

(19)



**Евразийское
патентное
ведомство**

(11) **041113**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**(45) Дата публикации и выдачи патента
2022.09.15(21) Номер заявки
201891127(22) Дата подачи заявки
2016.12.22(51) Int. Cl. **C12N 15/62** (2006.01)
A61K 38/00 (2006.01)
C07K 14/00 (2006.01)
C12N 15/11 (2006.01)
C12N 15/64 (2006.01)

**(54) НУКЛЕИНОВЫЕ КИСЛОТЫ, КОДИРУЮЩИЕ ПОВТОРЯЮЩИЕСЯ
АМИНОКИСЛОТНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ С ВЫСОКИМ СОДЕРЖАНИЕМ
ПРОЛИНОВЫХ И АЛАНИНОВЫХ ОСТАТКОВ, ИМЕЮЩИЕ НУКЛЕОТИДНЫЕ
ПОСЛЕДОВАТЕЛЬНОСТИ С НИЗКОЙ ПОВТОРЯЕМОСТЬЮ**

(31) **15202093.9**(32) **2015.12.22**(33) **EP**(43) **2019.01.31**(86) **PCT/EP2016/082407**(87) **WO 2017/109087 2017.06.29**(71)(73) Заявитель и патентовладелец:
**ИксЭль-ПРОТЕИН ГМБХ;
ТЕХНИШЕ УНИВЕРСИТЕТ
МЮНХЕН (DE)**(72) Изобретатель:
**Биндер Ули, Ахац Стефан, Скерра
Арне (DE)**(74) Представитель:
**Поликарпов А.В., Соколова М.В.,
Путинцев А.И., Черкас Д.А., Игнатъев
А.В. (RU)**(56) **WO-A1-2011144756**

M. SCHLAPSCHY ET AL.: "PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active

proteins", PROTEIN ENGINEERING DESIGN AND SELECTION, vol. 26, no. 8, 1 August 2013 (2013-08-01), pages 489-501, XP055195431, ISSN: 1741-0126, DOI: 10.1093/protein/gzt023, cited in the application, abstract page 490, 491

VOLKER MORATH ET AL.: "PASylation of Murine Leptin Leads to Extended Plasma Half-Life and Enhanced in Vivo Efficacy", MOLECULAR PHARMACEUTICS, vol. 12, no. 5, 4 May 2015 (2015-05-04), pages 1431-1442, XP055274064, US ISSN: 1543-8384, DOI: 10.1021/mp5007147, abstract, page 1432

DANIEL HARARI ET AL.: "Enhanced in Vivo Efficacy of a Type I Interferon Superagonist with Extended Plasma Half-life in a Mouse Model of Multiple Sclerosis", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 289, no. 42, 17 October 2014 (2014-10-17), pages 29014-29029, XP055274066, US ISSN: 0021-9258, DOI: 10.1074/jbc.M114.602474, abstract, page 29015

CLAUDIA T. MENDLER ET AL.: "High contrast tumor imaging with radio-labeled antibody Fab fragments tailored for optimized pharmacokinetics via PASylation", MABS, vol. 7, no. 1, 2 January 2015 (2015-01-02), pages 96-109, XP055274067, US ISSN: 1942-0862, DOI: 10.4161/19420862.2014.985522, abstract, page 106

(57) Изобретение относится к молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность с низкой повторяемостью, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина. Кодированный полипептид содержит повторяющуюся аминокислотную последовательность, образующую случайный клубок. Молекула нуклеиновой кислоты, содержащая указанные нуклеотидные последовательности с низкой повторяемостью, может дополнительно содержать нуклеотидную последовательность, кодирующую биологически или фармакологически активный белок. Кроме того, согласно настоящему изобретению предложены средства и способы отбора, позволяющие идентифицировать указанную молекулу нуклеиновой кислоты, содержащую указанную нуклеотидную последовательность с низкой повторяемостью. Настоящее изобретение также относится к способу получения указанных молекул нуклеиновых кислот. Кроме того, здесь предложены способы получения кодируемого полипептида или конъюгатов лекарственных средств с кодируемым полипептидом с использованием предложенных здесь молекул нуклеиновых кислот. Конъюгат лекарственного средства может содержать биологически или фармакологически активный белок или низкомолекулярное лекарственное средство. Кроме того, здесь предложены векторы и хозяева, содержащие такие молекулы нуклеиновых кислот.

B1**041113****041113****B1**

Настоящее изобретение относится к молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность с низкой повторяемостью, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина. Кодированный полипептид содержит повторяющуюся аминокислотную последовательность, образующую случайный клубок. Молекула нуклеиновой кислоты, содержащая указанные нуклеотидные последовательности с низкой повторяемостью, может дополнительно содержать нуклеотидную последовательность, кодирующую биологически или фармакологически активный белок. Кроме того, согласно настоящему изобретению предложены средства и способы отбора, позволяющие идентифицировать указанную молекулу нуклеиновой кислоты, содержащую указанную нуклеотидную последовательность с низкой повторяемостью. Настоящее изобретение также относится к способу получения указанной(ых) молекул(ы) нуклеиновой кислоты. Кроме того, здесь предложены способы получения кодируемого полипептида или конъюгата(ов) лекарственного средства с кодируемым полипептидом с использованием предложенных здесь молекул нуклеиновой кислоты. Конъюгат лекарственного средства может содержать биологически или фармакологически активный белок или низкомолекулярное лекарственное средство. Кроме того, здесь предложены векторы и хозяева, содержащие такие молекулы нуклеиновой кислоты.

Полипептиды, образующие случайные клубки, известны в уровне техники. Например, в WO 2008/155134 раскрыты белки, содержащие аминокислотную последовательность из по меньшей мере примерно 100 аминокислотных остатков, состоящие из пролиновых, аланиновых и сериновых (PAS) остатков. Аминокислотная последовательность, образующая случайный клубок, может содержать множество аминокислотных повторов. Эти повторы могут состоять из по меньшей мере 3-30 или более аминокислотных остатков. В WO 2011/144756 раскрыты полипептиды, содержащие повторяющиеся аминокислотные последовательности, состоящие исключительно из пролиновых и аланиновых (PA) остатков. Эти полипептиды также образуют случайные клубки и состоят из по меньшей мере 50 пролиновых и аланиновых остатков. В WO 2015/132004 раскрыт рекомбинантный клостридиальный нейротоксин, содержащий домен в виде случайного клубка, состоящего из PAS. В US 2006/0252120 A1 раскрыты гликопротеины с высоким содержанием гидроксипролина, содержащие сегменты, кодируемые как гликомодули с высоким содержанием пролина и мотивом аминокислотной последовательности $[(AP)_5]_n$. Кроме того, встречающиеся в природе полипептиды содержат последовательности с высоким содержанием пролина и аланина, такие как сверхкрупный оболочечный белок Macacine herpesvirus 1, ген которого опубликован в банке генов под номером доступа AAR41454.1. Способы оптимизации кодонов раскрыты в WO 2007/142954.

Повторяющиеся полипептиды из уровня техники, такие как PAS или PA последовательности, обычно кодируются соответствующими повторяющимися нуклеиновыми кислотами. Соответственно, нуклеиновые кислоты из уровня техники также отражают повторяющуюся структуру кодируемых ими аминокислотных последовательностей в своей нуклеотидной последовательности. Таким образом, нуклеиновые кислоты из уровня техники имеют высокую повторяемость на уровне своих последовательностей. Повторяемость нуклеиновых кислот из уровня техники может приводить к таким недостаткам, как частичная генетическая нестабильность, в частности когда они кодируют длинные PAS или PA последовательности, например 300 остатков или более.

Таким образом, техническая задача настоящего изобретения заключается в обеспечении средств и способов удобного и надежного получения полипептидов, содержащих аминокислотные повторы, в частности полипептидов, содержащих аминокислотные повторы, состоящие из пролина, аланина и возможно серина.

Техническая задача решена путем предложения воплощений, представленных ниже, и как описано в прилагаемой формуле изобретения.

Настоящее изобретение относится к нижеследующему.

1. Молекула нуклеиновой кислоты, содержащая нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина,

где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов,

где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 50000,

где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}},$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности,

n представляет собой длину повтора в указанной нуклеотидной последовательности и

$f_i(n)$ представляет собой частоту указанного повтора длиной n ,

где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора

длиной n равен 1.

2. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина и аланина.

3. Молекула нуклеиновой кислоты по п.2, где указанные пролиновые остатки составляют более чем примерно 10% и менее чем примерно 75% указанного кодируемого полипептида.

4. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина, аланина и серина.

5. Молекула нуклеиновой кислоты по п.4, где указанные пролиновые остатки составляют более 4% и менее 40% указанного кодируемого полипептида.

6. Молекула нуклеиновой кислоты по любому из пп.1-5, где указанный показатель нуклеотидных повторов (NRS) составляет менее 100.

7. Молекула нуклеиновой кислоты по любому из пп.1-6, имеющая повышенную генетическую стабильность.

8. Молекула нуклеиновой кислоты по любому из пп.1-7, где указанная нуклеотидная последовательность содержит указанные повторы, которые имеют максимальную длину n_{\max} , определенную по следующей формуле:

$$n_{\max} \leq 17 + \frac{N_{\text{tot}}}{600},$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности.

9. Молекула нуклеиновой кислоты по любому из пп.1-8, где указанные повторы имеют максимальную длину от примерно 14, 15, 16 или 17 нуклеотидов до примерно 55 нуклеотидов.

10. Молекула нуклеиновой кислоты по любому из пп.1-9, где указанные повторы имеют максимальную длину, соответствующую 50% длины указанной нуклеотидной последовательности.

11. Молекула нуклеиновой кислоты по любому из пп.1-10, где указанный кодируемый полипептид содержит повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 9 расположенных друг за другом аминокислотных остатков и где указанный полипептид образует случайный клубок.

12. Молекула нуклеиновой кислоты по любому из пп.1-11, выбранная из группы, состоящей из:

(а) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 и SEQ ID NO: 27;

(б) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36 и SEQ ID NO: 37;

(с) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 и SEQ ID NO: 41;

(д) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, состоящую из SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44 и/или SEQ ID NO: 45;

(е) молекулы нуклеиновой кислоты, гибридизующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (а) или (б);

(ф) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 56% идентичную нуклеотидной последовательности, определенной в любом из (а), (с) и (е);

(г) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 66,7% идентичную нуклеотидной последовательности, определенной в любом из (б), (д) и (е); и

(h) молекулы нуклеиновой кислоты, являющейся вырожденной в силу генетического кода по отношению к нуклеотидной последовательности, определенной в любом из (а)-(г).

13. Молекула нуклеиновой кислоты по любому из пп.1-12, содержащая два комплементарных "липких" 5'-конца, где "липкий" 5'-конец кодирующей цепи представляет собой 5'-GCC и где "липкий" 5'-конец некодирующей цепи представляет собой 5'-GGC.

14. Молекула нуклеиновой кислоты по любому из пп.1-13, функционально связанная в одной и той же рамке считывания с нуклеиновой кислотой, кодирующей биологически активный белок.

15. Молекула нуклеиновой кислоты по п.14, где указанный биологически активный белок представляет собой терапевтически эффективный белок.

16. Молекула нуклеиновой кислоты по п.14 или 15, где указанный биологически активный белок выбран из группы, состоящей из связывающего белка, фрагмента антитела, цитокина, фактора роста, гормона, фермента, белковой вакцины, пептидной вакцины, пептида, состоящего из 50 или менее аминокислотных остатков, или пептидомиметика.

17. Молекула нуклеиновой кислоты по п.16, где указанный связывающий белок выбран из группы, состоящей из антител, Fab-фрагментов, Fab'-фрагментов, F(ab')₂-фрагментов, одноцепочечных вариабельных фрагментов (scFv), (одно)доменных антител, выделенных вариабельных областей антител (областей VL и/или VH), CDR (гипервариабельные участки), иммуноглобулиновых доменов, пептидомиметиков, имеющих происхождение от CDR, лектинов, белковых каркасов, фибронектиновых доменов, те-

наскиновых доменов, доменов белка А, доменов SH3, доменов анкириновых повторов и липокалинов.

18. Молекула нуклеиновой кислоты по любому из пп.14-17, где указанный биологически активный белок выбран из группы, состоящей из антагониста рецептора интерлейкина-1, лептина, кислой сфингомиелиназы, аденозиндезаминазы, агалсидазы-альфа, альфа-1-антитрипсина, предсердного натрийуретического пептида альфа, альфа-галактозидазы, альфа-глюкозидазы, альфа-N-ацетилглюкозаминидазы, алтеплазы, амедилазы, амилина, аналога амилина, пептидного ингибитора слияния HIV (вирус иммунодефицита человека), аргининдезаминазы, аспарагиназы, фактора VIII с делецией домена В, морфогенетического костного белка, антагониста брадикинина, натрийуретического пептида В-типа, буганина, гормона роста, хорионического гонадотропина, антагониста рецептора CD3, антагониста CD19, антагониста CD20, антагониста CD40, антагониста CD40L (лиганд CD40), цереброзидсульфатазы, фактора свертывания VIIa, фактора свертывания XIII, фактора свертывания IX, фактора свертывания X, ингибитора компонента комплемента С3, антагониста компонента комплемента 5a, С-пептида, антагониста STLA-4 (цитотоксический Т-лимфоцитарный антиген 4), натрийуретического пептида С-типа, дефензина, дезоксирибонуклеазы I, антагониста рецептора EGFR (рецептор эпидермального фактора роста), эпидермального фактора роста, эритропоэтина, эксендина-4, эстринового пептида 1, антагониста рецептора FcγIIb, фактора роста фибробластов 21, фолликулстимулирующего гормона, желудочного ингибирующего полипептида (GIP), аналога GIP, глюкагона, агониста рецепторов глюкагона, глюкагоноподобного пептида 1 (GLP-1), аналога GLP-1, глюкагоноподобного пептида 2 (GLP-2), аналога GLP-2, гонадорелина, агониста гонадотропин-рилизинг гормона, антагониста гонадотропин-рилизинг гормона, gp120 (гликопротеин 120), gp160 (гликопротеин 160), гранулоцитарного колониестимулирующего фактора (G-CSF), гранулоцитарно-макрофагального колониестимулирующего фактора (GM-CSF), грелина, аналога грелина, гормона роста, рилизинг-гормона гормона роста, гематида, фактора роста гепатоцитов, антагониста рецептора фактора роста гепатоцитов (HGFR), антагониста гепсидина, миметика гепсидина, антагониста рецептора Her2/neu (рецептор эпидермального фактора роста человека 2), гистрелина, гирудина, антагониста hsp70 (белок теплового шока 70), гуманина, гиалуронидазы, гидролитического лизосомального глюкоцереброзид-специфичного фермента, идуронат-2-сульфатазы, антагонистов IgE (иммуноглобулин E), инсулина, аналога инсулина, инсулиноподобного фактора роста 1, инсулиноподобного фактора роста 2, интерферона-альфа, антагониста интерферона-альфа, супергониста интерферона-альфа, интерферона-альфа-п3, интерферона-бета, интерферона-гамма, интерферона-лямбда, интерферона-тау, интерлейкина, слитого белка интерлейкина-2, антагониста альфа-субъединицы рецептора интерлейкина-22 (IL-22ra), ирисина, белка, ассоциированного с островковым неогенезом, фактора роста кератиноцитов, антагонистов ионных каналов Kv1.3, лантипептида, липазы, лютеинизирующего гормона, лутропина альфа, лизостафина, маннозидазы, N-ацетилгалактозамин-6-сульфатазы, N-ацетилглюкозаминидазы, нейтрофильного желатиназоассоциированного липокалина, октреотида, ω-конотоксина, ингибитора комплемента *Opnithodoros moukata*, остеогенного белка 1, остеопротегерина, оксалатдекарбоксилазы, P128, паратиреоидного гормона, филомера, антагониста PD-1 (белок программируемой клеточной гибели-1), антагониста PDGF (фактор роста тромбоцитов), фенилаланинаммиаклиаза, тромбоцитарного фактора роста, проинсулина, белка С, релаксина, аналога релаксина, секретина, RGD-пептида (трипептид аргинин-глицин-аспарагиновая кислота), рибонуклеазы, сенреботазы (senrebotase), ингибитора сериновых протеаз, растворимого рецептора комплемента 1, растворимого рецептора DCC (рецептор, обнаруживаемый при колоректальном раке), растворимого рецептора TAC1 (трансмембранный активатор и партнер кальциевого модулятора и лиганда циклофилина), растворимого рецептора фактора некроза опухоли I (sTNF-RI), растворимого рецептора фактора некроза опухоли II (sTNF-RII), растворимого рецептора VEGF (фактор роста эндотелия сосудов) Flt-1, растворимого рецептора FcγIIb, соматостатина, аналога соматостатина, стрептокиназы, лиганда Т-клеточного рецептора, тенектеплазы, терипаратида, тромбомодулина-альфа, тимозина-альфа-1, ингибитора Toll-подобных рецепторов, фактора некроза опухоли (TNFα), антагониста фактора некроза опухоли α, уриказы, вазоактивного кишечного пептида, вазопрессина, аналога вазопрессина, антагониста VEGF, фактора фон Виллебранда.

19. Вектор, содержащий молекулу нуклеиновой кислоты по любому из пп.1-18.

20. Вектор по п.19, содержащий верхнюю последовательность распознавания, распознаваемую ферментом эндонуклеазой рестрикции, и нижнюю последовательность распознавания, распознаваемую ферментом эндонуклеазой рестрикции, где указанная верхняя последовательность распознавания и указанная нижняя последовательность распознавания расположены в обратно-комплементарной ориентации.

21. Вектор по п.20, где указанный фермент эндонуклеаза рестрикции, распознающий указанную нижнюю последовательность распознавания, отличается от фермента эндонуклеазы рестрикции, распознающего верхнюю последовательность распознавания.

22. Вектор по п.20 или 21, где указанная верхняя последовательность распознавания содержит две последовательности распознавания для двух разных рестриктаз.

23. Вектор по любому из пп.20-22, где указанная нижняя последовательность распознавания содержится в верхней последовательности распознавания.

24. Вектор по любому из пп.20-23, где указанная верхняя последовательность распознавания и/или указанная нижняя последовательность распознавания представляют собой последовательности распознавания для рестриктаз, приводящих к образованию "липких" нуклеотидных концов.

25. Вектор по любому из пп.20-24, где указанная верхняя последовательность распознавания и/или указанная нижняя последовательность распознавания представляют собой сайты распознавания для рестриктаз, расщепляющих за пределами последовательности распознавания.

26. Вектор по любому из пп.20-25, где указанная рестриктаза представляет собой рестриктазу типа III.

27. Вектор по любому из пп.20-26, где указанная верхняя последовательность распознавания имеет нуклеотидную последовательность "5'-GCTCTTC-3'" и/или где указанная нижняя последовательность распознавания имеет нуклеотидную последовательность "5'-CTCTTC-3'".

28. Вектор по любому из пп.0-27, где указанную верхнюю последовательность распознавания распознают SapI и EarI и/или где указанную нижнюю последовательность распознавания распознает EarI.

29. Вектор по любому из пп.20-28, имеющий последовательность, представленную в SEQ ID NO: 48 или SEQ ID NO: 55.

30. Хозяин, содержащий молекулу нуклеиновой кислоты по любому из пп.1-18 или трансформированный вектором по любому из пп.19-29.

31. Хозяин по п.30, выбранный из группы, состоящей из бактерии, клетки млекопитающего, клетки насекомого, клетки водоросли, ресничной клетки, клетки дрожжей и растительной клетки.

32. Хозяин по п.30 или 31, где бактерия принадлежит к роду *Escherichia*, *Corynebacterium*, *Pseudomonas* или *Bacillus*.

33. Хозяин по п.32, где указанная бактерия представляет собой *Escherichia coli*, *Corynebacterium glutamicum*, *Pseudomonas fluorescens* или *Bacillus megaterium*.

34. Хозяин по п.31, где указанная клетка млекопитающего представляет собой клетку хомяка.

35. Хозяин по п.34, где указанная клетка млекопитающего представляет собой клетку СНО (яичника китайского хомячка).

36. Хозяин по п.31, где указанные дрожжи принадлежат к роду *Saccharomyces*, *Pichia*, *Hansenula* или *Kluuyveromyces*.

37. Хозяин по п.36, где указанные дрожжи представляют собой *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*, *Hansenula polymorpha* или *Kluuyveromyces lactis*.

38. Способ получения указанной молекулы нуклеиновой кислоты по любому из пп.1-18 или указанного вектора по любому из пп.19-29, включающий культивирование/выращивание хозяина по любому из пп.30-37 и возможно выделение полученной молекулы нуклеиновой кислоты и/или вектора.

