu Leu Ala Ser Gln Ala Thr Ala Gln Gln Arg His Asp His Tyr Tyr 1136
 1137 Thr Leu Tyr Gln Asn Asn Ile Ser Ser Ala Glu Gln Leu Val Met Asp 1152
 1153 Thr Gln Thr Ser Ala Gln Ser Leu Ile Ser Ser Ser Thr Gly Val Gln 1168
 1169 Thr Ala Ser Gly Ala Leu Lys Val Ile Pro Asn Ile Phe Gly Leu Ala 1184
 1185 Asp Gly Gly Ser Arg Tyr Glu Gly Val Thr Glu Ala Ile Ala Ile Gly 1200
 1201 Leu Met Ala Ala Gly Gln Ala Thr Ser Val Val Ala Glu Arg Leu Ala 1216
 1217 Thr Thr Glu Asn Tyr Arg Arg Arg Arg Glu Glu Trp Gln Ile Gln Tyr 1232
 1233 Gln Gln Ala Gln Ser Glu Val Asp Ala Leu Gln Lys Gln Leu Asp Ala 1248
 1249 Leu Ala Val Arg Glu Lys Ala Ala Gln Thr Ser Leu Gln Gln Ala Lys 1264
 1265 Ala Gln Gln Val Gln Ile Arg Thr Met Leu Thr Tyr Leu Thr Thr Arg 1280
 1281 Phe Thr Gln Ala Thr Leu Tyr Gln Trp Leu Ser Gly Gln Leu Ser Ala 1296
 1297 Leu Tyr Tyr Gln Ala Tyr Asp Ala Val Val Ala Leu Cys Leu Ser Ala 1312
 1313 Gln Ala Cys Trp Gln Tyr Glu Leu Gly Asp Tyr Ala Thr Thr Phe Ile 1328
 1329 Gln Thr Gly Thr Trp Asn Asp His Tyr Arg Gly Leu Gln Val Gly Glu 1344
 1345 Thr Leu Gln Leu Asn Leu His Gln Met Glu Ala Ala Tyr Leu Val Arg 1360
 1361 His Glu Arg Arg Leu Asn Val Ile Arg Thr Val Ser Leu Lys Ser Leu 1376
 1377 Leu Gly Asp Asp Gly Phe Gly Lys Leu Lys Thr Glu Gly Lys Val Asp 1392
 1393 Phe Pro Leu Ser Glu Lys Leu Phe Asp Asn Asp Tyr Pro Gly His Tyr 1408
 1409 Leu Arg Gln Ile Lys Thr Val Ser Val Thr Leu Pro Thr Leu Val Gly 1424
 1425 Pro Tyr Gln Asn Val Lys Ala Thr Leu Thr Gln Thr Ser Ser Ser Ile 1440
 1441 Leu Leu Ala Ala Asp Ile Asn Gly Val Lys Arg Leu Asn Asp Pro Thr 1456
 1457 Gly Lys Glu Gly Asp Ala Thr His Ile Val Thr Asn Leu Arg Ala Ser 1472
 1473 Gln Gln Val Ala Leu Ser Ser Gly Ile Asn Asp Ala Gly Ser Phe Glu 1488
 1489 Leu Arg Leu Glu Asp Glu Arg Tyr Leu Ser Phe Glu Gly Thr Gly Ala 1504
 1505 Val Ser Lys Trp Thr Leu Asn Phe Pro Arg Ser Val Asp Glu His Ile 1520
 1521 Asp Asp Lys Thr Leu Lys Ala Asp Glu Met Gln Ala Ala Leu Leu Ala 1536
 1537 Asn Met Asp Asp Val Leu Val Gln Val His Tyr Thr Ala Cys Asp Gly 1552
 1553 Gly Ala Ser Phe Ala Asn Gln Val Lys Lys Thr Leu Ser 1565

U A 8 2 4 8 5 C 2

(2) ІНФОРМАЦІЯ ДЛЯ SEQ ID № 60:

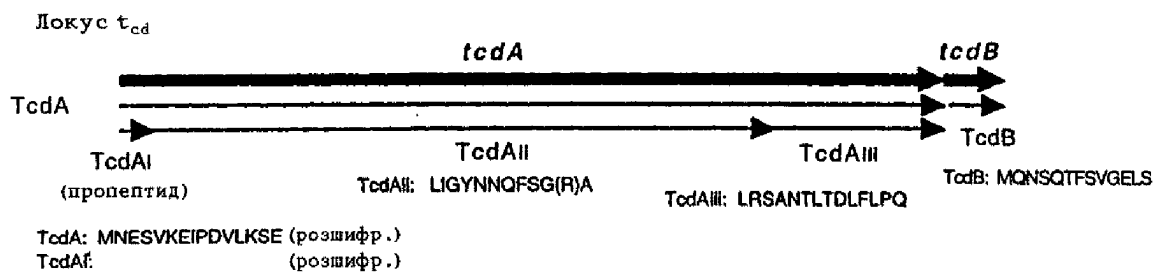
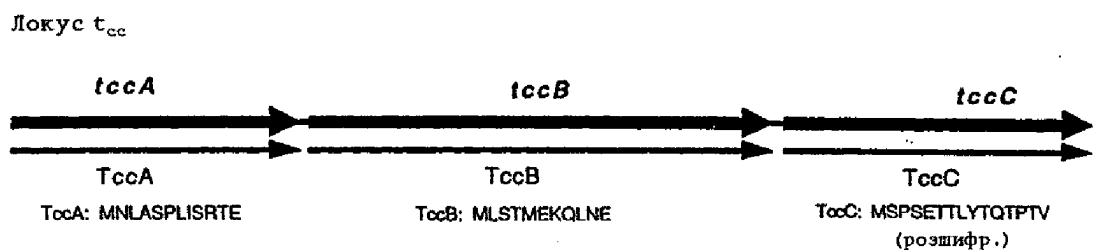
(1) ХАРАКТЕРИСТИКИ ПОСЛІДОВНОСТІ:

- (A) ДОВЖИНА: 3132 п.н.
- (B) ТИП: нуклеїнова кислота
- (C) КІЛЬКІСТЬ ЛАНЦЮГІВ: два
- (D) ТОПОЛОГІЯ: лінійна

(11) ТИП МОЛЕКУЛИ: ДНК (геномна)

(x1) ОПИС ПОСЛІДОВНОСТІ: SEQ ID № 60 (tccc):

1	ATG AGT CCC TCT CAG ACT ACT CTT TAT ACT CAA ACC CCA ACA GTC ACC	48
1	Met Ser Pro Ser Glu Thr Thr Leu Tyr Thr Gln Thr Pro Thr Val Ser	16
49	GTC TTA GAT AAT CCC GGT CTG TCC ATT CGT GAT ATT GGT TTT CAG CGT	96
17	Val Leu Asp Asn Arg Gly Leu Ser Ile Arg Asp Ile Gly Phe His Arg	12
97	ATT GTA ATC GGG GGG GAT ACT GAC ACC CGC GTC ACC CGT CAG CAG TAT	144
33	Ile Val Ile Gly Gly Asp Thr Asp Thr Arg Val Thr Arg His Gln Tyr	48
145	GAT CCC CGT GGA CAC CTG AAC TAC AGT ATT GAC CCA CGC TTG TAT GAT	192
49	Asp Ala Arg Gly His Leu Asn Tyr Ser Ile Asp Pro Arg Leu Tyr Asp	64
193	GCA AAG CAG GCT CAT AAC TCA GTA AAG CCT AAT TTT GTC TCG CAG CAT	240
65	Ala Lys Gln Ala Asp Asn Ser Val Lys Pro Asn Phe Val Trp Gln His	80
241	GAT CTG GCC GGT CAT GCC CTG CCG ACA GAG AGT GTC GAT GCT GGT CGT	288
81	Asp Leu Ala Gly His Ala Leu Arg Thr Glu Ser Val Asp Ala Gly Arg	96
289	ACT GTT GCA TTG AAT GAT ATT GAA GGT CGT TCG GTA ATG ACA ATG AAT	336
97	Thr Val Ala Leu Asn Asp Ile Glu Gly Arg Ser Val Met Thr Met Asn	112
337	CGC ACC GGT GTT CGT CAG ACC CCT CGC TAT GAA GGC AAC ACC TTG CCC	384
113	Ala Thr Gly Val Arg Gln Thr Arg Arg Tyr Glu Gly Asn Thr Leu Pro	128
385	GGT CGC TTG TTA TCT GTC AGC GAG CAA GTT TTC AAC CAA GAG AGT GCT	432
129	Gly Arg Leu Leu Ser Val Ser Glu Gln Val Phe Asn Gln Glu Ser Ala	144
433	AAA GTG ACA CAG CCC TTT ATC TCG CCT GCG AAT ACA ACC TCG GAG AAA	480
145	Lys Val Thr Glu Arg Phe Ile Trp Ala Gly Asn Thr Thr Ser Glu Lys	160
481	GAG TAT AAC CTC TCC GGT CTG TCT ATA CSC CAC TAC GAC ACA GCG GGA	528
161	Glu Tyr Asn Leu Ser Gly Leu Cys Ile Arg His Tyr Asp Thr Ala Gly	176
529	CTG ACC CGG TTG ATG AGT CAG TCA CTG GCG GCG GCG ATG CTA TCC CAA	576
177	Val Thr Arg Leu Met Ser Gln Ser Leu Ala Gly Ala Met Leu Ser Gln	192
577	TCT CAC CAA TTG CTC CCG CAA GCG CAG GAG GCT AAC TCG AGC GGT GAC	624
193	Ser His Gln Leu Leu Ala Glu Gly Gln Glu Ala Asn Trp Ser Gly Asp	208



Фіг. 6В

Офіційний бюлетень "Промислова власність". Книга 1 "Винаходи, корисні моделі, топографії інтегральних мікросхем", 2008, N 8, 25.04.2008. Державний департамент інтелектуальної власності Міністерства освіти і науки України.