39. Способ получения полипептида, кодируемого молекулой нуклеиновой кислоты по любому из пп.1-18, включающий культивирование/выращивание хозяина по любому из пп.30-37, содержащего молекулу нуклеиновой кислоты по любому из пп.1-18, и возможно выделение полученного полипептида.

40. Способ получения конъюгата лекарственного средства, содержащего полипептид, кодируемый молекулой нуклеиновой кислоты по любому из пп.1-18, и дополнительно содержащего (1) биологически активный белок, и/или (2) низкомолекулярное соединение, и/или (3) углевод.

41. Способ получения конъюгата лекарственного средства по п.40, где указанная молекула нуклеиновой кислоты имеет повышенную генетическую стабильность.

42. Способ получения конъюгата лекарственного средства по п.40 или 41, где указанный биологически активный белок представляет собой терапевтически эффективный белок.

43. Способ получения конъюгата лекарственного средства по любому из пп.40-42, где указанный полипептид опосредует повышенную стабильность указанного конъюгата лекарственного средства *in vivo* и/или *in vitro*.

44. Способ получения конъюгата лекарственного средства по любому из пп.40-43, где указанный биологически активный белок выбран из группы, состоящей из связывающего белка, фрагмента антитела, цитокина, фактора роста, гормона, фермента, белковой вакцины, пептидной вакцины, пептида, состоящего из 50 или менее аминокислотных остатков, или пептидомиметика.

45. Способ получения конъюгата лекарственного средства по любому из пп.40-44, где указанный связывающий белок выбран из группы, состоящей из антител, Fab-фрагментов, Fab'-фрагментов, F(ab')₂-фрагментов, одноцепочечных переменных фрагментов (scFv), (одно)доменных антител, выделенных переменных областей антител (областей VL и/или VH), CDR, иммуноглобулиновых доменов, пептидомиметиков, имеющих происхождение от CDR, лектинов, белковых каркасов, фибронектиновых доменов, тенаксиновых доменов, доменов белка А, доменов SH3, доменов анкириновых повторов и липокалинов.

46. Способ получения конъюгата лекарственного средства по любому из пп.40-45, где указанный биологически активный белок выбран из группы, состоящей из антагониста рецептора интерлейкина-1, лептина, кислой сфингомиелиназы, аденозиндезаминазы, агалсидазы-альфа, альфа-1-антитрипсина, предсердного натрийуретического пептида альфа, альфа-галактозидазы, альфа-глюкозидазы, альфа-N-ацетилглюкозаминидазы, алтеплазы, амеидплазы, амилина, аналога амилина, пептидного ингибитора слияния HIV, аргининдезаминазы, аспарагиназы, фактора VIII с делецией домена В, костного морфоген-

нетического белка, антагониста брадикинина, натрийуретического пептида В-типа, буганина, гормона роста, хорионического гонадотропина, антагониста рецептора CD3, антагониста CD19, антагониста CD20, антагониста CD40, антагониста CD40L, церебродисульфатазы, фактора свертывания VIIa, фактора свертывания XIII, фактора свертывания IX, фактора свертывания X, ингибитора компонента комплемента C3, антагониста компонента комплемента 5a, С-пептида, антагониста CTLA-4, натрийуретического пептида С-типа, дефензина, дезоксирибонуклеазы I, антагониста рецептора EGFR, эпидермального фактора роста, эритропоэтина, эксендина-4, эзринового пептида 1, антагониста рецептора FcγIIb, фактора роста фибробластов 21, фолликул-стимулирующего гормона, желудочного ингибирующего полипептида (GIP), аналога GIP, глюкагона, агониста рецепторов глюкагона, глюкагоноподобного пептида 1 (GLP-1), аналога GLP-1, глюкагоноподобного пептида 2 (GLP-2), аналога GLP-2, гонадорелина, агониста гонадотропин-рилизинг гормона, антагониста гонадотропин-рилизинг гормона, gp120, gp160, гранулоцитарного колониестимулирующего фактора (G-CSF), гранулоцитарно-макрофагального колониестимулирующего фактора (GM-CSF), грелина, аналога грелина, гормона роста, рилизинг-гормона гормона роста, гематида, фактора роста гепатоцитов, антагониста рецептора фактора роста гепатоцитов (HGFR), антагониста гепсидина, миметика гепсидина, антагониста рецептора Her2/neu, гистрелина, гирудина, антагониста hsp70, гуманина, гиалуронидазы, гидролитического лизосомального глюкоцереброзид-специфического фермента, идуронат-2-сульфатазы, антагонистов IgE, инсулина, аналога инсулина, инсулиноподобного фактора роста 1, инсулиноподобного фактора роста 2, интерферона-альфа, антагониста интерферона-альфа, суперагониста интерферона-альфа, интерферона-альфа-n3, интерферона-бета, интерферона-гамма, интерферона-лямбда, интерферона-тау, интерлейкина, слитого белка интерлейкина-2, антагониста альфа-субъединицы рецептора интерлейкина-22 (IL-22ra), ирисина, белка, ассоциированного с островковым неогенезом, фактора роста кератиноцитов, антагонистов ионных каналов Kv1.3, лантисептида, липазы, лютеинизирующего гормона, лутропина альфа, лизостафина, маннозидазы, N-ацетилгалактозамин-6-сульфатазы, N-ацетилглюкозаминидазы, нейтрофильного желатиназоассоциированного липокалина, октреотида, ω-коноксина, ингибитора комплемента *Ornithodoros moubata*, остеогенного белка 1, остеопротегерина, оксалатдекарбоксилазы, P128, паратиреоидного гормона, филомера, антагониста PD-1, антагониста PDGF, фенилаланинаммиаклиаза, тромбоцитарного фактора роста, проинсулина, белка С, релаксина, аналога релаксина, секретина, RGD-пептида, рибонуклеазы, сенреботазы, ингибитора сериновых протеаз, растворимого рецептора комплемента 1, растворимого рецептора DCC, растворимого рецептора TAC1, растворимого рецептора фактора некроза опухоли I (sTNF-RI), растворимого рецептора фактора некроза опухоли II (sTNF-RII), растворимого рецептора VEGF Flt-1, растворимого рецептора FcγIIb, соматостатина, аналога соматостатина, стрептокиназы, лиганда Т-клеточного рецептора, тенектеплазы, терипаратида, тромбомодулина-альфа, тимозина-альфа-1, ингибитора Toll-подобных рецепторов, фактора некроза опухоли (TNFα), антагониста фактора некроза опухоли α, уриказы, вазоактивного кишечного пептида, вазопрессина, аналога вазопрессина, антагониста VEGF, фактора фон Виллебранда.

47. Способ получения конъюгата лекарственного средства по любому из пп.40-46, где указанное низкомолекулярное соединение выбрано из группы, состоящей из ингибиторов ангиогенеза, противоаллергических лекарственных средств, противорвотных лекарственных средств, антидепрессантов, антигипертензивных лекарственных средств, противовоспалительных лекарственных средств, противоинфекционных лекарственных средств, антипсихотических лекарственных средств, антипролиферативных (цитотоксических и цитостатических) лекарственных средств, антагонистов кальция и других лекарственных средств, действующих на органы кровообращения, холинергических агонистов, лекарственных средств, действующих на центральную нервную систему, лекарственных средств, действующих на дыхательную систему, гормонов, стероидов, поликетидов, углеводов, олигосахаридов, нуклеиновых кислот, производных нуклеиновых кислот, бессмысловых нуклеиновых кислот, малых интерферирующих РНК (siRNA), ингибиторов микро-РНК (miR), миметиков микро-РНК, ДНК-аптамеров и РНК-аптамеров.

48. Полипептид, полученный или получаемый способом по п.39, конъюгат лекарственного средства, полученный способом по любому из пп.40-47.

49. Способ секвенирования молекулы нуклеиновой кислоты по любому из пп.1-18.

50. Способ амплификации молекулы нуклеиновой кислоты по любому из пп.1-18.

51. Способ клонирования молекулы нуклеиновой кислоты по любому из пп.1-18.

52. Способ отбора генетически стабильной молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, где указанная нуклеотидная последовательность имеет длину по меньшей мере 300 нуклеотидов, включающий стадию отбора молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую показатель нуклеотидных повторов (NRS) менее 50000, где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности,

n представляет собой длину повтора в указанной нуклеотидной последовательности и $f_i(n)$ представляет собой частоту указанного повтора длиной n , где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1.

В определенных аспектах настоящее изобретение относится к нижеследующему.

1. Молекула нуклеиновой кислоты, содержащая нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина,

где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов,

где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 50000,

где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности,

n представляет собой длину повтора в указанной нуклеотидной последовательности и

$f_i(n)$ представляет собой частоту указанного повтора длиной n ,

где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1.

2. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина и аланина, где указанные пролиновые остатки составляют более чем примерно 10% и менее чем примерно 75% указанного кодируемого полипептида.

3. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина, аланина и серина, где указанные пролиновые остатки составляют более 4% и менее 40% указанного кодируемого полипептида.

4. Молекула нуклеиновой кислоты по любому из пп.1-3, где указанный показатель нуклеотидных повторов (NRS) составляет менее 100.

5. Молекула нуклеиновой кислоты по любому из пп.1-4, имеющая повышенную генетическую стабильность.

6. Молекула нуклеиновой кислоты по любому из пп.1-5, где указанная нуклеотидная последовательность содержит указанные повторы, имеющие максимальную длину n_{max} , определенную по следующей формуле:

$$n_{max} \leq 17 + \frac{N_{tot}}{600},$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности.

7. Молекула нуклеиновой кислоты по любому из пп.1-6, где указанные повторы имеют максимальную длину от примерно 14, 15, 16 или 17 нуклеотидов до примерно 55 нуклеотидов.

8. Молекула нуклеиновой кислоты по любому из пп.1-7, где указанный кодируемый полипептид содержит повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 9 расположенных друг за другом аминокислотных остатков и где указанный полипептид образует случайный клубок.

9. Молекула нуклеиновой кислоты по любому из пп.1-8, выбранная из группы, состоящей из:

(a) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 и SEQ ID NO: 27;

(b) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36 и SEQ ID NO: 37;

(c) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40 и SEQ ID NO: 41;

(d) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, состоящую из SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44 и/или SEQ ID NO: 45;

(e) молекулы нуклеиновой кислоты, гибридизирующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (a) или (b);

(f) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 56% идентичную нуклеотидной последовательности, определенной в любом из (a), (c) и (e);

(g) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 66,7% идентичную нуклеотидной последовательности, определенной в любом из (b), (d) и (e); и

(h) молекулы нуклеиновой кислоты, являющейся вырожденной в силу генетического кода по отношению к нуклеотидной последовательности, определенной в любом из (a)-(d).

10. Молекула нуклеиновой кислоты по любому из пп.1-9, функционально связанная в одной и той же рамке считывания с нуклеиновой кислотой, кодирующей биологически активный белок.

11. Молекула нуклеиновой кислоты по п.10, где указанный биологически активный белок выбран из группы, состоящей из связывающего белка, фрагмента антитела, цитокина, фактора роста, гормона, фермента, белковой вакцины, пептидной вакцины, пептида, состоящего из 50 или менее аминокислотных остатков, или пептидомиметика, где указанный связывающий белок выбран из группы, состоящей из антител, Fab-фрагментов, Fab'-фрагментов, F(ab')₂-фрагментов, одноцепочечных переменных фрагментов (scFv), (одно)доменных антител, выделенных переменных областей антител (областей VL и/или VH), CDR, иммуноглобулиновых доменов, пептидомиметиков, имеющих происхождение от CDR, лектинов, белковых каркасов, фибронектиновых доменов, тенасциновых доменов, доменов белка А, доменов SH3, доменов анкириновых повторов и липокалина.

12. Молекула нуклеиновой кислоты по п.10 или 11, где указанный биологически активный белок выбран из группы, состоящей из кислой сфингомиелиназы, аденозиндезаминазы, агалсидазы-альфа, альфа-1-антитрипсина, предсердного натрийуретического пептида альфа, альфа-галактозидазы, альфа-глюкозидазы, альфа-N-ацетилглюкозаминидазы, алтеплазы, амедиплазы, амилина, аналога амилина, пептидного ингибитора слияния HIV, аргининдезаминазы, аспарагиназы, фактора VIII с делецией домена В, морфогенетического костного белка, антагониста брадикинина, натрийуретического пептида В-типа, буганина, гормона роста, хорионического гонадотропина, антагониста рецептора CD3, антагониста CD19, антагониста CD20, антагониста CD40, антагониста CD40L, цереброзидсульфатазы, фактора свертывания VIIa, фактора свертывания XIII, фактора свертывания IX, фактора свертывания X, ингибитора компонента комплемента C3, антагониста компонента комплемента 5a, С-пептида, антагониста CTLA-4, натрийуретического пептида С-типа, дефензина, дезоксирибонуклеазы I, антагониста рецептора EGFR, эпидермального фактора роста, эритропоэтина, эксендина-4, эриного пептида 1, антагониста рецептора FcγIIb, фактора роста фибробластов 21, фолликулстимулирующего гормона, желудочного ингибирующего полипептида (GIP), аналога GIP, глюкагона, агониста рецепторов глюкагона, глюкагоноподобного пептида 1 (GLP-1), аналога GLP-1, глюкагоноподобного пептида 2 (GLP-2), аналога GLP-2, гонадотропина, агониста гонадотропин-рилизинг гормона, антагониста гонадотропин-рилизинг гормона, gp120, gp160, гранулоцитарного колониестимулирующего фактора (G-CSF), гранулоцитарно-макрофагального колониестимулирующего фактора (GM-CSF), грелина, аналога грелина, гормона роста, рилизинг-гормона, гематида, фактора роста гепатоцитов, антагониста рецептора фактора роста гепатоцитов (HGFR), антагониста гепсидина, миметика гепсидина, антагониста рецептора Her2/neu, гистрелина, гирудина, антагониста hsp70, гуманина, гиалуронидазы, гидролитического лизосомального глюкоцереброзид-специфичного фермента, идуронат-2-сульфатазы, антагонистов IgE, инсулина, аналога инсулина, инсулиноподобного фактора роста 1, инсулиноподобного фактора роста 2, интерферона-альфа, антагониста интерферона-альфа, суперагониста интерферона-альфа, интерферона-альфа-n3, интерферона-бета, интерферона-гамма, интерферона-лямбда, интерферона-тау, интерлейкина, антагониста рецептора интерлейкина-1, слитого белка интерлейкина-2, антагониста альфа-субъединицы рецептора интерлейкина-22 (IL-22ra), ирисина, белка, ассоциированного с островковым неогенезом, фактора роста кератиноцитов, антагонистов ионных каналов Kv1.3, лантипептида, лептина, липазы, лютеинизирующего гормона, лутропина альфа, лизостафина, маннозидазы, N-ацетилгалактозамин-6-сульфатазы, N-ацетилглюкозаминидазы, нейтрофильного желатиназоассоциированного липокалина, октреотида, ω-конотоксина, ингибитора комплемента *Ornithodoros moubata*, остеогенного белка 1, остеопротегерина, оксалатдекарбоксилазы, P128, паратиреоидного гормона, филомера, антагониста PD-1, антагониста PDGF, фенилаланинаммиаклиазы, тромбоцитарного фактора роста, проинсулина, белка С, релаксина, аналога релаксина, секретина, RGD-пептида, рибонуклеазы, сенреботазы, ингибитора сериновых протеаз, растворимого рецептора комплемента 1, растворимого рецептора DCC, растворимого рецептора TAC1, растворимого рецептора фактора некроза опухоли I (sTNF-RI), растворимого рецептора фактора некроза опухоли II (sTNF-RII), растворимого рецептора VEGF Flt-1, растворимого рецептора FcγIIb, соматостатина, аналога соматостатина, стрептокиназы, лиганда Т-клеточного рецептора, тенектеплазы, терипаратида, тромбомодулина-альфа, тимозина-альфа-1, ингибитора Toll-подобных рецепторов, фактора некроза опухоли (TNFα), антагониста фактора некроза опухоли α, уриказы, вазоактивного кишечного пептида, вазопрессина, аналога вазопрессина, антагониста VEGF, фактора фон Виллебранда.

13. Хозяин, содержащий молекулу нуклеиновой кислоты по любому из пп.1-12 или трансформированный вектором, содержащим молекулу нуклеиновой кислоты по любому из пп.1-12.

14. Способ получения полипептида, кодируемого молекулой нуклеиновой кислоты по любому из пп.1-12, включающий культивирование/выращивание хозяина по п.13 и возможно выделение полученного полипептида.

15. Способ получения конъюгата лекарственного средства, содержащего полипептид, кодируемый молекулой нуклеиновой кислоты по любому из пп.1-12, и дополнительно содержащего (1) биологически

активный белок, и/или (2) низкомолекулярное соединение, и/или (3) углевод.

Настоящее изобретение относится к молекулам нуклеиновой кислоты с нуклеотидными последовательностями с низкой повторяемостью, кодирующими полипептиды, состоящие из пролина, аланина и возможно серина. Такие полипептиды здесь также называют полипептидами с высоким содержанием РА или с высоким содержанием пролина/аланина. Молекулы нуклеиновой кислоты по изобретению, предложенные здесь, имеют мало или не имеют нуклеотидных повторов определенной максимальной длины, такой как максимальная длина примерно 14, 15, 16 или 17 нуклеотидов на последовательность нуклеотидного повтора. Кроме того, нуклеотидная последовательность, кодирующая полипептид с высоким содержанием РА, имеет общую длину по меньшей мере 300 нуклеотидов, и отдельные нуклеотидные повторы в этой кодирующей последовательности имеют максимальную длину 14, 15, 16, 17, примерно 20, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45, примерно 50 или примерно 55 нуклеотидов.

В одном аспекте молекула нуклеиновой кислоты по изобретению содержит нуклеотидную последовательность, кодирующую повторяющуюся аминокислотную последовательность (например, полипептид с высоким содержанием РА), где указанная нуклеотидная последовательность указанной молекулы нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов, где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 50000 и где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

В данной формуле N_{tot} представляет собой длину указанной нуклеотидной последовательности, n представляет собой длину повтора в указанной нуклеотидной последовательности и $f_i(n)$ представляет собой частоту указанного повтора длиной n , и где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1. Определение приведено ниже.

В определенном аспекте молекула нуклеиновой кислоты по изобретению кодирует полипептид, состоящий из пролина, аланина и возможно серина, где участок нуклеотидной последовательности, повторяющийся в нуклеотидной последовательности указанной молекулы нуклеиновой кислоты (то есть "повтор") имеет максимальную длину 14, 15, 16, 17, примерно 20, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45, примерно 50 или примерно 55 нуклеотидов. Иными словами, молекула нуклеиновой кислоты содержит нуклеотидную последовательность, кодирующую полипептид с высоким содержанием РА, где указанная кодирующая нуклеотидная последовательность содержит нуклеотидные повторы, имеющие максимальную длину 14, 15, 16, 17, примерно 20, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45, примерно 50 или примерно 55 нуклеотидов. Молекулы нуклеиновой кислоты/последовательности по изобретению могут также содержать дополнительные кодирующие последовательности, такие как, среди прочего, биологически или фармакологически активные белки.

В прилагаемых примерах продемонстрировано, что молекулы нуклеиновой кислоты с низкой повторяемостью по изобретению предпочтительны по сравнению с молекулами нуклеиновой кислоты с высокой повторяемостью по предшествующему уровню техники. В частности, генетическая стабильность предложенных здесь молекул нуклеиновой кислоты с низкой повторяемостью выше, что подтверждено здесь и показано в прилагаемых примерах. Для оценки генетической стабильности повторяющихся молекул нуклеиновой кислоты по предшествующему уровню техники, кодирующих полипептиды с высоким содержанием РА, был сконструирован вектор, содержащий повторяющуюся молекулу нуклеиновой кислоты по предшествующему уровню техники, кодирующую повторяющуюся последовательность с высоким содержанием пролина/аланина (PAS#1a(600); SEQ ID NO: 12; пример 6, фиг. 2B), составленную из множества 60-мерных элементов нуклеотидной последовательности, как раскрыто в WO 2008/155134. Данный вектор назван здесь "pASK75-PAS#1a(600)-IL1Ra" (SEQ ID NO: 51). Хозяина (*E. coli*) трансформировали указанным вектором и культивировали в течение нескольких суток, например 7 суток. На 7 сутки, после непрерывного роста примерно в 70 поколениях, клетки высевали на агар LB/Amp, отбирали клоны и получали плазмиды. Плазмиды анализировали, используя рестриктазы с последующим электрофорезом а агарозном геле (фиг. 5). Четыре из пяти проанализированных клонов pASK75-PAS#1a(600)-IL1Ra продемонстрировали укороченные фрагменты нуклеиновой кислоты, кодирующие последовательности с высоким содержанием пролина/аланина (фиг. 5, дорожки 1-5).

Таким образом, повторяющиеся молекулы нуклеиновой кислоты по предшествующему уровню техники, кодирующие повторяющиеся последовательности с высоким содержанием пролина/аланина, генетически нестабильны. Иными словами, молекулы нуклеиновой кислоты по предшествующему уровню техники имеют низкую стабильность *in vivo*. Безотносительно теории, генетическая нестабильность может быть результатом гомологичной рекомбинации. Из-за генетической нестабильности и укорочения геновой кассеты полученная в результате аминокислотная последовательность с высоким содержанием пролина/аланина также будет изменена. Поэтому при длительном культивировании аминокислотная последовательность с высоким содержанием пролина/аланина, кодируемая нестабильными плазмидами,

будет отличаться от последовательности, кодируемой исходной плазмидой. Следовательно, есть существенный риск того, что полипептиды, полученные при длительном культивировании с использованием повторяющихся молекул нуклеиновой кислоты по предшествующему уровню техники, не будут желаемыми полипептидами.

Более того, есть риск, что полученная полипептидная композиция может содержать множество различных полипептидов (например, полипептидов с высоким содержанием пролина/аланина различных размеров, длин и/или последовательностей) в противоположность желаемому единообразию биологических продуктов, особенно для биофармацевтического применения. Таким образом, генетическая нестабильность повторяющихся молекул нуклеиновой кислоты по предшествующему уровню техники, кодирующих полипептиды с высоким содержанием пролина/аланина, может приводить к снижению качества конечного биологического продукта, делая его получение неудобным и ненадежным, особенно с точки зрения нормативных требований к терапевтическому применению.

В настоящем изобретении проблему генетической нестабильности решают посредством разработки молекул нуклеиновой кислоты с малым числом внутренних нуклеотидных повторов. Тем не менее, эта задача была непростой, с учетом малого числа доступных триплетных нуклеотидных кодонов, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина.

Как показано в прилагаемых неограничивающих примерах, молекулы нуклеиновой кислоты по изобретению позволяют избежать указанных выше недостатков; см., например, примеры 5 и 6, а также фиг. 4 и фиг. 5, дорожки 6-10. Тем не менее, как и молекулы нуклеиновой кислоты по предшествующему уровню техники, молекулы нуклеиновой кислоты по изобретению кодируют полипептиды с высоким содержанием пролина/аланина, содержащие множество повторов с высоким содержанием пролина/аланина. Однако, в отличие от молекул нуклеиновой кислоты по предшествующему уровню техники, молекулы нуклеиновой кислоты по настоящему изобретению имеют нуклеотидную последовательность с низкой повторяемостью (то есть они содержат мало нуклеотидных повторов и/или только короткие нуклеотидные повторы).

В иллюстративном примере 5 показано получение типичного вектора по изобретению, содержащего нуклеотидную последовательность с низкой повторяемостью, кодирующую полипептид с высоким содержанием пролина/аланина. Нуклеотидная последовательность с низкой повторяемостью, называемая "PAS#1f/1c/1b(600)", использованная в этом типичном векторе, показана на SEQ ID NO: 38. Полученная плазида была обозначена "pASK75-PAS#1f/1c/1b(600)-IL1Ra" (SEQ ID NO: 50) и показана на фиг. 4.

Плазмиду по изобретению "pASK75-PAS#1f/1c/1b(600)-IL1Ra" подвергали такому же культивированию, как описано выше в отношении вектора "pASK75-PAS#1a(600)-IL1Ra", содержащего повторяющуюся молекулу нуклеиновой кислоты по предшествующему уровню техники, кодирующую повторяющуюся последовательность с высоким содержанием пролина/аланина (PAS#1a(600); SEQ ID NO: 12), составленную из множества 60-мерных элементов нуклеотидной последовательности, как раскрыто в WO 2008/155134 (фиг. 1A). В отличие от "pASK75-PAS#1a(600)-IL1Ra", плазида по изобретению "pASK75-PAS#1f/1c/1b(600)-IL1Ra" продемонстрировала высокую генетическую стабильность: все проанализированные клоны "pASK75-PAS#1f/1c/1b(600)" продемонстрировали только ожидаемые полосы длиной 3093 п.о. (пары оснований) и 2377 п.о. (фиг. 5, дорожки 6-10), указывая на высокую генетическую стабильность генной кассеты с низкой повторяемостью PAS#1f/1c/1b(600), содержащей 1800 пар оснований и кодирующей последовательность с высоким содержанием пролина/аланина PAS#1. Эта высокая генетическая стабильность позволяет преодолеть недостатки молекул нуклеиновой кислоты по предшествующему уровню техники. Это ясно демонстрирует, что молекулы нуклеиновой кислоты по изобретению, содержащие нуклеотидную последовательность с низкой повторяемостью, полезны для удобного и надежного биосинтеза полипептидов с высоким содержанием пролина/аланина и/или соответствующих слитых белков.

В иллюстративном примере 4 продемонстрировано еще одно преимущество молекул нуклеиновой кислоты, предложенных согласно настоящему изобретению. В этом примере типичную молекулу нуклеиновой кислоты, содержащую нуклеотидную последовательность с низкой повторяемостью по изобретению (называемую кассетой PAS#1f/1c/1b(600); SEQ ID NO: 38; см., например, пример 1), подвергали автоматическому секвенированию ДНК. В результате была получена четкая и безошибочная электрофореграмма, включавшая более 900 пар оснований (фиг. 3), без признаков неспецифического связывания праймеров. Таким образом, в отличие от повторяющихся нуклеотидных последовательностей, которые можно надежно секвенировать только с использованием праймеров, гибридизирующихся выше или ниже соответствующей клонированной генной кассеты, длинные фрагменты ДНК с низкой повторяемостью, кодирующие последовательности с высоким содержанием пролина/аланина, можно легко секвенировать полностью. В этом случае можно также использовать праймеры с внутренним связыванием, получая таким образом, при желании, множество перекрывающихся считанных последовательностей; следует отметить, что в случае повторяющихся нуклеотидных последовательностей использование таких праймеров с внутренней гибридизацией не обеспечивает уникальных паттернов при секвенировании. Таким образом, молекулы нуклеиновой кислоты по изобретению, содержащие нуклеотидную последовательность с низкой повторяемостью, позволяют обойти эти проблемы при секвенировании.

В итоге, настоящее изобретение имеет, среди прочего, следующие преимущества по сравнению с молекулами нуклеиновой кислоты по предшествующему уровню техники, содержащими последовательности с высокой повторяемостью. Предпочтительные нуклеотидные последовательности с низкой повторяемостью по настоящему изобретению можно быстро и легко полностью секвенировать, в отличие от молекул нуклеиновой кислоты по предшествующему уровню техники. Другим преимуществом молекул нуклеиновой кислоты по настоящему изобретению являются их улучшенные свойства при амплификации, например, посредством полимеразной цепной реакции (PCR), благодаря низкой повторяемости. Кроме того, молекулы нуклеиновой кислоты по изобретению позволяют усовершенствовать методику клонирования по сравнению с нуклеотидными последовательностями, содержащими повторяемые/повторяющиеся последовательности. Особым преимуществом предложенных здесь молекул нуклеиновой кислоты является то, что они имеют повышенную генетическую стабильность по сравнению с молекулами нуклеиновой кислоты с высокой повторяемостью по предшествующему уровню техники. Это позволяет надежным образом получать полипептиды с высоким содержанием пролина/аланина и/или их слитые белки.

Характерный признак молекул нуклеиновой кислоты по настоящему изобретению состоит в том, что нуклеотидные последовательности, кодирующие полипептид, состоящий из пролина, аланина и возможно серина, представляют собой "нуклеотидные последовательности с низкой повторяемостью", что обеспечивает полезные технические эффекты, описанные выше. В прилагаемых примерах продемонстрированы способы, которые могут быть использованы для анализа того, содержит ли молекула нуклеиновой кислоты нуклеотидную последовательность с низкой повторяемостью по изобретению. В частности, в прилагаемых Примерах представлен показатель, называемый здесь "показателем нуклеотидных повторов" ("Nucleotide Repeat Score (NRS)"). Этот показатель нуклеотидных повторов (NRS) определен здесь по формуле, обсуждаемой ниже:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{r=1}^{k(n)} f_r(n)}}{N_{tot}}$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности, n представляет собой длину повтора в указанной нуклеотидной последовательности и $f_r(n)$ представляет собой частоту указанного повтора длиной n , и где при наличии более чем одного повтора длиной n , $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1. Этот показатель подробно описан ниже и наглядно продемонстрирован в прилагаемых примерах.

NRS позволяет специалисту в данной области выбрать нуклеотидные последовательности с низкой повторяемостью для использования в настоящем изобретении. Иными словами, NRS обеспечивает инструмент для определения степени повторяемости нуклеотидной последовательности. Для автоматического определения повторов и расчета NRS может быть применен алгоритм, называемый калькулятором NRS (NRS-Calculator), предложенный здесь.

Как продемонстрировано в прилагаемых иллюстративных примерах, например примере 13, несколько нуклеиновых кислот по предшествующему уровню техники, кодирующих последовательности с высоким содержанием пролина/аланина, сравнивали с нуклеиновыми кислотами с низкой повторяемостью, кодирующими последовательности с высоким содержанием пролина/аланина, по настоящему изобретению с использованием калькулятора NRS, описанного в примере 14. Например, определяли NRS следующих последовательностей по предшествующему уровню техники: PAS#1a(200), раскрытой в WO 2008/155134; PA#1a(200), раскрытой в WO 2011/144756, [(AP)₅]₂₀APA, раскрытой в US 20060252120, [AAPAPAPAP]₁₀AS, опубликованной в GenBank под номером доступа DQ399411a; крупного оболочечного белка Macacine herpesvirus 1, ген которого опубликован в GenBank под номером доступа NP851896. Кроме того, были построены гистограммы, на которых показаны частоты нуклеотидных повторов против длины соответствующего повтора, для этих нуклеотидных последовательностей по предшествующему уровню техники и для нуклеотидных последовательностей с низкой повторяемостью по настоящему изобретению, таких как PAS#1b(200) (SEQ ID NO: 19) или PA#1e/1d/1c/1b(800) (SEQ ID NO: 44) (фиг. 9). Гистограммы нуклеотидных последовательностей по предшествующему уровню техники выявили высокую повторяемость. В отличие от этого, гистограммы нуклеотидных последовательностей с низкой повторяемостью по изобретению, например PAS#1b(200) и PA#1e/1d/1c/1b(800), демонстрируют лишь небольшое число повторов с максимальной длиной 14 нуклеотидов; см., например, фиг. 9F-I.

Различия повторяемости нуклеотидных последовательностей по предшествующему уровню техники и нуклеотидных последовательностей по изобретению становятся еще более очевидными при сравнении их показателей нуклеотидных повторов. Проанализированные здесь последовательности по предшествующему уровню техники имеют NRS более 8000 (табл. 2). В отличие от этого, типичные нуклеотидные последовательности по изобретению демонстрируют низкие показатели нуклеотидных повторов, например менее 34; см. табл. 1. Следовательно, здесь ясно продемонстрировано, что применительно к повторяемости качество нуклеотидных последовательностей, кодирующих последовательности с высоким содержанием пролина/аланина, по изобретению намного выше, чем у последовательностей по пред-

шествующему уровню техники, с меньшим числом и длиной повторов нуклеотидной последовательности. Таким образом, молекулы нуклеиновой кислоты по настоящему изобретению имеют последовательности с низкой повторяемостью.

Как указано выше, специалисту в данной области известны несколько альтернатив для анализа степени повторяемости молекулы нуклеиновой кислоты. В прилагаемых примерах показано, что повторяемость молекул нуклеиновой кислоты по изобретению и по предшествующему уровню техники также анализировали по точечным диаграммам; см., например, пример 3. По точечным диаграммам анализировали нуклеиновые кислоты, кодирующие повторяющуюся последовательность с высоким содержанием пролина/аланина, PA#3a (SEQ ID NO: 15; фиг. 2A), раскрытую в WO 2011/144756, PAS#1 (SEQ ID NO: 11; фиг. 2B), раскрытую в WO 2008/155134, кодируемый мультимер $[(AP)_5]_n$ (SEQ ID NO: 16), раскрытый в US 2006/0252120, и ген повторяющейся области с высоким содержанием пролина/аланина сверхкрупного оболочечного белка Macacine herpesvirus 1, опубликованный в GenBank под номером доступа AAR41454.1 (SEQ ID NO: 18; фиг. 2C). В прилагаемых примерах был использован инструмент "dottup" пакета программного обеспечения Geneious версии 8.1 (Biomatters, Окленд, Новая Зеландия). Этот алгоритм выравнивает соответствующую последовательность с самой собой и применяет окно повтора из, например, 14 или 15 нуклеотидов. Точечные диаграммы нуклеотидных последовательностей по предшествующему уровню техники сравнивали с точечными диаграммами нуклеотидных последовательностей с низкой повторяемостью по изобретению, например элементов PA#3b(200) (SEQ ID NO: 36), PA#1b (SEQ ID NO: 28) или составных нуклеотидных последовательностей с низкой повторяемостью PAS#1f/1c/1b(600) (SEQ ID NO: 38) и PAS#1d/1f/1c/1b(800) (SEQ ID NO: 39). В то время как все проанализированные последовательности по предшествующему уровню техники продемонстрировали высокую повторяемость на уровне нуклеотидной последовательности, на что указывают черные параллельные диагональные линии (фиг. 2A-C), точечные диаграммы типичных нуклеотидных последовательностей по изобретению демонстрируют отсутствие или лишь небольшое число разрозненных повторов с максимальной длиной 14 нуклеотидов (черные линии) на протяжении всей нуклеотидной последовательности длиной 600 нуклеотидов (фиг. 2A,C), 1800 нуклеотидов или 2400 нуклеотидов, соответственно (фиг. 2B). Таким образом, здесь продемонстрировано, что предложенные здесь нуклеотидные последовательности представляют собой нуклеотидные последовательности с низкой повторяемостью, не содержащие повторов или содержащие лишь небольшое число коротких повторов.

В итоге, прилагаемые примеры ясно демонстрируют, что молекулы нуклеиновой кислоты по изобретению имеют нуклеотидные последовательности с низкой повторяемостью, кодируя в то же время повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина. Здесь также продемонстрировано, что "низкая повторяемость" молекул нуклеиновой кислоты, предложенных здесь, может легко быть оценена альтернативными способами, например по показателю нуклеотидных повторов, предложенному здесь, или способами, известными специалисту в данной области, такими как анализ точечных диаграмм. Альтернативно, специалист в данной области может определить повторы нуклеотидной последовательности и, таким образом, степень повторяемости нуклеотидной последовательности вручную или с помощью комплексного программного обеспечения, такого как Visual Gene Developer (Jung (2011) BMC Bioinformatics 12:340) или инструмент Repfind (Betley (2002) CurrBiol 12:1756-1761). Посредством этого молекулы нуклеиновой кислоты по изобретению, имеющие неожиданные предпочтительные свойства, можно легко отличить от молекул нуклеиновой кислоты по предшествующему уровню техники, не имеющих этих свойств.

Как описано выше, нуклеотидная последовательность с низкой повторяемостью по изобретению может иметь NRS менее 50000, или она может иметь повторы с максимальной длиной примерно 17, примерно 20, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45, примерно 50 или примерно 55 нуклеотидов. Нуклеотидная последовательность с низкой повторяемостью по изобретению имеет длину по меньшей мере 100 нуклеотидов, предпочтительно по меньшей мере 150, более предпочтительно по меньшей мере 200, еще более предпочтительно по меньшей мере 300 нуклеотидов, еще более предпочтительно по меньшей мере 350 нуклеотидов, еще более предпочтительно по меньшей мере 600 нуклеотидов, еще более предпочтительно по меньшей мере 900, еще более предпочтительно по меньшей мере 1200, еще более предпочтительно по меньшей мере 1500 нуклеотидов или, наиболее предпочтительно, по меньшей мере 1800 нуклеотидов. Иными словами, молекула нуклеиновой кислоты по изобретению содержит или состоит из нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 100 нуклеотидов, предпочтительно по меньшей мере 150, более предпочтительно по меньшей мере 200, еще более предпочтительно по меньшей мере 300 нуклеотидов, еще более предпочтительно по меньшей мере 350 нуклеотидов, еще более предпочтительно по меньшей мере 400 или 500 нуклеотидов, еще более предпочтительно по меньшей мере 600 нуклеотидов, еще более предпочтительно по меньшей мере 700 или 800 нуклеотидов, еще более предпочтительно по меньшей мере 900 нуклеотидов, еще более предпочтительно по меньшей мере 1000 или 1100, еще более предпочтительно по меньшей мере 1200 нуклеотидов (например, 1203 нуклеотида), еще более предпочтительно по меньшей мере 1300 или 1400 нуклеотидов, еще более предпочтительно по меньшей мере

1500 нуклеотидов, еще более предпочтительно по меньшей мере 1600 или 1700 нуклеотидов или, наиболее предпочтительно по меньшей мере 1800 нуклеотидов.

Молекула нуклеиновой кислоты по изобретению может содержать или состоять из нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет максимальную длину 5000 нуклеотидов, предпочтительно 4800 нуклеотидов, 3600 нуклеотидов или 2400 нуклеотидов. Молекула нуклеиновой кислоты по изобретению может содержать или состоять из нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет максимальную длину 5000 нуклеотидов, 4900 нуклеотидов, 4800 нуклеотидов, 4700 нуклеотидов, 4600 нуклеотидов, 4500 нуклеотидов, 4400 нуклеотидов, 4300 нуклеотидов, 4200 нуклеотидов, 4100 нуклеотидов, 4000 нуклеотидов, 3900 нуклеотидов, 3800 нуклеотидов, 3700 нуклеотидов, 3600 нуклеотидов, 3500 нуклеотидов, 3400 нуклеотидов, 3300 нуклеотидов, 3200 нуклеотидов, 3100 нуклеотидов, 3000 нуклеотидов, 2900 нуклеотидов, 2800 нуклеотидов, 2700 нуклеотидов, 2600 нуклеотидов, 2500 нуклеотидов, 2400 нуклеотидов, 2300 нуклеотидов, 2200 нуклеотидов, 2100 нуклеотидов, 2000 нуклеотидов или 1900 нуклеотидов.

В особенно предпочтительном аспекте молекула нуклеиновой кислоты по изобретению может содержать или состоять из нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину от 1200 до 3600 нуклеотидов.

Кроме того, нуклеотидная последовательность с низкой повторяемостью имеет показатель нуклеотидных повторов (NRS) менее 50000, предпочтительно менее 40000, более предпочтительно менее 30000, более предпочтительно менее 20000, более предпочтительно менее 10000, более предпочтительно менее 1000, более предпочтительно менее 500, еще более предпочтительно менее 100. Особенно предпочтительны нуклеотидные последовательности с низкой повторяемостью, имеющие показатель нуклеотидных повторов (NRS) менее 50, более предпочтительно менее 48, более предпочтительно менее 45, более предпочтительно менее 43, более предпочтительно менее 40, более предпочтительно менее 38 или наиболее предпочтительно менее 35. Иными словами, молекула нуклеиновой кислоты по изобретению содержит нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 50000, предпочтительно менее 40000, более предпочтительно менее 30000, более предпочтительно менее 20000, более предпочтительно менее 10000, более предпочтительно менее 1000, более предпочтительно менее 500, еще более предпочтительно менее 400, 300, 200 и еще более предпочтительно менее 100. Особенно предпочтительны молекулы нуклеиновой кислоты, содержащие нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 90, 80, 70, 60, более предпочтительно менее 50, более предпочтительно менее 48, более предпочтительно менее 45, более предпочтительно менее 43, более предпочтительно менее 40, более предпочтительно менее 39, 38, 37, или 36, или менее 39, 38, 37 или 36 или, наиболее предпочтительно, показатель NRS 35 или менее 35. Еще более предпочтительны молекулы нуклеиновой кислоты, содержащие нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) 34, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9 или 8 или NRS менее 34, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9 или 8.

Как обсуждено выше, "показатель нуклеотидных повторов" или "NRS" может быть определен по следующей формуле:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{r=1}^{k(n)} f_r(n)}}{N_{tot}}$$

Таким образом, NRS определяют как сумму, по всем возможным длинам повторов (от n равно 4 вплоть до N_{tot} минус 1), квадратов длин каждого повтора (n), умноженную на квадратный корень их общей частоты, разделенную на общую длину анализируемой нуклеотидной последовательности (N_{tot}). Иными словами, NRS представляет собой сумму квадратов длин повторов в указанной нуклеотидной последовательности, умноженную на квадратный корень суммы частот указанных повторов длиной n ($f_r(n)$), где, при наличии более чем одного повтора длиной n, $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n, в противном случае $k(n)$ для указанного повтора длиной n равен 1, и где эта сумма разделена на общую длину указанной нуклеотидной последовательности.

При использовании здесь " N_{tot} " представляет собой общую длину указанной нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина. Длина N_{tot} также представляет собой число нуклеотидов в указанной нуклеотидной последовательности. Таким образом, N_{tot} представляет собой общую длину анализируемой нуклеотидной последовательности.

При использовании здесь "n" представляет собой длину повтора в указанной нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина. Длина n также представляет собой число нуклеотидов в указанном повторе. Таким образом, n представляет собой длину повтора в анализируемой нуклеотидной последовательности. По определению, наибольший возможный повтор может иметь длину, которая на один нуклеотид меньше общей длины анализируемой нуклеотидной последовательности (N_{tot}), то есть n равно N_{tot} минус 1. С другой стороны, наименьшей длиной повтора, рассматриваемой при анализе NRS, является $n=4$, что соответствует самому короткому участку из нуклеотидов, длина которого превышает длину одного триплетного кодона для аминокислоты. Поскольку такие кодоны повторяются много раз просто в силу кодируемой аминокислотной последовательности, их не следует рассматривать как повторы нуклеотидной последовательности в отношении решаемой технической задачи.

При использовании здесь термин "повтор" означает, что нуклеотидная последовательность содержит более чем одну идентичную непрерывную последовательность нуклеотидов длиной n (то есть повтор). Иными словами, нуклеотидная последовательность содержит по меньшей мере две или множество копий непрерывной части/участка/последовательности нуклеотидов определенной длины. Иными словами, термин "повтор" относится к последовательностям нуклеотидов длиной n, встречающимся в нуклеотидной последовательности более одного раза. Здесь подразумевают, что на всем протяжении анализируемой нуклеотидной последовательности может быть только один тип повтора длиной n или могут быть более одного разных повторов одинаковой длины n. Так, нуклеотидная последовательность может, например, иметь повтор длиной n, который во всех случаях имеет одну и ту же последовательность; следует понимать, что на всем протяжении анализируемой нуклеотидной последовательности указанный повтор встречается по меньшей мере два раза, но может также встречаться множество раз. Альтернативно, существуют повторы одинаковой длины n, имеющие разные последовательности, то есть повторы имеют одинаковую длину (n), но не являются идентичными на уровне последовательности. В этом случае последовательность каждого из разных повторов составляет повтор другого типа.

При использовании здесь " $f_i(n)$ " представляет собой частоту повтора длиной n. Иными словами, $f_i(n)$ представляет собой число встречающихся повторов длиной n. При наличии только одного типа повтора длиной n, $k(n)$ равен 1. Альтернативно, при наличии более одного разных повторов длиной n, $k(n)$ представляет собой число разных последовательностей указанных повторов длиной n. Иными словами, при наличии более чем одного повтора длиной n с разными последовательностями $k(n)$ представляет собой число указанных разных последовательностей указанных повторов длиной n. Таким образом, при наличии двух или более повторов длиной n разного типа $k(n)$ представляет собой число указанных разных последовательностей указанных повторов длиной n. В противном случае, если все повторы длиной n имеют одинаковую (то есть идентичную) последовательность, $k(n)$ для указанного повтора длиной n равен 1.

Например, частота повтора длиной 14 нуклеотидов (n равна 14) в нуклеотидной последовательности (длиной N_{tot}) может составлять пять. Это может означать, что все пять повторов длиной 14 нуклеотидов имеют идентичную последовательность (принадлежат к одному и тому же типу), то есть последовательность повтора встречается в нуклеотидной последовательности пять раз. В таком случае $k(n)$ равен 1, и $f_i(n)$ равен 5. В другом варианте пять повторов длиной 14 нуклеотидов могут иметь разные последовательности. В этом варианте предполагают, что два из пяти повторов имеют одну идентичную последовательность (составляют один тип), а три из пяти повторов имеют другую идентичную последовательность (составляют второй тип), где последовательность первых двух повторов, имеющих одну последовательность, и последовательность последних трех повторов, имеющих другую последовательность, отличаются друг от друга. Таким образом, в таком варианте число указанных разных последовательностей длиной n составляет 2, то есть $k(n)$ равен 2, $f_1(n)$ равен 2 и $f_2(n)$ равен 3. Индекс "i" в обозначении $f_i(n)$ можно понимать как указание типа повтора в группе разных повторов, имеющих длину n.

Формула определения NRS включает сумму квадратов длин повторов (n^2), где n представляет собой индекс суммирования, 4 представляет собой нижнюю границу суммирования, и $N_{tot}-1$ представляет собой верхнюю границу суммирования. Таким образом, минимальная рассматриваемая длина повтора составляет 4. Повтор длиной 4 нуклеотида включает все последовательности длиннее одного триплета кодонов, кодирующего аминокислоту. Для каждого последующего значения индекс n увеличивается с шагом 1 до n равного $N_{tot}-1$.

Кроме того, формула для определения NRS включает квадратный корень суммы частот указанных повторов длиной n ($f_i(n)$), где i представляет собой индекс суммирования, $f_i(n)$ представляет собой индексную переменную, представляющую каждое последующее значение в ряду, 1 представляет собой нижнюю границу суммирования, и $k(n)$ представляет собой верхнюю границу суммирования. Для каждого последующего значения индекс i увеличивается с шагом 1 до i равного $k(n)$. Таким образом, при наличии только одного типа повтора длиной n, то есть если все повторы длиной n имеют идентичную последовательность, $k(n)$ для указанного повтора длиной n равен 1, и вместо суммы анализируют только частоту $f_i(n)$ данного повтора длиной n на всем протяжении нуклеотидной последовательности (N_{tot}).

Например, для определения NRS согласно изобретению может быть применен следующий сцена-

рий калькулятора NRS:

```

import math
import sys
class NRSCalculator:
    def __init__(self):
        self.repeats = dict()
        self.sums = dict()
        self.seq = None
        self.range_min = None
        self.range_max = None
    def _match_at(self, row, column):
        return self.seq[row] == self.seq[column]
    def _get_repeats_at(self, row, column):
        length = 1
        search_row = row
        search_column = column
        while True:
            if not 0 <= search_row < len(self.seq):
                break
            if not 0 <= search_column < search_row:
                break
            if length > self.range_max:
                break
            if not self._match_at(search_row, search_column):
                break
            if length >= self.range_min:
                repeats =
self.repeats.setdefault(self.seq[row:row + length], set())
                repeats.add(row)
                repeats.add(column)
                search_row += 1
                search_column += 1
                length += 1
    def _get_repeats(self):
        self.repeats = dict()
        for row in xrange(len(self.seq)):
            for column in xrange(row):
                self._get_repeats_at(row, column)
    def _get_sums(self):
        self.sums = dict()
        for (seq, repeats) in self.repeats.iteritems():
            length = len(seq)
            self.sums[length] = self.sums.get(length, 0) +
len(repeats)
    def set_range(self, range_min, range_max):
        self.range_min = range_min
        self.range_max = range_max
    def set_sequence(self, seq):
        self.seq = seq
    def work(self):
        if not self.seq and not self.range_min and not
self.range_max:
            raise RuntimeError('Can not work without
initialization')
        self._get_repeats()
        self._get_sums()
    def print_repeats(self):
        print('Sequence (Length bp) : NumRepeats (Positions)')
        for seq, repeats in sorted(self.repeats.iteritems(),
key=lambda t: len(t[0])):
            list = [seq, len(seq), len(repeats)]
            list.extend(map(lambda value: value + 1,
sorted(repeats)))
            print('%s Ntot = %u : %u (%s)' % (seq, len(seq),
len(repeats), ', '.join(map(lambda value: str(value + 1),
sorted(repeats)))))
    def print_sums(self):
        print('Length\tFrequency')
        for item in self.sums.iteritems():
            print('%u\t%u' % item)
    def print_score(self):
        sum = 0
        for length, count in self.sums.iteritems():
            sum += (length ** 2) * math.sqrt(count)
        print('NRS = %.0f' % (sum / len(self.seq)))
def handle_sequence(finder, name, sequence):
    finder.set_range(4, len(sequence))
    finder.set_sequence(sequence)
    finder.work()
    print('%s: Ntot = %u' % (name, len(sequence)))
    #finder.print_repeats()
    finder.print_sums()
    finder.print_score()
if len(sys.argv) != 2:
    print('Usage: %s FILENAME' % sys.argv[0])
    sys.exit(1)
finder = NRSCalculator()
with open(sys.argv[1], 'r') as infile:
    name = 'Unnamed'
    seq = ''
    for line in infile:

```

```

line = line.strip()
if line.startswith('>'):
    if len(seq) > 0:
        handle_sequence(finder, name, seq)
    name = line
    seq = ''
    continue
seq += line.upper()
handle_sequence(finder, name, seq)

```

Кроме того, изобретение также относится к нуклеотидной последовательности, содержащей нуклеотидные повторы, имеющие максимальную длину n_{\max} , определенную по следующей формуле:

$$n_{\max} \leq 17 + \frac{N_{\text{tot}}}{600},$$

где N_{tot} представляет собой общую длину указанной нуклеотидной последовательности. При использовании здесь термин "максимальная длина" или " n_{\max} " определяет число нуклеотидов в наиболее длинной непрерывной части/участке/последовательности нуклеотидов, присутствующей в указанной нуклеотидной последовательности или молекуле нуклеиновой кислоты в по меньшей мере двух копиях. Иными словами, при использовании здесь термин "максимальная длина" или " n_{\max} " означает, что нуклеотидная последовательность молекулы нуклеиновой кислоты по настоящему изобретению не имеет повторов длиннее, чем указанная длина.

В прилагаемых примерах продемонстрировано, что типичные молекулы нуклеиновой кислоты по изобретению содержат лишь небольшое число повторов, например, длиной 14 нуклеотидов; см., например, прилагаемый пример 2. Как объяснено выше, анализ повторов может быть проведен с использованием любого подходящего инструмента, такого как анализ NRS, предложенный здесь, вручную или с помощью комплексного программного обеспечения, такого как анализ точечных диаграмм, например с применением Visual Gene Developer (Jung (2011), см. выше) или инструмента Repfind (Betley (2002), см. выше). Точечная диаграмма является графическим представлением сходства двух последовательностей.

В прилагаемых примерах нуклеотидные последовательности, предложенные здесь, выровнены сами с собой. Каждая ось прямоугольной диаграммы отражает одну из двух (полностью идентичных) сравниваемых нуклеотидных последовательностей. Все положения первой введенной последовательности сравнивают со всеми положениями второй введенной последовательности с присвоением им значений с использованием определенной подстановочной матрицы. Это приводит к получению матрицы значений, по которой определяют местные области сходства/идентичности (соответствующие диагоналям на точечной диаграмме). По всем возможным диагоналям перемещается окно/порог повтора, длину которых указывает пользователь. Каждое положение в окне/пороге повтора соответствует парному значению из матрицы значений. Значение всего окна представляет собой сумму значений отдельных положений в нем. Если значение окна превышает определенный пользователем порог, то на точечной диаграмме отображается линия, соответствующая этому окну (см., например, <http://emboss.sourceforge.net/apps/cvs/emboss/apps/dotmatcher.html>).

Анализ точечных диаграмм применен в прилагаемых примерах, например в примере 3, для анализа повторяемости нуклеотидных последовательностей молекул нуклеиновой кислоты. В них доказано, что, в случае применения окна/порога повтора 14, молекулы нуклеиновой кислоты по изобретению, например SEQ ID NO: 36 (называемая здесь PA#3b(200)) или SEQ ID NO: 28 (называемая здесь PA#1b), имеют лишь небольшое число разрозненных повторов длиной 14 нуклеотидов на протяжении всей нуклеотидной последовательности длиной 600 нуклеотидов. Увеличение окна/порога повтора 14 на один нуклеотид, то есть окно/порог повтора 15 нуклеотидов, не приводит к обнаружению дополнительных повторов во всей анализируемой нуклеотидной последовательности (см. фиг. 2 и пример 3). Это означает, что максимальная длина повтора во всей исследуемой последовательности нуклеиновой кислоты составляет 14 нуклеотидов (включая также более короткие повторы). В целом, можно предполагать, что чем короче такие нуклеотидные повторы, тем меньше их отрицательное влияние на генетическую стабильность.

Тем не менее, при применении окна/порога повтора, например, 14 или 15 нуклеотидов, как использовано в прилагаемых примерах (см. пример 3), молекулы нуклеиновой кислоты, кодирующие последовательности с высоким содержанием пролина/аланина, раскрытые в предшествующем уровне техники, демонстрируют более длинные и/или более многочисленные повторяющиеся последовательности. Например, последовательность с высоким содержанием пролина/аланина, раскрытая в WO 2011/144756 (SEQ ID NO: 15; называемая здесь PA#3a(200)), имеет множество нуклеотидных повторов длиной 60, которые расположены друг за другом и даже перекрываются, приводя таким образом к большому количеству длинных параллельных линий. Другие молекулы нуклеиновой кислоты, раскрытые в предшествующем уровне техники, такие как PAS#1a(600), показанная в SEQ ID NO: 12 и раскрытая в WO 2008/155134, мультимер $[(AP)_5]_n$, показанный в SEQ ID NO: 16 и раскрытый в US 2006/0252120, или крупный оболочечный белок Масацине herpesvirus 1, показанный в SEQ ID NO: 18 и раскрытый в GenBank под номером AAR41454.1, также демонстрируют высокую повторяемость своих нуклеотидных последовательностей.

В этом состоит их основное отличие от молекул нуклеиновой кислоты по настоящему изобретению,

которые демонстрируют низкую повторяемость на уровне нуклеотидной последовательности. Например, при сравнении числа повторов в последовательности по предшествующему уровню техники РА#3а(200) (SEQ ID NO: 15) с числом повторов в нуклеотидной последовательности с низкой повторяемостью по изобретению РА#3b(200) (SEQ ID NO: 36) с применением окна/порога повтора 14 нуклеотидов, можно наблюдать, что последовательность по изобретению имеет 29 повторов (на 600 нуклеотидных остатков) в сравнении с намного более чем 100 повторами (на 600 нуклеотидных остатков) в последовательности по предшествующему уровню техники. При применении окна 15 нуклеотидов типичная анализируемая молекула нуклеиновой кислоты или нуклеотидная последовательность по изобретению не содержит никаких повторов. В отличие от этого, в молекулах нуклеиновых кислот по предшествующему уровню техники можно обнаружить множество повторов длиной 15 нуклеотидов и более. В прилагаемых примерах доказано, что окно/порог повтора от 14 до 20, в частности, например, 14, 15, 16, 17, 18, 19 или 20 нуклеотидов, является подходящим для подтверждения того, что молекула нуклеиновой кислоты по изобретению имеет нуклеотидную последовательность с низкой повторяемостью. Следует понимать, что число повторов, обнаруживаемых в определенной нуклеотидной последовательности, обратно пропорционально длине окна/порога повтора. Например, если длина окна/порога повтора составляет 1, то число повторов может быть равно числу всех нуклеотидных остатков (А, Т, G и/или С) в нуклеотидной последовательности (при условии, что нуклеотид каждого типа представлен в анализируемой последовательности по меньшей мере двумя копиями). При увеличении длины окна/порога повтора число повторов в анализируемой нуклеотидной последовательности будет уменьшаться специфичным для последовательности образом. Следовательно, нуклеотидные последовательности "с низкой повторяемостью" по изобретению также могут содержать повторы; но, при сравнении с последовательностями по предшествующему уровню техники с применением одних и тех же параметров анализа, длина и число этих повторов меньше.

Таким образом, предложенная(ые) здесь молекула(ы) нуклеиновой кислоты содержит(ат) повторы, имеющие максимальную длину 14, 15, 16, 17, примерно 18, примерно 20, примерно 21, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45, примерно 50 или примерно 55 нуклеотидов. Кроме того, предложенная здесь молекула нуклеиновой кислоты содержит повторы, имеющие максимальную длину от примерно 17 нуклеотидов до примерно 55 нуклеотидов. В контексте максимальной длины повтора термин "примерно" означает, что максимальная длина повтора может отличаться от указанной длины повтора не более чем на плюс/минус 4 нуклеотида. Иными словами, в данном контексте "примерно" относится к диапазону, где длина повтора может быть на 4 нуклеотида больше или меньше указанной длины повтора. Например, максимальная длина повтора примерно 55 нуклеотидов относится к нуклеотидной последовательности, содержащей повторы, имеющие максимальную длину от 51 до 59 нуклеотидов. Кроме того, максимальная длина повтора примерно 17 нуклеотидов относится к нуклеотидной последовательности, содержащей повторы, имеющие максимальную длину от 13 до 21 нуклеотида.

Кроме того, изобретение относится к молекуле нуклеиновой кислоты, содержащей повторы, имеющие максимальную длину 59 нуклеотидов, предпочтительно 54 нуклеотида, более предпочтительно 50 нуклеотидов, более предпочтительно 48 нуклеотидов, более предпочтительно 40 нуклеотидов, более предпочтительно 36 нуклеотидов, более предпочтительно 35 нуклеотидов, более предпочтительно 30 нуклеотидов, более предпочтительно 25 нуклеотидов, более предпочтительно 24 нуклеотида, более предпочтительно 21 нуклеотид, более предпочтительно 20 нуклеотидов, более предпочтительно 19 нуклеотидов, более предпочтительно 18 нуклеотидов, более предпочтительно 16 нуклеотидов, более предпочтительно 15 нуклеотидов, более предпочтительно 14 нуклеотидов или наиболее предпочтительно 17 нуклеотидов.

При использовании здесь молекула нуклеиновой кислоты, содержащая повторы, имеющие максимальную длину, например, 17 нуклеотидов, относится к молекуле нуклеиновой кислоты, содержащей повторяющиеся последовательности длиной 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 или 17 нуклеотидов. Сходным образом, молекула нуклеиновой кислоты, содержащая повторы, имеющие максимальную длину 14 нуклеотидов, относится к молекуле нуклеиновой кислоты, содержащей повторы длиной до 14 нуклеотидов, то есть "14 нуклеотидов или менее" или "1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 или 14 нуклеотидов". Иными словами, молекула нуклеиновой кислоты, содержащая повторы, имеющие максимальную длину 14 нуклеотидов, не содержит повторов, которые длиннее 14 нуклеотидов, то есть длиной "более 14 нуклеотидов".

Кроме того, изобретение относится к молекуле нуклеиновой кислоты с низкой повторяемостью, где низкая повторяемость означает, что участок нуклеотидной последовательности, повторяющийся в нуклеотидной последовательности большей длины, кодирующей повторяющуюся аминокислотную последовательность, соответствует максимум 0,05%, предпочтительно 0,1%, более предпочтительно 0,5%, более предпочтительно 1%, более предпочтительно 2%, более предпочтительно 3%, более предпочтительно 4%, более предпочтительно 5%, более предпочтительно 6%, более предпочтительно 7%, более предпочтительно 8%, более предпочтительно 9%, более предпочтительно 10%, более предпочтительно 15%, более предпочтительно 20%, более предпочтительно 25%, более предпочтительно 30%, более предпочтительно 40% или наиболее предпочтительно 50% длины нуклеотидной последовательности, кодирующей повторяющийся участок аминокислотной последовательности. Иными словами, нуклеиновая кислота по

настоящему изобретению содержит повторы, имеющие максимальную длину, соответствующую 0,05%, предпочтительно 0,1%, более предпочтительно 0,5%, более предпочтительно 1%, более предпочтительно 2%, более предпочтительно 3%, более предпочтительно 4%, более предпочтительно 5%, более предпочтительно 6%, более предпочтительно 7%, более предпочтительно 8%, более предпочтительно 9%, более предпочтительно 10%, более предпочтительно 15%, более предпочтительно 20%, более предпочтительно 25%, более предпочтительно 30%, более предпочтительно 40% или наиболее предпочтительно 50% длины указанной нуклеотидной последовательности, кодирующей повтор аминокислотной последовательности в полипептиде, состоящем из пролина, аланина и возможно серина.

Очевидно, что согласно настоящему изобретению предложены не только новые нуклеотидные последовательности по изобретению, кодирующие последовательности с высоким содержанием РА и имеющие длину по меньшей мере 300 нуклеотидов (что соответствует 100 аминокислотным остаткам), но также согласно настоящему изобретению предложены средства и способы отбора рекомбинантных и/или синтетических молекул нуклеиновой кислоты, позволяющие получать генетически стабильные последовательности и/или обеспечивающие удобное клонирование, секвенирование и/или амплификацию. Этот способ отбора основан на NRS, предложенном здесь, и/или максимальной длине повторов, описанной здесь. Соответственно настоящее изобретение относится к способу отбора генетически стабильной молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, где указанная нуклеотидная последовательность имеет длину по меньшей мере 300 нуклеотидов, включающему стадию отбора молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую показатель нуклеотидных повторов (NRS) менее 50000, где указанный показатель нуклеотидных повторов (NRS) определен по формуле, приведенной здесь выше. Кроме того, настоящее изобретение относится к способу отбора генетически стабильной молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, включающему стадию отбора указанной нуклеотидной последовательности, содержащей повторы, имеющие максимальную длину n_{max} . Кроме того, настоящее изобретение относится к способу отбора генетически стабильной молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, включающему стадию отбора указанной нуклеотидной последовательности, содержащей повторы, имеющие максимальную длину от примерно 17 нуклеотидов до примерно 55 нуклеотидов. Воплощения и параметры формулы NRS и нуклеотидных последовательностей с низкой повторяемостью, предложенных здесь выше в контексте нуклеиновых кислот по изобретению, кодирующих последовательности с высоким содержанием РА, применимы, с необходимыми поправками, к предложенным и наглядно продемонстрированным здесь способам отбора генетически стабильных молекул нуклеиновой кислоты, кодирующих последовательности с высоким содержанием РА, а также к другим воплощениям.

Здесь было неожиданно обнаружено и продемонстрировано, что молекулы нуклеиновой кислоты по изобретению имеют повышенную стабильность *in vivo*. Как показано в прилагаемых иллюстративных примерах, например в примере 6, предложенные здесь молекулы нуклеиновой кислоты более стабильны, чем молекулы нуклеиновой кислоты, содержащие больше повторяющихся нуклеотидных последовательностей. В примере 6 проведено сравнение молекул нуклеиновой кислоты, кодирующих полипептиды, состоящие из пролина, аланина и серина, различающихся повторяемостью нуклеотидных последовательностей. Молекулы нуклеиновой кислоты по изобретению, содержащие нуклеотидные последовательности с низкой повторяемостью, предложенные здесь (см. также фиг. 2B), например SEQ ID NO: 50, имеют более высокую генетическую стабильность, чем нуклеотидные последовательности с высокой повторяемостью, такие как по предшествующему уровню техники, например SEQ ID NO: 51. Таким образом, при использовании здесь термин "стабильность *in vivo*" относится к "генетической стабильности". При использовании здесь термин "генетическая стабильность" означает, что нуклеиновая кислота стабильно поддерживается в клетке-хозяине и что последовательность не претерпевает генетических модификаций, например посредством мутации, вставки или делеции. При использовании здесь мутации относятся к изменениям нуклеотидной последовательности, например заменам, делециям, вставкам или удлинением. В прилагаемых примерах показано, что молекулы нуклеиновой кислоты, содержащие нуклеотидные последовательности с высокой повторяемостью, склонны к делециям или укорочению; см., например, пример 6. Поэтому молекулы нуклеиновой кислоты, содержащие нуклеотидные последовательности с высокой повторяемостью/повторяющиеся генные кассеты, имеют меньшую генетическую стабильность, что может возникать, например, из-за рекомбинации при делении клеток. Параметры NRS и n_{max} , определенные здесь, обеспечивают объективное средство различения нуклеотидных последовательностей с высокой повторяемостью, таких как последовательности по предшествующему уровню техники, имеющих высокие показатели NRS и/или n_{max} , и нуклеотидных последовательностей с низкой повторяемостью по изобретению, имеющих низкие показатели NRS и/или n_{max} .

Молекулы нуклеиновой кислоты, предложенные здесь, кодируют полипептиды, содержащие повторяющиеся аминокислотные последовательности. Следует отметить, что кодируемая повторяющаяся аминокислотная последовательность может также быть частью/фрагментом слитого белка. Например,

молекула нуклеиновой кислоты, раскрытая здесь, может кодировать полипептид, состоящий из повторов с высоким содержанием пролина/аланина, например повторов, состоящих из пролина, аланина и возможно серина. Полипептид с высоким содержанием пролина/аланина может образовывать случайный клубок. В определенных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, состоящий из пролина, аланина и серина, образующий случайный клубок. В определенных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, состоящий из пролина и аланина, образующий случайный клубок.

В частности, нуклеотидная последовательность кодирует полипептид, состоящий из пролина, аланина и возможно серина. Этот кодируемый полипептид образует случайный клубок. Указанная нуклеотидная последовательность входит в состав молекулы нуклеиновой кислоты, предложенной здесь. Таким образом, в определенных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность, состоящий из пролина, аланина и возможно серина и образующий случайный клубок. В определенных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность, состоящий из пролина, аланина и серина и образующий случайный клубок. В определенных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность, состоящий из пролина и аланина и образующий случайный клубок.

При использовании здесь термин "случайный клубок" относится к любой конформации полимерной молекулы, включая аминокислотные полимеры, в частности полипептиды из L-аминокислот, где отдельные мономерные элементы, образующие указанную полимерную структуру, ориентированы, по существу, случайным образом относительно соседнего мономерного элемента или элементов, оставаясь при этом химически связанными. В частности, кодируемый полипептид или аминокислотный полимер, принимающий/имеющий/образующий "конформацию в виде случайного клубка", по существу не имеет определенной вторичной и третичной структуры. Природа кодируемых полипептидных случайных клубков и способы их экспериментального определения известны специалисту в данной области и описаны в научной литературе (Cantor (1980) *Biophysical Chemistry*, 2nd ed., W. H. Freeman and Company, New York; Creighton (1993) *Proteins - Structures and Molecular Properties*, 2nd ed., W. H. Freeman and Company, New York; Smith (1996) *Fold. Des.* 1:R95-R106) и патентной литературе, например WO 2011/144756 и WO 2008/155134.

Кодируемые полипептиды, образующие случайные клубки, по настоящему изобретению принимают/образуют конформацию в виде случайного клубка, например, в водном растворе и/или в физиологических условиях. Термин "физиологические условия" известен в данной области и относится к условиям, в которых белки обычно принимают свою нативную свернутую конформацию. Конкретнее, термин "физиологические условия" относится к биофизическим условиям окружающей среды, типичным для высших форм жизни и особенно для млекопитающих, наиболее предпочтительно для людей. Термин "физиологические условия" может относиться к биохимическим и биофизическим параметрам, обычно наблюдаемым в организме, особенно в биологических жидкостях, млекопитающих и особенно у людей. Указанные "физиологические условия" могут относиться к соответствующим параметрам, наблюдаемым в здоровом организме, а также к параметрам, наблюдаемым при болезненных состояниях или у пациентов-людей. Например, больное млекопитающее или пациент-человек могут иметь повышенную, но "физиологическую" температуру тела (то есть температурные условия), когда указанное млекопитающее или указанный человек страдает от лихорадки. Применительно к "физиологическим условиям", в которых белки принимают свою нативную конформацию/состояние, наиболее важными параметрами являются температура (37°C для организма здорового человека), pH (7,35-7,45 для человеческой крови), осмоляльность (280-300 ммоль/кг H₂O) и, по необходимости, общее содержание белка (66-85 г/л сыворотки).

Тем не менее, специалисту в данной области известно, что в физиологических условиях эти параметры могут варьировать, например температура, pH, осмоляльность и содержание белка в заданных биологических или тканевых жидкостях, таких как кровь, спинномозговая жидкость, перитонеальная жидкость и лимфа, могут различаться (Klinke (2005) *Physiologie*, 4th edition., Georg Thieme Verlag, Stuttgart). Например, в спинномозговой жидкости осмоляльность может составлять примерно 290 ммоль/кг H₂O, а концентрация белка может составлять от 0,15 до 0,45 г/л, в то время как в лимфе pH может составлять примерно 7,4, а содержание белка может составлять от 3 до 5 г/л. При определении того, образует/принимает ли кодируемый полипептид/аминокислотная последовательность конформацию в виде случайного клубка в экспериментальных условиях, биофизические параметры, такие как температура, pH, осмоляльность и содержание белка, могут отличаться от физиологических условий, обычно наблюдаемых *in vivo*. Температуры, составляющие от 1 до 42°C или предпочтительно от 4 до 25°C, можно считать применимыми для анализа и/или подтверждения биофизических свойств и биологической активности белка в физиологических условиях *in vitro*.

Принято считать, что некоторые буферы, которые могут содержать растворители и/или эксципиенты для фармацевтических композиций, представляют собой "физиологические растворы"/"физиологические условия" *in vitro*, в частности в экспериментальных условиях, например в контек-

сте измерения CD или других способов, позволяющих специалисту в данной области определять структурные свойства белка/аминокислотной последовательности. Примерами таких буферов являются, например, забуференный фосфатом физиологический раствор (PBS, например, 115 mM NaCl, 4 mM KH_2PO_4 , 16 mM Na_2HPO_4 , pH 7,4), трис-буферы, ацетатные буферы, цитратные буферы или сходные с ними буферы. Обычно pH буфера, соответствующего "условиям физиологического раствора", должен находиться в диапазоне от 6,5 до 8,5, предпочтительно в диапазоне от 7,0 до 8,0, наиболее предпочтительно в диапазоне от 7,2 до 7,7, а осмоляльность должна находиться в диапазоне от 10 до 1000 ммоль/кг H_2O , более предпочтительно в диапазоне от 50 до 500 ммоль/кг H_2O и наиболее предпочтительно в диапазоне от 200 до 350 ммоль/кг H_2O . Возможно содержание белка в физиологическом растворе может находиться в диапазоне от 0 до 100 г/л без учета самого исследуемого белка, принимающего конформацию в виде случайного клубка; кроме того, могут присутствовать типичные стабилизирующие белки, например человеческий или бычий сывороточный альбумин.

Полипептиды, кодируемые молекулами нуклеиновой кислоты по изобретению, образуют конформацию в виде случайного клубка не только в физиологических условиях, но, в более широком смысле, в водном растворе; например, см. WO 2011/144756. Термин "водный раствор" хорошо известен в данной области. "Водный раствор" может представлять собой раствор с содержанием воды (H_2O) по меньшей мере примерно 20%, по меньшей мере примерно 30%, по меньшей мере примерно 40%, по меньшей мере примерно 50%, по меньшей мере примерно 60%, по меньшей мере примерно 70%, по меньшей мере примерно 80% или по меньшей мере примерно 90% H_2O (мас./мас.). Соответственно кодируемые полипептиды, предложенные в настоящем изобретении, могут образовывать конформацию в виде случайного клубка в водном растворе, возможно содержащем другие смешивающиеся растворители, или в водных дисперсиях с более широким диапазоном температур, значений pH, осмоляльности или содержания белка. Это особенно важно для применения полипептида, образующего случайный клубок, вне медицинской терапии или диагностики *in vivo*, например, в косметике, питании или пищевых технологиях.

В контексте данного изобретения также подразумевают, что конформация кодируемого полипептида в виде случайного клубка сохраняется в фармацевтических композициях, таких как жидкие фармацевтические/биологические композиции или лиофилизированные фармацевтические композиции. Это особенно важно в контексте кодируемых биологически активных белков или конъюгатов лекарственных средств, предложенных здесь, содержащих, среди прочего, полипептид, образующий случайный клубок. Предпочтительно, в соответствующих буферных системах, растворителях и/или эксципиентах следует использовать "физиологические условия". Однако подразумевают, что, например, в лиофилизированных или высушенных композициях (таких как, например, фармацевтические/биологические композиции) конформация в виде случайного клубка у предложенного здесь кодируемого полипептида, образующего случайный клубок, может временно отсутствовать и/или не поддаваться определению. Тем не менее, после восстановления в соответствующих буферах/растворах/эксципиентах/растворителях или после введения в организм пациента или животного указанный кодируемый полипептид, образующий случайный клубок, будет снова принимать/образовывать свою конформацию в виде случайного клубка.

В определенных аспектах настоящего изобретения молекулы нуклеиновой кислоты, раскрытые здесь, кодируют полипептиды, состоящие (главным образом или исключительно) из пролина, аланина и возможно серина, где идентичны не более 9 расположенных друг за другом аминокислотных остатков. Такие кодируемые полипептиды образуют случайный клубок. В определенном аспекте кодируемые аминокислотные последовательности/полипептиды, принимающие конформацию в виде случайного клубка, могут содержать множество аминокислотных повторов, состоящих главным образом или исключительно из пролиновых, аланиновых и возможно сериновых аминокислотных остатков, где идентичны не более 9 расположенных друг за другом аминокислотных остатков. В определенном аспекте кодируемые аминокислотные последовательности/полипептиды, принимающие конформацию в виде случайного клубка (полипептид, образующий случайный клубок, как определено здесь), могут содержать множество аминокислотных повторов, состоящих главным образом или исключительно из пролиновых, аланиновых и сериновых аминокислотных остатков, где идентичны не более 9 расположенных друг за другом аминокислотных остатков. В определенном аспекте кодируемые аминокислотные последовательности/полипептиды, принимающие конформацию в виде случайного клубка, могут содержать множество аминокислотных повторов, состоящих главным образом или исключительно из пролиновых и аланиновых аминокислотных остатков, где идентичны не более 9 расположенных друг за другом аминокислотных остатков.

В предпочтительных аспектах молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 8 расположенных друг за другом аминокислотных остатков и где указанный полипептид образует случайный клубок. Предпочтительно молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 7 расположенных друг за другом аминокислотных остатков, и где указанный полипептид образует случайный клубок. Более предпочтительно молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся

аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 6 расположенных друг за другом аминокислотных остатков, и где указанный полипептид образует случайный клубок. Особенно предпочтительно, молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 5 расположенных друг за другом аминокислотных остатков, и где указанный полипептид образует случайный клубок. Еще предпочтительнее, молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 4 расположенных друг за другом аминокислотных остатков, и где указанный полипептид образует случайный клубок. Наиболее предпочтительно, молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, содержащий повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 3 расположенных друг за другом аминокислотных остатков, и где указанный полипептид образует случайный клубок.

Неограничивающий пример аминокислотного повтора, состоящего исключительно из пролиновых, аланиновых и сериновых остатков, приведен ниже; см., например, SEQ ID NO: 6. Неограничивающий пример кодируемого аминокислотного повтора, состоящего исключительно из пролиновых и аланиновых остатков, приведен ниже; см., например, SEQ ID NO: 8. Кодируемый полипептид может содержать множество копий одной и той же последовательности или разных последовательностей.

Молекула нуклеиновой кислоты, раскрытая здесь, кодирует полипептид, состоящий главным образом или исключительно из трех аминокислотных остатков: пролина (Pro, P), аланина (Ala, A) и возможно серина (Ser, S). При использовании здесь термин "возможно" означает, что кодируемый полипептид по настоящему изобретению либо состоит главным образом или исключительно из пролина, аланина и серина, либо состоит главным образом или исключительно из пролина и аланина. Кодируемый полипептид, состоящий главным образом или исключительно из трех аминокислотных остатков, пролина, аланина и серина, назван здесь "PAS-полипептидом". Кодируемый полипептид, состоящий главным образом или исключительно из двух аминокислотных остатков, пролина и аланина, назван здесь "PA-полипептидом". Неограничивающий пример кодируемого полипептида, состоящего из пролина и аланина, приведен в SEQ ID NO: 8. Неограничивающий пример кодируемого полипептида, состоящего из пролина, аланина и серина, приведен в SEQ ID NO: 6. При использовании здесь термин "главным образом" означает, что предпочтительно по меньшей мере примерно 90% или по меньшей мере примерно 95% кодируемых аминокислот представляют собой пролин, аланин и возможно серин, где пролин, аланин и серин в сумме составляют большую часть аминокислотных остатков, но могут не быть единственными аминокислотными остатками; таким образом, кодируемые аминокислотные последовательности не обязательно на 100% представляют собой пролин, аланин и возможно серин. Поэтому кодируемые полипептиды/аминокислотные последовательности могут также содержать другие аминокислоты, отличные от пролина, аланина и возможно серина, в качестве минорных компонентов, при условии, что аминокислотная последовательность образует/принимает/имеет конформацию в виде случайного клубка. Такая конформация в виде случайного клубка может быть легко определена средствами и способами, описанными здесь. Соответственно, в одном воплощении настоящее изобретение относится к молекуле нуклеиновой кислоты, кодирующей полипептид, образующий случайный клубок, где аминокислотная последовательность состоит главным образом из пролина, аланина и возможно серина.

В случае если кодируемый полипептид состоит из пролина и аланина, указанные пролиновые остатки составляют более чем примерно 10% и менее чем примерно 75% указанной кодируемой аминокислотной последовательности. Соответственно кодируемый полипептид, образующий случайный клубок, состоит главным образом из пролина и аланина, где пролиновые остатки составляют более чем примерно 10% и менее чем примерно 75% аминокислотной последовательности. Оставшиеся по меньшей мере 25-90% указанной аминокислотной последовательности представлены аланиновыми остатками.

Предпочтительно кодируемая аминокислотная последовательность содержит более чем примерно 10%, предпочтительно более чем примерно 12%, более предпочтительно более чем примерно 14, 18, 20%, более предпочтительно более чем примерно 22, 23, 24 или 25%, более предпочтительно более чем примерно 27, 29 или 30%, более предпочтительно более чем примерно 32, 33 или 34% и наиболее предпочтительно более чем примерно 35% пролиновых остатков. Аминокислотная последовательность предпочтительно содержит менее чем примерно 75%, более предпочтительно менее 65%, более предпочтительно менее 60%, более предпочтительно менее 55%, более предпочтительно менее 50% пролиновых остатков, при этом предпочтительны меньшие значения. Еще предпочтительнее, аминокислотная последовательность содержит менее чем примерно 48, 46, 44, 42% пролиновых остатков. Более предпочтительны аминокислотные последовательности, содержащие менее чем примерно 41, 40, 39, 38, 37 или 36% пролиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительны аминокислотные последовательности, содержащие менее чем примерно 34, 32 или 30%. Более предпочтительны аминокислотные последовательности, содержащие менее чем примерно 28, 26 или 25%. Наиболее предпочтительно, аминокислотные последовательности содержат менее чем примерно 35% пролиновых остатков.

Наоборот, аминокислотная последовательность предпочтительно содержит менее чем примерно 90%, более предпочтительно менее 88%, 86%, 84%, 82% или 80% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно, аминокислотная последовательность содержит менее чем примерно 79%, 78%, 77%, 76% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно, аминокислотная последовательность содержит менее чем примерно 74%, 72% или 70% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно, аминокислотная последовательность содержит менее чем примерно 69%, 67% или 65% аланиновых остатков, при этом предпочтительны меньшие значения. Наиболее предпочтительно, аминокислотная последовательность содержит менее чем примерно 75% аланиновых остатков. Кроме того, здесь предпочтительна аминокислотная последовательность, содержащая более чем примерно 25%, предпочтительно более чем примерно 30%, более предпочтительно более чем примерно 35%, более предпочтительно более чем примерно 40%, более предпочтительно более чем примерно 45%, более предпочтительно более чем примерно 50%, более предпочтительно более чем примерно 52, 54, 56, 58 или 59% аланиновых остатков, при этом предпочтительны большие значения. Еще предпочтительнее, аминокислотная последовательность содержит более чем примерно 60, 61, 62, 63 или 64% аланиновых остатков. Более предпочтительно аминокислотная последовательность содержит более чем примерно 66, 67, 69 или 70% аланиновых остатков. Более предпочтительно, аминокислотная последовательность содержит более чем примерно 72, 74 или 75% аланиновых остатков. Наиболее предпочтительно аминокислотная последовательность содержит более чем примерно 65% аланиновых остатков.

Таким образом, полипептид, образующий случайный клубок, может содержать аминокислотную последовательность, состоящую из примерно 25% или 30% пролиновых остатков примерно 75% или 70% аланиновых остатков, соответственно. Альтернативно, полипептид, образующий случайный клубок, может содержать аминокислотную последовательность, состоящую из примерно 35% пролиновых остатков, и примерно 65% аланиновых остатков. При использовании здесь термин "примерно X%" не ограничен точным числом процентов, но также включает значения с 10-20% дополнительными остатками или с на 10-20% меньшим количеством остатков. Например, термин 10% может также относиться к 11 или 12% и к 9 или 8% соответственно.

В случае если кодируемый полипептид состоит из пролина, аланина и серина, указанные пролиновые остатки составляют более чем примерно 4% и менее чем примерно 40% указанной кодируемой аминокислотной последовательности. Оставшуюся часть указанной аминокислотной последовательности составляют аланиновые и сериновые остатки.

Предпочтительно, кодируемая аминокислотная последовательность содержит более чем примерно 4%, предпочтительно более чем примерно 6%, более предпочтительно более чем примерно 10%, более предпочтительно более чем примерно 15%, более предпочтительно более чем примерно 20%, более предпочтительно более чем примерно 22, 23 или 24%, более предпочтительно более чем примерно 26, 29 или 30%, более предпочтительно более чем примерно 31, 32, 33, 34 или 35% и наиболее предпочтительно более чем примерно 25% пролиновых остатков. Кодируемая аминокислотная последовательность предпочтительно содержит менее чем примерно 40%, более предпочтительно менее чем 38, 35, 30 или 26% пролиновых остатков, при этом предпочтительны меньшие значения.

Кодируемая аминокислотная последовательность предпочтительно содержит менее чем примерно 95%, более предпочтительно менее 90, 86, 84, 82 или 80%, при этом предпочтительны меньшие значения. Более предпочтительно, кодируемая аминокислотная последовательность содержит менее чем примерно 79, 78, 77, 76% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 75, 73, 71 или 70% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 69, 67, 66 или 65% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 64, 63, 62 или 60% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 59, 57, 56 или 55% аланиновых остатков, при этом предпочтительны меньшие значения. Более предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 54, 53 или 51% аланиновых остатков, при этом предпочтительны меньшие значения. Наиболее предпочтительно кодируемая аминокислотная последовательность содержит менее чем примерно 50% аланиновых остатков.

Кроме того, здесь предпочтительна кодируемая аминокислотная последовательность, содержащая более чем примерно 10%, предпочтительно более чем примерно 15, 17, 19 или 20%, более предпочтительно более чем примерно 22, 24 или 25%, более предпочтительно более чем примерно 27, 29 или 30%, более предпочтительно более чем примерно 32, 34 или 35%, более предпочтительно более чем примерно 37, 39 или 40%, более предпочтительно более чем примерно 42, 44 или 45%, более предпочтительно более чем примерно 46, 47 или 49% аланиновых остатков, при этом предпочтительны большие значения. Наиболее предпочтительно кодируемая аминокислотная последовательность содержит более чем примерно 50% аланиновых остатков. Как указано выше, оставшуюся часть указанной аминокислотной по-

следовательности составляют сериновые остатки.

Соответственно кодируемый полипептид, образующий случайный клубок, может содержать аминокислотную последовательность, состоящую из примерно 35% пролиновых остатков, примерно 50% аланиновых остатков и примерно 15% сериновых остатков. Примеры нуклеотидных последовательностей и кодируемых ими полипептидов приведены в табл. 1. При использовании здесь термин "примерно X%" не ограничен точным числом процентов, но также включает значения с 10-20% дополнительными остатками или с на 10-20% меньшим количеством остатков. Например, термин 10% может также относиться к 11 или 12% или к 9 и 8% соответственно.

Тем не менее, как указано выше и описано более подробно ниже, указанный кодируемый полипептид, образующий случайный клубок, и, в частности, аминокислотная последовательность могут также содержать дополнительные аминокислоты, отличные от пролина, аланина и возможно серина, в качестве минорных компонентов. Как уже обсуждено выше, содержание указанного(ых) минорного(ых) компонента(ов), то есть аминокислот(ы), отличной(ых) от пролина, аланина или возможно серина, в кодируемом полипептиде, образующем случайный клубок, по настоящему изобретению, может составлять менее чем примерно 10% или менее чем примерно 5%.

Специалисту в данной области известно, что кодируемая аминокислотная последовательность/полипептид может также образовывать конформацию в виде случайного клубка, когда указанная аминокислотная последовательность/полипептид (фрагмент полипептида) содержит другие остатки, отличные от пролина, аланина и возможно серина в качестве минорных компонентов. При использовании здесь термин "минорный компонент" означает, что в кодируемых полипептидах, образующих случайный клубок, по настоящему изобретению от пролина, аланина или серина отличаются максимум 5% или максимум 10% аминокислотных остатков. Это означает, что от пролина, аланина и возможно серина могут отличаться максимум 10 из 100 аминокислот; предпочтительно от пролина, аланина и возможно серина могут отличаться максимум 8%, то есть максимум 8 из 100 аминокислот; более предпочтительно от пролина, аланина и возможно серина могут отличаться максимум 6%, то есть максимум 6 из 100 аминокислот; еще предпочтительнее от пролина, аланина и возможно серина могут отличаться максимум 5%, то есть максимум 5 из 100 аминокислот; особенно предпочтительно от пролина, аланина и возможно серина могут отличаться максимум 4%, то есть максимум 4 из 100 аминокислот; еще предпочтительнее от пролина, аланина и возможно серина могут отличаться максимум 3%, то есть максимум 3 из 100 аминокислот; еще предпочтительнее от пролина, аланина и возможно серина могут отличаться максимум 2%, то есть максимум 2 из 100 аминокислот; и наиболее предпочтительно от пролина, аланина и возможно серина может отличаться максимум 1%, то есть максимум 1 из 100 аминокислот, входящих в состав полипептида, образующего случайный клубок. Указанные аминокислоты, отличные от пролина, аланина и возможно серина, могут быть выбраны из группы, состоящей из Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr и Val, включая посттрансляционно модифицированные аминокислоты или искусственные аминокислоты (см., например, Budisa (2004) *Angew Chem Int Ed Engl* 43:6426-6463; Young (2010) *J Biol Chem* 285:11039-11044; Liu (2010) *Annu Rev Biochem* 79: 413-444; Wagner (1983) *Angew Chem Int Ed Engl* 22:816-828; Walsh (2010) *Drug Discov Today* 15: 773-780). В определенных случаях последовательности с высоким содержанием РА могут также содержать Ser в качестве минорного компонента. Например, в случае, если кодируемый полипептид, образующий случайный клубок, состоит из пролина и аланина, в качестве минорного компонента можно также рассматривать серин.

В целом, здесь предпочтительно, чтобы в кодируемом полипептиде, образующем случайный клубок, описанном здесь, или кодируемом полипептиде, образующем случайный клубок, являющемся частью/фрагментом слитого белка, не было этих "минорных" аминокислот (отличных от пролина, аланина и возможно серина). Согласно изобретению кодируемая аминокислотная последовательность/полипептид, образующий случайный клубок, могут, в частности, состоять исключительно из пролина, аланина и возможно серина (то есть в кодируемом полипептиде, образующем случайный клубок, или в аминокислотной последовательности нет других аминокислотных остатков).

В контексте настоящего изобретения молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, или нуклеотидная последовательность, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, может состоять из по меньшей мере 300 нуклеотидов. Тем не менее, специалисту в данной области очевидно, что длина нуклеотидной последовательности, предложенной здесь, не ограничена, при условии, что кодируемый полипептид образует случайный клубок. В прилагаемых примерах подтверждено, что неожиданным образом молекулы нуклеиновой кислоты, предложенные здесь, содержащие нуклеотидные последовательности с низкой повторяемостью, могут быть синтезированы независимо от их длины. Например, доказано, что могут быть синтезированы нуклеотидные последовательности, имеющие длину примерно 600 нуклеотидов. Такие последовательности могут быть использованы для сборки нуклеотидных последовательностей еще большей длины. Иными словами, эти нуклеотидные последовательности могут быть использованы как элементы/модули/структурные блоки для составления/сборки нуклеотидных последовательностей большей длины, входящих в состав молекулы нуклеиновой кислоты по изобретению. В одном воплощении один элемент/модуль/структурный блок нуклеотидной последовательности также соответствует молекуле

нуклеиновой кислоты по изобретению.

Согласно настоящему изобретению идентичные или неидентичные элементы/модули/структурные блоки нуклеотидной последовательности можно сочетать друг с другом при условии, что собранная нуклеотидная последовательность кодирует полипептид, состоящий из пролина, аланина и возможно серина. Кроме того, согласно настоящему изобретению идентичные или неидентичные элементы/модули/структурные блоки нуклеотидной последовательности можно сочетать друг с другом при условии, что собранная нуклеотидная последовательность кодирует полипептид с повторяющейся аминокислотной последовательностью, образующей случайный клубок. Как указано выше, сборка молекул нуклеиновой кислоты, предложенных здесь, содержащих нуклеотидные последовательности с низкой повторяемостью, с использованием этих элементов/модулей/структурных блоков особенно предпочтительна. Ниже подтверждена возможность сборки длинных нуклеотидных последовательностей длиной по меньшей мере 2400 нуклеотидов. Нуклеотидные последовательности по изобретению можно сочетать друг с другом или с дополнительными неидентичными нуклеотидными последовательностями. Таким образом, молекула нуклеиновой кислоты, предложенная здесь, может быть собрана из идентичных или неидентичных нуклеотидных последовательностей, представляющих собой нуклеотидные последовательности с низкой повторяемостью. В определенных аспектах молекула нуклеиновой кислоты, предложенная здесь, содержит по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 192, SEQ ID NO: 193, SEQ ID NO: 194 и SEQ ID NO: 195. Поскольку эти типичные нуклеотидные последовательности могут быть использованы для сборки нуклеотидных последовательностей большей длины, они могут быть названы элементами, модулями или структурными блоками. Таким образом, здесь подразумевают, что молекулы нуклеиновой кислоты по изобретению могут содержать множество этих нуклеотидных модулей или нуклеотидных последовательностей, собранных в нуклеотидную последовательность большей длины, где сама указанная нуклеотидная последовательность большей длины представляет собой нуклеотидную последовательность с низкой повторяемостью, как описано выше. Специалисту в данной области ясно, что молекула нуклеиновой кислоты по изобретению может также содержать фрагменты данных модулей нуклеотидной последовательности. Иными словами, молекула нуклеиновой кислоты, предложенная здесь, содержит или представляет собой по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 192, SEQ ID NO: 193, SEQ ID NO: 194 и SEQ ID NO: 195, или по меньшей мере один фрагмент этих последовательностей.

Здесь подразумевают, что нуклеотидные последовательности с низкой повторяемостью или элементы, модули или структурные блоки, предложенные здесь, можно менять местами или сочетать друг с другом в любой комбинации при условии, что собранная нуклеотидная последовательность содержит нуклеотидную последовательность с низкой повторяемостью по настоящему изобретению. Примеры собранных нуклеотидных последовательностей приведены ниже, например SEQ ID NO: 38, обозначенная

здесь как PAS#1f/1c/1b(600), SEQ ID NO: 39, обозначенная здесь как PAS#1d/1f/1c/1b(800), SEQ ID NO: 40, обозначенная здесь как PAS#1h/1e/1i(600), SEQ ID NO: 41, обозначенная здесь как PAS#1j/1h/1e/1i(800), SEQ ID NO: 42, обозначенная здесь как PA#1d/1c/1b(600), SEQ ID NO: 43, обозначенная здесь как PA#1i/1h/1g/1f(800), SEQ ID NO: 44, обозначенная здесь как PA#1e/1d/1c/1b(800), SEQ ID NO: 45, обозначенная здесь как PA#1i/1h/1g/1f/1e/1d/1c/1b(1600), SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 173; SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190 и/или SEQ ID NO: 191.

Таким образом, в предпочтительных аспектах изобретения молекула нуклеиновой кислоты, предложенная здесь, имеет, содержит или представляет собой указанную собранную нуклеотидную последовательность.

Как подтверждено в прилагаемых примерах (см. пример 1), длинная нуклеотидная последовательность по изобретению может быть собрана постадийно. Собранная нуклеотидная последовательность имеет низкую повторяемость. В прилагаемых примерах показана сборка типичной нуклеотидной последовательности, представленной в SEQ ID NO: 39 или обозначенной здесь как PAS#1d/1f/1c/1b(800), содержащей 2400 нуклеотидов и кодирующей 800 аминокислот.

Здесь подразумевают, что эти собранные нуклеотидные последовательности представляют собой нуклеотидные последовательности с низкой повторяемостью. Например, в прилагаемых примерах подтверждено, что типичная нуклеотидная последовательность, представленная в SEQ ID NO: 39 или обозначенная здесь как последовательность PAS#1d/1f/1c/1b(800), не имеет повторов при окне повтора 15 или имеет только один повтор длиной 14 нуклеотидов на протяжении всей нуклеотидной последовательности длиной 2400 нуклеотидов (см. фиг. 2B). Для сравнения, длинная нуклеотидная последовательность, раскрытая в предшествующем уровне техники, содержит повторяющиеся нуклеотидные последовательности, что продемонстрировано на примере PAS#1a(600), представленной здесь в SEQ ID NO: 12. Соответственно длинные молекулы нуклеиновой кислоты по настоящему изобретению имеют нуклеотидные последовательности с низкой повторяемостью и, таким образом, позволяют преодолеть технические трудности, ассоциированные с повторяющимися участками нуклеотидов, как указано выше.

Молекулы нуклеиновой кислоты и родственные им молекулы нуклеиновой кислоты (такие как варианты, фрагменты, молекулы нуклеиновой кислоты, по меньшей мере на 66%, например по меньшей мере на 66,6%, идентичные определенным нуклеотидным последовательностям, кодирующим полипептид, состоящий из пролина и аланина, или аналогичные варианты, фрагменты, молекулы нуклеиновой кислоты, по меньшей мере на 56%, например по меньшей мере на 56,6%, идентичные определенным нуклеотидным последовательностям, кодирующим полипептид, состоящий из пролина, аланина и серина, предложенным и определенным здесь, и тому подобное) содержат или представляют собой нуклеотидные последовательности с низкой повторяемостью, кодирующие полипептид, образующий конформацию в виде случайного клубка, что повышает стабильность *in vivo/in vitro*.

Молекулы нуклеиновой кислоты и родственные им молекулы нуклеиновой кислоты имеют, содержат или представляют собой нуклеотидные последовательности с низкой повторяемостью, кодирующие полипептиды, образующие конформацию в виде случайного клубка и повышающие стабильность биологически или фармакологически активного белка или лекарственного средства *in vivo/in vitro*. Указанные родственные молекулы нуклеиновой кислоты содержат или представляют собой варианты и фрагменты указанных молекул нуклеиновой кислоты. Указанные родственные молекулы нуклеиновой кислоты по меньшей мере на 66%, например по меньшей мере на 66,6%, идентичны определенной нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина и аланина, или по меньшей мере на 56%, например по меньшей мере на 56,6%, идентичны определенной нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и серина, и так далее.

В определенных аспектах молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина и аланина, по меньшей мере на 66%, например по меньшей мере на 66,6%, гомологична/идентична нуклеотидной последовательности, показанной, среди прочего, в SEQ ID NO: 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 42, 43, 44, 45, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 192 и 193. Более предпочтительно молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина и аланина, по меньшей мере на 68, 70, 75, 80, 85, 90, 92, 93, 94, 95, 96, 97 или 98% гомологична/идентична нуклеотидной последовательности, показанной, среди прочего, в 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 42, 43, 44, 45, 87, 88, 89, 90, 91, 92, 93, 94,

95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 192 и 193, при этом предпочтительны большие значения. Наиболее предпочтительно, молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина и аланина, по меньшей мере на 99% гомологична/идентична нуклеотидной последовательности, показанной, среди прочего, в SEQ ID NO: 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 42, 43, 44, 45, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 192 и 193.

В других аспектах молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, по меньшей мере на 56%, например по меньшей мере на 56,6%, гомологична/идентична/сходна с нуклеотидной последовательностью, показанной, среди прочего, в SEQ ID NO: 19, 20, 21, 22, 23, 24, 25, 26, 27, 38, 39, 40, 41, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 174, 175, 176, 177, 178, 179, 180, 181, 182, 184, 185, 186, 187, 188, 189, 190, 191, 194 и 195. Более предпочтительно, молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, по меньшей мере на 58, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97 или 98% гомологична/идентична нуклеотидной последовательности, показанной, среди прочего, в 19, 20, 21, 22, 23, 24, 25, 26, 27, 38, 39, 40, 41, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 174, 175, 176, 177, 178, 179, 180, 181, 182, 184, 185, 186, 187, 188, 189, 190, 191, 194 и 195. Наиболее предпочтительно, молекула нуклеиновой кислоты, содержащая ортологичную/гомологичную/идентичную/сходную (и, таким образом, родственную) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, по меньшей мере на 99% гомологична/идентична/сходна с нуклеотидной последовательностью, показанной, среди прочего, в SEQ ID NO: 19, 20, 21, 22, 23, 24, 25, 26, 27, 38, 39, 40, 41, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 174, 175, 176, 177, 178, 179, 180, 181, 182, 184, 185, 186, 187, 188, 189, 190, 191, 194 и 195. Определенные выше родственные нуклеотидные последовательности могут также быть включены в изоформы большей или меньшей длины, сплайсированные варианты или слитые конструкции.

В определенных аспектах молекула нуклеиновой кислоты, предложенная здесь, может гибридизоваться в жестких условиях с комплементарной цепью нуклеотидной последовательности, показанной, среди прочего, в SEQ ID NO: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194 и 195. При использовании здесь термин "гибридизация" или "гибридуется" может относиться к гибридизации в жестких или нежестких условиях. Если не указано дополнительно, условия предпочтительно являются жесткими. Такие условия гибридизации можно определить, следуя общепринятым протоколам, описанным, например, в Sambrook (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 3rd ed., New York; Ausubel (1989) *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, New York; или Higgins (1985) *Nucleic acid hybridization, a practical approach*, IRL Press Oxford, Washington DC. Установление подходящих условий входит в компетенцию специалиста в данной области, и они могут быть определены по протоколам, описанным в данной области. Так, для определения только специфично гибридизующихся последовательностей обычно будут необходимы жесткие условия гибридизации и промывки, такие как, например, 0,1× цитратно-солевой буфер (saline sodium citrate buffer, SSC), 0,1% мас./об. SDS при 65°C или 2×SSC, 60°C, 0,1% мас./об. SDS. Условия гибридизации низкой жесткости для выявления гомологичных или неточно комплементарных последовательностей могут, например, быть установлены как 6×SSC, 1% мас./об. SDS при 65°C. Хорошо известно, что длина нуклеиновой кислоты, используемой в качестве зонда, и состав нуклеиновой кислоты, подлежащей определению, также являются параметрами условий гибридизации.

Согласно настоящему изобретению термин "гомология", или "процент гомологии", или "идентичность", или "процент идентичности", или "идентичность последовательности" в контексте двух или более нуклеотидных последовательностей относится к двум или более последовательностям или подпоследовательностям, являющимся одинаковыми или имеющими определенный процент нуклеотидов, являющихся одинаковыми (предпочтительно, идентичность составляет по меньшей мере 66%, например по

меньшей мере 66,6% в случае молекулы нуклеиновой кислоты, кодирующей полипептид, состоящий из пролина и аланина, более предпочтительно по меньшей мере 68, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97 или 98%, наиболее предпочтительно по меньшей мере 99%; или предпочтительно по меньшей мере 56%, например по меньшей мере 56,6% в случае молекулы нуклеиновой кислоты, кодирующей пролин, аланин и серин, более предпочтительно по меньшей мере 58, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97 или 98%, наиболее предпочтительно по меньшей мере 99%) при их сравнении и выравнивании для максимального соответствия в окне сравнения (предпочтительно по всей длине) или в обозначенной области, как определено с применением алгоритма сравнения последовательностей, известного в данной области, или посредством выравнивания вручную и визуальной проверки.

Последовательности, идентичные, например, на 70-90% или более, можно рассматривать как по существу идентичные или сходные. Такое определение также применимо к комплементу анализируемой последовательности нуклеиновой кислоты. Предпочтительно описанная идентичность присутствует в области длиной по меньшей мере примерно 15-25 нуклеотидов, более предпочтительно в области длиной по меньшей мере примерно 50-100 нуклеотидов, более предпочтительно в области длиной по меньшей мере примерно 125-200 нуклеотидов, более предпочтительно в области длиной по меньшей мере примерно 225-300 нуклеотидов, еще более предпочтительно в области длиной по меньшей мере примерно 325-600 нуклеотидов, еще более предпочтительно в области длиной по меньшей мере примерно 625-800 нуклеотидов и наиболее предпочтительно в области длиной по меньшей мере примерно 825-1200 нуклеотидов. Специалистам в данной области будет известно, как определить процент идентичности последовательностей с применением, например, таких алгоритмов, как алгоритмы на основе компьютерных программ CLUSTALW (Thompson (1994) *Nucl. Acids Res.* 2: 4673-4680), CLUSTAL Omega (Sievers (2014) *Curr. Protoc. Bioinformatics* 48:3.13.1-3.13.16) или FASTDB (Brutlag (1990) *Comp App Biosci* 6: 237-245). Также специалистам в данной области доступны алгоритмы BLAST, что расшифровывается как "Basic Local Alignment Search Tool" ("основной инструмент поиска локального выравнивания"), и BLAST 2.0 (Altschul, (1997) *Nucl. Acids Res.* 25:3389-3402; Altschul (1990) *J. Mol. Biol.* 215:403-410). В программе BLASTN для нуклеиновых кислот по умолчанию используется длина слова (W) 11, ожидание (E) 10, M, равное 5, N, равное 4, и сравнение обеих цепей. В матрице значений BLOSUM62 (Henikoff (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:10915-10919) используются выравнивания (B) 50, ожидание (E) 10, M, равное 5, N, равное 4, и сравнение обеих цепей.

Для определения того, соответствует ли нуклеотидный остаток данной нуклеотидной последовательности определенному положению в другой нуклеотидной последовательности, например SEQ ID NO: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194 и 195, соответственно, специалист в данной области может применить средства и способы, хорошо известные в данной области, например выравнивание вручную или с применением компьютерных программ, таких как указанные здесь. Например, для поиска локальных выравниваний последовательностей можно применять BLAST 2.0. BLAST или BLAST 2.0, обсуждаемые выше, проводят выравнивание нуклеотидных последовательностей для определения сходства последовательностей. Ввиду локальной природы выравниваний BLAST или BLAST 2.0 особенно полезны при определении точных совпадений или при определении сходных или идентичных последовательностей.

Как показано в прилагаемых примерах, здесь предложены средства и способы быстрого и легкого клонирования длинных нуклеотидных последовательностей с низкой повторяемостью или длинных молекул нуклеиновой кислоты с низкой повторяемостью, содержащих указанные нуклеотидные последовательности, см., например, примеры 1, 2, 5, 7 и 10. Как продемонстрировано в прилагаемых примерах, здесь предложены типичные векторы, особенно подходящие для сборки нуклеотидных последовательностей по изобретению в конструкции большей длины. Как указано выше, из молекул нуклеиновой кислоты или нуклеотидных последовательностей, предложенных здесь, постадийно могут быть собраны длинные молекулы нуклеиновой кислоты или длинные нуклеотидные последовательности. Здесь предполагают, что, в лучшем случае, размер собранной молекулы нуклеиновой кислоты или нуклеотидной последовательности не ограничен или ограничен практическими обстоятельствами, такими как размер плазмиды или эффективность трансформации/трансфекции.

Типичная сборка молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность или молекулу нуклеиновой кислоты с низкой повторяемостью, кодирующую повторяющуюся аминокислотную последовательность из пролина, аланина и серина, длиной примерно 2400 нуклеотидов (кодирующей полипептид длиной 800 аминокислотных остатков), продемонстрирована в прилагаемом Примере 2. Там показано, что в типичном векторе rXL2 проводили клонирование и сборку сначала элемента (структурного блока) нуклеотидной последовательности PAS#1b(200), представленного в SEQ ID NO: 19, затем элемента нуклеотидной последовательности PAS#1c(200), представленного в SEQ ID NO: 20, затем

элемента нуклеотидной последовательности PAS#1f(200), представленного SEQ ID NO: 23, и, в завершение, элемента нуклеотидной последовательности PAS#1d(200), представленного в SEQ ID NO: 21. Эта типичная собранная молекула нуклеиновой кислоты или собранная нуклеотидная последовательность обозначена здесь как PAS#1d/1f/1c/1b(800) и представлена в SEQ ID NO: 39. Кроме того, здесь доказано, что эти собранные молекулы нуклеиновой кислоты по изобретению имеют или содержат нуклеотидные последовательности с низкой повторяемостью. В частности, в примере 2 показано, что эта типичная собранная нуклеотидная последовательность, представленная в SEQ ID NO: 39, имеет только повторяющиеся последовательности максимальной длиной 14 нуклеотидов.

Изобретение также относится к способу обеспечения молекул нуклеиновой кислоты по изобретению; см., например, фиг. 1E, где показан типичный способ сборки нуклеотидных последовательностей большей длины. Кроме того, изобретение относится к вектору, особенно подходящему для сборки молекул нуклеиновой кислоты с низкой повторяемостью, кодирующих полипептиды с высоким содержанием РА.

В соответствии с описанным выше, молекула нуклеиновой кислоты, предложенная здесь, может кодировать полипептид, состоящий из пролина и аланина.

Соответственно, молекула нуклеиновой кислоты, предложенная здесь, может быть выбрана из группы, состоящей из:

(a) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 192 и SEQ ID NO: 193;

(b) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, состоящую из SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172 и/или SEQ ID NO: 173;

(c) молекулы нуклеиновой кислоты, гибридизирующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (a) или (b);

(d) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 66,7% идентичную нуклеотидной последовательности, определенной в любом из (a), (b) и (c); и

(g) молекулы нуклеиновой кислоты, являющейся вырожденной в силу генетического кода по отношению к нуклеотидной последовательности, определенной в (a) или (b).

В соответствии с описанным выше молекула нуклеиновой кислоты, предложенная здесь, может кодировать полипептид, состоящий из пролина, аланина и серина.

Соответственно, молекула нуклеиновой кислоты, предложенная здесь, может быть выбрана из группы, состоящей из:

(a) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 194 и SEQ ID NO: 195;

(b) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190 и SEQ ID NO: 191;

(c) молекулы нуклеиновой кислоты, гибридизирующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (a) или (b);

(d) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 56% идентичную нуклеотидной последовательности, определенной в любом из (a), (b) и (c);

(e) молекулы нуклеиновой кислоты, являющейся вырожденной в силу генетического кода по отно-

шению к нуклеотидной последовательности, определенной в (a) или (b).

В определенных аспектах изобретение относится к способу получения молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность с низкой повторяемостью по изобретению, включающему:

(a) обеспечение вектора, содержащего верхнюю последовательность распознавания, распознаваемую ферментом эндонуклеазой рестрикции, и нижнюю последовательность распознавания, распознаваемую ферментом эндонуклеазой рестрикции,

где возможно указанный фермент эндонуклеаза рестрикции, распознающий указанную нижнюю последовательность распознавания, отличается от фермента эндонуклеазы рестрикции, распознающего верхнюю последовательность распознавания,

где указанная верхняя последовательность распознавания и указанная нижняя последовательность распознавания расположены в обратно-комплементарной ориентации,

где указанная верхняя последовательность распознавания содержит две последовательности распознавания для двух разных рестриктаз,

где указанная нижняя последовательность распознавания содержится в верхней последовательности распознавания, и/или

где указанная верхняя последовательность распознавания и/или указанная нижняя последовательность распознавания представляют собой сайты распознавания для рестриктаз, расщепляющих за пределами последовательности распознавания;

(b) расщепление указанного вектора стадии (a) рестриктазой(ами), распознающей(ими) указанную верхнюю и/или указанную нижнюю последовательность распознавания;

(c) возможно дефосфорилирование указанного вектора стадии (b) для предотвращения гибридизации комплементарных "липких" концов;

(d) обеспечение молекулы нуклеиновой кислоты, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где концы нуклеотидной последовательности гибридизуются с расщепленными концами вектора стадии (b) или (c); и

(e) введение указанной молекулы нуклеиновой кислоты стадии (d) в указанный расщепленный вектор в присутствии ДНК-лигазы.

В определенных аспектах изобретения способ, предложенный здесь, может быть применен для получения молекул нуклеиновой кислоты большей длины, где молекула нуклеиновой кислоты содержит собранную нуклеотидную последовательность с низкой повторяемостью, где способ получения молекулы нуклеиновой кислоты, описанный выше, дополнительно включает:

(f) расщепление указанного вектора стадии (e) рестриктазой, распознающей указанную верхнюю или указанную нижнюю последовательность распознавания;

(g) возможно дефосфорилирование указанного вектора стадии (f) для предотвращения гибридизации комплементарных "липких" концов;

(h) обеспечение молекулы нуклеиновой кислоты, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, где концы нуклеотидной последовательности гибридизуются с расщепленными концами вектора стадии (f) или (g);

(i) введение указанной нуклеотидной последовательности стадии (h) в указанный расщепленный вектор в присутствии ДНК-лигазы;

(j) повторение стадий (f)-(i) до достижения желаемой длины нуклеотидной последовательности.

Специалисту в данной области ясно, что стадии этого способа можно многократно повторять до достижения желаемой длины собранной нуклеотидной последовательности или собранной молекулы нуклеиновой кислоты, кодирующей полипептид с высоким содержанием РА.

Для обеспечения средств и способов для предпочтительного клонирования молекул нуклеиновой кислоты, предложенных здесь, подходящий вектор для клонирования содержит верхнюю последовательность распознавания, распознаваемую эндонуклеазой рестрикции, и нижнюю последовательность распознавания, распознаваемую эндонуклеазой рестрикции, где возможно указанная эндонуклеаза рестрикции, распознающая указанную нижнюю последовательность распознавания, отличается от указанной эндонуклеазы рестрикции, распознающей указанную верхнюю последовательность распознавания, где указанная верхняя последовательность распознавания и указанная нижняя последовательность распознавания расположены в обратно-комплементарной ориентации. Кроме того, указанная верхняя последовательность распознавания может содержать две последовательности распознавания для двух разных рестриктаз. Кроме того, указанная нижняя последовательность распознавания может содержаться в верхней последовательности распознавания. Таким образом, указанная нижняя последовательность распознавания может содержаться в одной из верхних последовательностей распознавания. В прилагаемых иллюстративных примерах, в частности, см. фиг. 1, 4, 6 и 8, показаны типичные векторы и молекулы нуклеиновой кислоты, а также способы их обеспечения.

В прилагаемых примерах подтверждена особая предпочтительность того, чтобы нижняя последовательность распознавания содержалась в верхней последовательности распознавания. Использование такой стратегии позволяет использовать верхний сайт распознавания для присоединения последующих

нуклеотидных последовательностей к вектору, предложенному здесь. Разумеется, положения верхнего и нижнего сайтов рестрикции в таком векторе взаимозаменяемы. Типичная область для клонирования с введенной нуклеотидной последовательностью по изобретению показана на фиг. 1D. В таком случае рестриктаза, распознающая нижнюю последовательность распознавания, также распознает и расщепляет верхнюю последовательность распознавания. Таким образом, нуклеотидная последовательность или молекула нуклеиновой кислоты по изобретению, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, может быть вырезана из вектора, предложенного здесь, с использованием рестриктазы, распознающей нижнюю и верхнюю последовательность распознавания, что позволяет использовать ее для лигирования с другими нуклеиновыми кислотами или векторами, например, чтобы обеспечить экспрессию генов или создать кодирующую область для слитого белка. Ниже показан пример такого способа клонирования для сборки длинной нуклеотидной последовательности с низкой повторяемостью; см., например, фиг. 1E.

В определенных аспектах первый "липкий" конец может содержать нуклеотидный триплет/кодон, кодирующий аланин, в частности GCC, а также GCT, GCA или GCG. В предпочтительном аспекте молекула нуклеиновой кислоты по изобретению содержит два комплементарных "липких" 5'-конца, то есть один "липкий" 5'-конец расположен на кодирующей цепи, а другой "липкий" 5'-конец расположен на некодирующей цепи. В особенно предпочтительном воплощении "липкий" 5'-конец кодирующей цепи представляет собой GCC, а "липкий" 5'-конец некодирующей цепи представляет собой GGC.

В определенных других аспектах первый "липкий" конец может также содержать нуклеотидные триплеты/кодоны, кодирующие пролин или серин, например CCT, CCC, CCA, CCG, TCT, TCC, TCA, TCG, AGT или AGC. Тем не менее, специалисту в данной области ясно, что способ получения молекулы нуклеиновой кислоты, предложенный здесь, не ограничен "липкими" концами, а молекулу нуклеиновой кислоты или нуклеотидную последовательность можно, например, также лигировать через тупые концы. При использовании здесь термин "липкий" конец относится к концевой части цепи нуклеиновой кислоты как части двуцепочечной молекулы ДНК без присоединенной комплементарной цепи. При использовании здесь термин "тупой конец" относится к концевой части цепи ДНК без "липкого" конца. Предполагают, что длина "липкого" конца не ограничена; тем не менее, "липкий" конец, содержащий от 1 до 10 нуклеотидов, по-видимому, является наиболее подходящим. В прилагаемых примерах был использован "липкий" конец из 3 нуклеотидов, кодирующий аминокислоту аланин. Этот тип триплетного "липкого" конца обладает тем преимуществом, что он непосредственно совместим с рамкой считывания для аминокислотной трансляции молекулы нуклеиновой кислоты по настоящему изобретению.

Следует отметить, что этот "липкий" конец вводит дополнительный триплет в нуклеотидную последовательность или молекулу нуклеиновой кислоты, кодирующую полипептид, состоящий из пролина, аланина и возможно серина. Соответственно, способ клонирования по изобретению вводит одну дополнительную аминокислоту, например аланин. Эту дополнительную аминокислоту или соответствующий триплет/кодон можно рассматривать как часть полипептида, состоящего из пролина, аланина и возможно серина, или кодирующей его нуклеиновой кислоты, соответственно. Следовательно, как будет ясно специалисту в данной области, при применении способа клонирования, описанного выше, для нуклеотидной последовательности, содержащей, например, 300 нуклеотидов, вырезанная молекула нуклеиновой кислоты будет содержать 303 нуклеотида, если также учитывать оба "липких" 5'-конца из трех нуклеотидов. Здесь предполагают, что, в зависимости от длины "липких" концов, используемых здесь, данный способ клонирования может приводить к включению еще большего числа триплетов/кодонов или, соответственно, большего числа аминокислотных остатков.

Кроме того, при клонировании на векторе или в виде части синтетического фрагмента (например, цепи) ДНК большей длины молекула нуклеиновой кислоты или нуклеотидная последовательность, предложенная здесь, содержит указанную верхнюю последовательность распознавания и/или указанную нижнюю последовательность распознавания, где указанная верхняя последовательность распознавания и/или указанная нижняя последовательность распознавания представляют собой сайты распознавания для рестриктаз, расщепляющих за пределами последовательности распознавания. Рестриктазы, используемые здесь, например, рестриктазы Типа IIS, предпочтительно расщепляют за пределами их последовательностей распознавания с одной стороны, результатом чего является расщепленная молекула нуклеиновой кислоты или нуклеотидная последовательность, не содержащая сайта(ов) распознавания, в зависимости от правильной ориентации асимметричной последовательности распознавания.

Кроме того, в предложенном здесь векторе расщепление такими рестриктазами сохраняет последовательности распознавания для рестриктаз. Сайт клонирования типичного вектора рXL2 показан на фиг. 1C. Молекула нуклеиновой кислоты или нуклеотидная последовательность, вырезанная из этого вектора, не имеет сайтов распознавания для рестриктаз, используемых для ее клонирования или вырезания, что особенно предпочтительно для сборки молекул нуклеиновой кислоты большей длины или нуклеотидных последовательностей большей длины, например способом, описанным здесь, для клонирования на вектор экспрессии или для создания кодирующей области для слитого белка.

В другом воплощении предложен типичный вектор рXL1 (SEQ ID NO: 55) (показанный на фиг. 1B). В данном случае клонированная/введенная молекула нуклеиновой кислоты или нуклеотидная последова-

тельность по настоящему изобретению фланкирована двумя сайтами рестрикции SapI. Таким образом, молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, может быть просто вырезана из этого вектора посредством обработки/расщепления одной рестриктазой, то есть SapI (или также EarI), что особенно предпочтительно для последующего клонирования на вектор экспрессии и/или для создания кодирующей области для слитого белка.

Как показано в прилагаемых примерах, молекулы нуклеиновой кислоты или нуклеотидные последовательности, предложенные здесь, собирают в вектор pXL2 (SEQ ID NO: 48) повторяющимся образом; см., например, пример 2. Поэтому один или оба сайта распознавания, присутствующие в данном векторе, можно использовать для введения дополнительной нуклеотидной последовательности в вектор (с одной стороны существующей вставки) или, альтернативно, для вырезания всей (собранный) нуклеотидной последовательности или молекулы нуклеиновой кислоты из вектора.

В качестве дополнительного преимущества в способе получения молекулы нуклеиновой кислоты по изобретению, предложенной здесь, отсутствие сайтов распознавания для рестриктаз в молекуле нуклеиновой кислоты по настоящему изобретению (1) предотвращает внутреннее расщепление собранной нуклеотидной последовательности большей длины, предложенной здесь, например между элементами/модулями собранных нуклеотидных последовательностей, и (2) приводит к непрерывному переходу между аминокислотными последовательностями, кодируемыми соседними элементами/модулями собранных нуклеотидных последовательностей или между аминокислотными последовательностями, кодируемыми нуклеотидной последовательностью по изобретению и нуклеотидной последовательностью биологически активного белка. Эта методика может быть названа "бесследовая" или "бесшовная" сборка или клонирование.

В прилагаемых примерах продемонстрирована возможность использования рестриктаз Типа IIS; см. примеры 1 и 2. Сайты распознавания рестриктаз этого класса отделены от их сайтов расщепления, и некоторые из них, например SapI и EarI, расщепляют за пределами своей последовательности распознавания с одной стороны. Ниже показано, что указанную верхнюю последовательность распознавания на pXL2 распознают SapI и EarI, а указанную нижнюю последовательность распознавания распознает EarI. Таким образом, указанная верхняя последовательность распознавания имеет нуклеотидную последовательность "5'-GCTCTTC-3'", а указанная нижняя последовательность распознавания имеет нуклеотидную последовательность "5'-CTCTTC-3'". Специалисту в данной области ясно, что особых ограничений по рестриктазам здесь нет, при условии, что они соответствуют задаче или ограничениям, определенным здесь.

При использовании здесь термин "фермент эндонуклеаза рестрикции" или "рестриктаза" относится к ферменту эндонуклеазе, разрезающему/расщепляющему/гидролизующему олиго/полинуклеотиды по фосфодиэфирной связи, соединяющему определенные нуклеотиды в молекуле нуклеиновой кислоты или нуклеотидной последовательности, например ДНК, в или рядом с определенной последовательностью распознавания, содержащей некоторый нуклеотидный участок. Таким образом, рестриктазы катализируют гидролиз фосфодиэфирных связей в нуклеотидной последовательности или молекуле нуклеиновой кислоты. Рестриктазы обычно классифицируют на три типа, различающиеся по своей структуре и по разрезанию/расщеплению их субстрата в их последовательности распознавания или в отдельном сайте. Для разрезания/расщепления двуцепочечной ДНК рестриктазы обычно расщепляют две фосфодиэфирные связи, по одной в каждом сахарофосфатном остове (то есть в каждой цепи) двойной спирали ДНК.

При использовании здесь термин "последовательность распознавания" относится к определенной последовательности нуклеотидов, например длиной от 4 до 8 определенных пар оснований, распознаваемой рестриктазой.

При использовании здесь термин "расщепление" означает, что рестриктаза разрезает/расщепляет/гидролизует молекулу нуклеиновой кислоты и/или вектор. Как указано выше, рестриктаза расщепляет фосфодиэфирную связь в полинуклеотидной цепи.

При использовании здесь термин "введение" или "вставка" относится к лигированию молекулы нуклеиновой кислоты в вектор под действием фермента. При этом концы полинуклеотидов соединяются друг с другом посредством образования фосфодиэфирных связей между 3'-гидроксильной группой на конце одного полинуклеотида и 5'-фосфорильной группой другого. Концы нуклеиновой кислоты, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, или нуклеотидной последовательности могут гибридизоваться с расщепленными концами вектора. В предпочтительных аспектах изобретения такие концы представляют собой "липкие" концы, способные гибридизоваться с соответствующими "липкими" концами расщепленного вектора.

Для введения молекулы нуклеиновой кислоты в вектор его предпочтительно дефосфорилируют во избежание большого количества рециркуляризованной векторной ДНК без вставки. Типичным ферментом для дефосфорилирования может быть кишечная щелочная фосфатаза телят (CIP или CIAP) или щелочная фосфатаза креветки, обе из которых удаляют фосфатную группу с 5'-концов расщепленных полинуклеотидов.

При использовании здесь термины "верхний" и "нижний" относятся к относительному положению в молекуле нуклеиновой кислоты или в нуклеотидной последовательности. Каждая цепь молекулы нук-

леиновой кислоты или нуклеотидной последовательности имеет 5'-конец и 3'-конец, названные так по атомам углерода в дезоксирибозном (или рибозном) сахаре. Обычно верх и низ относятся к 5'-3'-направлению кодирующей цепи, в котором происходит транскрипция с образованием РНК. В случае двуцепочечной ДНК верх расположен ближе к 5'-концу кодирующей цепи рассматриваемого гена или рамки считывания, а низ ближе к 3'-концу. Ввиду антипараллельности двуцепочечной ДНК это означает, что 3'-конец некодирующей цепи расположен выше данного гена, а 5'-конец ниже.

Предполагают, что при использовании здесь термин "молекула нуклеиновой кислоты" или "нуклеотидная последовательность" включают молекулы нуклеиновой кислоты, такие как молекулы ДНК и молекулы РНК. Здесь подразумевают, что термин "нуклеотидная последовательность" эквивалентен термину "молекула нуклеиновой кислоты" и что эти термины могут быть использованы здесь взаимозаменяемо. Указанная молекула нуклеиновой кислоты или указанная нуклеотидная последовательность может быть одноцепочечной или двуцепочечной, но предпочтительно представляет собой двуцепочечную ДНК. Специалисту в данной области известно, что двуцепочечная ДНК фактически содержит две разные молекулы нуклеиновой кислоты, последовательности которых большей частью комплементарны (без учета "липких" концов, если они есть), и которые нековалентно связаны/гибридизованы друг с другом с образованием двойной цепи.

В одном аспекте изобретения нуклеотидная последовательность или молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, функционально связана, в одной и той же рамке считывания, с нуклеиновой кислотой, кодирующей биологически или фармакологически активный белок. В предпочтительных аспектах изобретения нуклеотидная последовательность или молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, функционально связана, в одной и той же рамке считывания, с нуклеиновой кислотой, кодирующей биологически или фармакологически активный белок. Таким образом, молекула нуклеиновой кислоты кодирует гетерологичный конъюгат лекарственного средства, содержащий полипептид, состоящий из пролина, аланина и возможно серина, и биологически или фармакологически активный белок. При использовании здесь "гетерологичная" означает, что молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, и биологически или фармакологически активный белок, не встречается в природе.

При использовании здесь термин "функционально связанный" относится к близкому расположению, где рассматриваемые компоненты связаны друг с другом таким образом, который позволяет каждому из них выполнять свою предполагаемую функцию.

Нуклеотидная последовательность, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, может быть конъюгирована с нуклеотидной последовательностью, кодирующей биологически или фармакологически активный белок, "бесшовным" образом, то есть без линкеров между этими двумя последовательностями. Альтернативно, между полипептидом, образующим случайный клубок, и биологически или фармакологически активным белком может быть расположена линкерная или спейсерная структура. Таким образом, в определенных аспектах изобретения между нуклеотидной последовательностью, кодирующей полипептид с высоким содержанием пролина/аланина, и нуклеотидной последовательностью, кодирующей биологически или фармакологически активный белок, может быть введена нуклеотидная последовательность, кодирующая аминокислотный линкер. Примерами линкеров могут быть сайт расщепления, чувствительный к протеазе, серин/глициновый линкер, аффинная метка, такая как His₆-метка или Strep-метка II, сигнальный пептид, ретенционный пептид (retention peptide), пептид направленного действия, такой как мембранный транслокационный пептид (membrane translocation peptide), или дополнительные эффекторные домены, например, фрагменты антител для направленного воздействия на опухоль, связанные с противоопухолевым токсином или ферментом для активации пролекарства, и так далее. Полипептид, содержащий линкер/спейсер, может иметь сайт расщепления протеазой плазмы, что сделает возможным контролируемое высвобождение указанного биологически активного белка. Можно без труда определить линкеры/спейсеры различных типов и длины для обеспечения оптимальной биологической активности конкретных белков.

Линкеры/спейсеры могут вызывать иммунный ответ у субъекта, получающего слитый белок, несущий линкерную или спейсерную структуру. Поэтому в предпочтительных аспектах изобретения нуклеотидная последовательность, кодирующая полипептид с высоким содержанием пролина/аланина, конъюгирована с нуклеотидной последовательностью, кодирующей биологически или фармакологически активный белок, "бесшовным" образом. При использовании здесь "бесшовный" означает, что нуклеотидная последовательность, кодирующая полипептид, образующий случайный клубок, конъюгирована с нуклеотидной последовательностью, кодирующей биологически или фармакологически активный белок, напрямую. Таким образом, не происходит введения дополнительных нуклеотидов, кодирующих аминокислотные остатки, отличные от пролина, аланина и возможно серина. Как продемонстрировано в прилагаемых примерах, "бесшовное" клонирование проводили с использованием "липких" концов, кодирующих аланиновый аминокислотный остаток; см., например, пример 7. Таким образом, настоящее изобретение также относится к способу получения молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую пролин, аланин и возможно серин, и нуклеотидную последователь-

ность, кодирующую биологически или фармакологически активный белок или пептид, и где указанная нуклеотидная последовательность конъюгирована с указанной нуклеотидной последовательностью, кодирующей биологически или фармакологически активный белок, "бесшовным" образом. В прилагаемых примерах и выше также продемонстрировано, что "бесшовное" клонирование может быть применено для сборки нуклеотидных последовательностей большей длины, кодирующих полипептиды с высоким содержанием РА. Таким образом, при применении метода "бесшовного" клонирования полученная нуклеотидная последовательность, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, представляет собой нуклеотидную последовательность с низкой повторяемостью, описанную здесь.

При использовании здесь термин "биологическая активность" описывает биологический эффект вещества на живую материю, например, на организм. Соответственно при использовании здесь термин "биологически активный белок" или "биологически активный пептид" относится к белкам или пептидам, способным оказывать биологический эффект на живые клетки/организмы, подверженные воздействию указанным белком, полипептидом или пептидом. В контексте изобретения нуклеотидная последовательность или молекула нуклеиновой кислоты, кодирующая полипептид, состоящий из пролина, аланина и возможно серина, функционально связана, в одной и той же рамке считывания, с гетерологичной нуклеиновой кислотой, кодирующей биологически активный белок.

Здесь подразумевают, что кодируемый биологически активный белок или пептид представляет собой "аминокислотную последовательность, имеющую и/или опосредующую биологическую активность", или представляет собой "аминокислотную последовательность с биологической активностью", и/или представляет собой аминокислотную последовательность, имеющую и/или опосредующую фармакологическую активность. Термины "биологически активный белок", "аминокислотная последовательность, имеющая и/или опосредующая биологическую активность", или "аминокислотная последовательность с биологической активностью", и/или "аминокислотная последовательность, имеющая и/или опосредующая фармакологическую активность," также включают любые белки или пептиды, представляющие интерес (и их функциональные фрагменты, такие как фрагменты антител, фрагменты, содержащие внеклеточный или внутриклеточный домен (домены) мембранного рецептора, усеченные формы фактора роста или цитокина и тому подобное), для которых полезно увеличить период полувыведения *in vivo* или *in vitro*. Специалисту в данной области известно, что конформация полипептида, состоящего из пролина, аланина и возможно серина, в виде случайного клубка опосредует повышенную стабильность биологически/фармакологически активного(ых) ("функционального(ых)") белка(ов) или пептида(ов) *in vivo* или *in vitro*, в частности увеличенный период полувыведения из плазмы.

В одном воплощении данного изобретения кодируемая аминокислотная последовательность, имеющая и/или опосредующая биологическую активность, согласно настоящему изобретению может быть выведена из "любого белка, представляющего интерес", то есть любого белка, представляющего интерес с фармацевтической или биологической точки зрения, или любого белка, применимого в качестве терапевтически эффективного белка. Соответственно указанный биологически активный белок или пептид может представлять собой фармакологически активный или терапевтически эффективный белок или пептид. Фармакологически активные или терапевтически эффективные белки или пептиды представляют собой любые белки или пептиды, оказывающие желаемый фармакологический, фармацевтический и/или физиологический эффект. Эффект может быть (1) профилактическим в смысле полного или частичного предотвращения и/или уменьшения степени выраженности заболевания/патологического состояния/расстройства или его симптома; и/или (2) терапевтическим в смысле полного или частичного ингибирования заболевания/патологического состояния/расстройства, то есть остановки его развития, или ослабления заболевания/патологического состояния/расстройства, то есть обеспечения обратного развития заболевания/патологического состояния/расстройства. Наиболее предпочтительно, указанный биологически активный белок представляет собой терапевтически эффективный белок, например, для применения в качестве вакцины. Таким образом, указанный биологически активный белок может также быть использован при вакцинации.

Кроме того, указанный биологически активный белок может представлять собой диагностически значимый белок. При использовании здесь "диагностически значимый белок" относится к белку или полипептиду, используемому в диагностике. В контексте настоящего изобретения диагностика относится к распознаванию и (раннему) выявлению заболевания, например, рака и опухоли, или клинического состояния у субъекта. Она может также включать дифференциальную диагностику. Кроме того, в определенных воплощениях термин "диагностика" может включать оценку тяжести заболевания или клинического состояния. В частности, полипептиды и/или конъюгаты лекарственных средств, кодируемые молекулами нуклеиновой кислоты по настоящему изобретению, могут быть использованы для диагностики *in vitro* или *in vivo*. Например, кодируемые полипептиды и/или конъюгаты лекарственных средств могут быть использованы в способах медицинской визуализации. Кодируемые полипептиды и/или конъюгаты лекарственных средств с высоким содержанием РА являются особенно подходящими для этих способов, поскольку они лучше проникают в опухолевые клетки. Кроме того, кодируемые полипептиды и/или конъюгаты лекарственных средств с высоким содержанием РА демонстрируют больший контраст между опухолевыми клетками и клетками крови или здоровыми клетками/тканями.

Кодируемый биологически активный белок выбран из группы, состоящей из связывающего белка, фрагмента антитела, цитокина, фактора роста, гормона, фермента, белковой вакцины, пептидной вакцины, пептида или пептидомиметика. При использовании здесь "пептид" предпочтительно содержит/состоит из 50 или менее аминокислотных остатков, в то время как "белок" предпочтительно содержит/состоит из 50 или более аминокислотных остатков.

При использовании здесь термин "связывающий белок" относится к молекуле, способной специфично взаимодействовать с потенциальным партнером (партнерами) по связыванию, таким образом, что она способна распознавать указанного потенциального партнера (партнеров) по связыванию среди множества молекул, отличных от указанного потенциального партнера (партнеров) по связыванию, в такой степени, что из совокупности с указанным множеством молекул, не таких как указанный потенциальный партнер (партнеры) по связыванию, она связывается или существенным образом связывается только с указанным потенциальным партнером (партнерами) по связыванию. Способы измерения связывающей активности связывающего белка в отношении потенциального партнера по связыванию известны в данной области, и могут быть рутинным образом применены, например, посредством твердофазного иммуноферментного анализа (ELISA), изотермической титрационной калориметрии (ИТС), равновесного диализа, анализов осаждения, микромасштабного термофореза, флуоресцентного титрования или спектроскопии методом поверхностного плазмонного резонанса (SPR) с использованием, например, прибора Biacore.

Типичные связывающие белки/связывающие молекулы, которые могут быть применены в контексте настоящего изобретения, включают, без ограничения, антитела, фрагменты антител, такие как Fab-фрагменты, Fab'-фрагменты, F(ab')₂-фрагменты, одноцепочечные переменные фрагменты (scFv), (одно)доменные антитела, в частности имеющие происхождение от представителей семейства Верблюдовых, лам или акул, выделенные переменные области антител (области VL и/или VH), в частности от людей и приматов, CDR, иммуноглобулиновые домены, пептидомиметики, имеющие происхождение от CDR, лектины, фибронектиновые домены, тенасциновые домены, домены белка А, домены SH3, домены анкириновых повторов и липокалина или различные типы каркасных связывающих белков, как описано, например, в Skerra (2000) *J. Mol. Recognit.* 13: 167-187, Gebauer (2009) *Curr. Opin. Chem. Biol.* 13:245-255 или Binz (2005) *Nat. Biotechnol.* 23: 1257-1268.

Другие типичные кодируемые биологически или фармакологически активные белки или терапевтически эффективные белки, представляющие интерес, которые могут быть применены в контексте настоящего изобретения, включают, без ограничения, антагонист рецептора интерлейкина, антагонист рецептора интерлейкина-1, такой как EBI-005 или анакинра, лептин, ацетилхолинэстераза, активированный белок С (дротрекогин), антагонист рецептора активина ПВ-типа, аденозиндезаминазу, агалсидазу альфа, агонист Toll-подобного рецептора-5, такой как энтолимод, альфа-1-антитрипсин, ингибитор альфа-1-протеиназы, альфа-галактозидазу, человеческий предсердный натрийуретический пептид альфа, альфа-N-ацетилглюкозаминидазу, алтеплазу, амедилазу, амилин, аналог амилина, ANF-Rho, ангиотензин (1-7), ангиотензин II, ангиотензинпревращающий фермент 2, одноцепочечный фрагмент антитела против эпителиальной молекулы клеточной адгезии, антитромбин альфа, антитромбин III, апоптоз-индуцирующий фермент mi-APO, аргининдезаминазу, аспарагиназы, такие как каласпаргаза (calaspargase), пэгаспаргаза, крисантаспаза, фактор VIII с делецией домена В, такой как бероктоког альфа или октофактор, бектумаб (лимфоскан), липазы, стимулируемые желчными кислотами, такие как буцелипаза альфа (bucelipase alfa), связывающий белок, направленный против респираторно-синцитиального вируса, такой как паливизумаб, морфогенетические костные белки, такие как BMP-2 (диботермин альфа) или BMP-6, буганин, бычий карбоксигемоглобин, бычий гормон роста, ингибитор эстеразы C1, белок-экзофермент C3, карбоксигемоглобин, антагонист CD19, антагонист CD20, такой как ритуксан, антагонист рецептора CD3, антагонист CD40, антагонист CD40L, такой как дапиролизумаб или антова (Antova), цереброзидсульфатазу, цетрин, такой как VGX-210, хондроитинлиазу, фактор свертывания крови IX, такой как нонаког гамма, конаког бета (conacog beta), албутрепенаког альфа (albutrepenonacog alfa), фактор свертывания крови VIIa, такой как эптаког альфа, марзептаког альфа (marzeptacog alfa), ватрептаког альфа (vatreptacog alfa), орептаког альфа (ogertacog alfa), фактор свертывания крови VIII, такой как сусоктоког альфа (susococog alfa), дамоктоког альфа (damococog alfa), туроктоког альфа, руриоктоког альфа (rurioctocog alfa), эфмороктоког альфа, эфралоктоког альфа (efralococog alfa), симоктоког альфа, фактор свертывания крови X, фактор свертывания крови XIII, такой как катридекаког, коллагеназа *Clostridium histolyticum*, ингибитор компонента комплемента C3, антагонист рецептора комплемента 5a, кортикотропин-рилизинг-фактор, антагонисты рецепторов CSF1, такие как FPA008, антагонист CSF1R, антагонист CTLA-4, такой как ипилимумаб, циановирин-N, дезоксирибонуклеаза I, такая как дорназа альфа, антагонист рецептора EGFR, эластазы, такие как человеческая панкреатическая эластаза типа I, такая как вонапанитаза (vonapanitase), эндостатин, энкастим (enkastim), эпидермальный фактор роста, эритропоэтин альфа, эритропоэтин дзета (erythropoietin zeta), антагонисты рецептора FcγIIb, фибриногеназу, фибринолитический фермент, такой как бриназа, фактор роста фибробластов 1 (человеческий кислый фактор роста фибробластов), фактор роста фибробластов 18, фактор роста фибробластов 2 (человеческий основной фактор роста фибробластов), фактор роста фибробластов 21, антагонисты рецептора фактора роста фибробластов 2,

такие как FPA144, Fms-подобный лиганд тирозинкиназы 3, фолликул-стимулирующие гормоны, такие как фоллитропин альфа или фоллитропин бета, фрагмент человеческого бактерицидно-го/пермеабилзирующего белка 21 (опебакан (orebacan)/rBPI 21), гелонин, агонист рецепторов глюкогона, антагонист гликопротеина Pb/IIIa, такие как абциксимаб, ферменты, разрушающие гликозаминогликаны, такие как кондолиаза (condoliase), gp120, gp160, гранулоцитарный колониестимулирующий фактор (G-CSF), гранулоцитарно-макрофагальный колониестимулирующий фактор (GM-CSF), белок теплового шока hsp65 *Mycobacterium* BCG, слитый с транскрипционным фактором E7 (верпасеп кальтеспен (verpaser caltespen)), фактор роста гепатоцитов, антагонист рецептора фактора роста гепатоцитов (HGFR), антагонист гепсидина, антагонист рецептора Her2/neu, такой как герцептин, гетеродимерный 15:IL-15Ra (hetIL-15), гирудин, антагонист hsp70, человеческую кислую сфингомиелиназу, хорионический гонадотропин человека, такой как хориогонадотропин альфа, человеческие кислые α -глюкозидазы, такие как ревеглюкозидаза альфа (reveglucosidase alfa) или алглюкозидаза альфа, человеческий гормон роста, человеческий фактор роста кератиноцитов (KGF), человеческую матриксную металлопротеиназу, фрагмент основного белка миелина человека, человеческий остеогенный белок 1, человеческий остеогенный белок-1, человеческий паратиреоидный гормон, человеческий тромбомодулин альфа, гиалуронидазу, такую как gHuPH20, гиалуронидазы, такие как человеческая гиалуронидаза PH-20 (воргиалуронидаза альфа (vorhyaluronidase alfa)), гиалозидаза (hyalosidase) или бовгиалуронидаза, гидролитические лизосомальные глюкоцереброзид-специфичные ферменты, такие как глюкоцереброзидаза, велаглюцераза альфа или талиглюцераза альфа, идуронат-2-сульфатазу, антагонисты IgE, такие как омализумаб, инсулина, белок ирокез-гомеобокса 2 (iroquois homeobox protein 2, IRX-2), инсулин, аналог инсулина, антагонист интегрин $\alpha 4\beta 1$, интерферон-гау, интерферон-альфа, антагонист интерферона-альфа, суперагонист интерферона-альфа, интерферон-альфа-n3 (Alferon N Injection), интерферон-бета, интерферон-гамма, интерферон-лямбда, слитые белки интерлейкина-2, такие как DAB(389)IL-2, интерлейкин-11, такой как опрелвекин, интерлейкин-12, антагонист рецептора интерлейкина-17, белок, связывающийся с интерлейкином-18, интерлейкин-2, интерлейкин-22, интерлейкин-4, такой как питракинра (pitakinra), мутеин интерлейкина-4, антагонист рецептора интерлейкина-6, интерлейкин-7, антагонист альфа-субъединицы рецептора интерлейкина-22 (IL-22ra), ирисин, белок, ассоциированный с островковым неогенезом, каллидиногеназу, лактоферрин, фрагмент лактоферрина, ланотеплазу, липазные ферменты, такие как бурлулипаза (burulipase), ризолипаза (rizolipase), эпафипаза (epafipase) или себелипаза альфа, лютеинизирующий гормон, лутропин альфа, молекула лимфоцитарной экспансии (lymphocyte expansion molecule), лизостафин, желудочную липазу млекопитающих (Merispace), маннозидазы, такие как велманаза альфа, агонист рецептора меланокортина-4, пептид из 23 аминокислот, имеющий происхождение от MERE, метиониловый фактор стволовых клеток человека (анцестим), микроплазмин, N-ацетилгалактозамин-6-сульфатазу, такую как элосульфаза альфа, N-ацетилглюкозаминидазу, насаруплазу бета (nasaruplase beta), фактор роста нервов, нейрегулин-1, нейротоксин (например, клостридиальный нейротоксин, такой как нейротоксин *Clostridium botulinum* (такой как нейротоксин *Clostridium botulinum* серотипа A, B, C, D, E, F или G, особенно нейротоксин *Clostridium botulinum* серотипа A), нейтрофильный желатиназоассоциированный липокалин, окриплазмин, ингибитор комплемента *Ornithodoros moubata* (OmCI/Coversin), остеопротегерин, P128 (StaphTAME), памитеплазу, паратиреоидный гормон (PTH), антагонист PD-1, антагонист PDGF, белок пентраксин-2, фаговый лизин, такой как HY133, фенилаланинаммиаклиазу, такую как валиаза, фосфатазы, такие как тканенеспецифическая щелочная фосфатаза альфа, плазминоген, вариант плазминогена, такой как V10153, тромбоцитарный фактор роста BB, свиной гормон роста, пептид направленного действия на прохитин I, проинсулин, белок A, белок C, такой как дротрекогин, белок, связывающийся с лигандами рецептора фактора роста фибробластов, такой как FP-1039, рекомбинантный ингибитор пути тканевого фактора (тифакогин), релаксин, аналог релаксина, такой как серелаксин, ретеплазу, rhPDGF-BB, рибонуклеазу, такую как онконаза или амфиназа, сенреботаза (senrebotase), ингибиторы сериновых протеаз, такие как конестат альфа, сфериказу (sfericase), сиалидазу, растворимый рецептор DCC (обнаружен при колоректальном раке), растворимый рецептор TACI (атацицепт), растворимый рецептор фактора некроза опухоли I (sTNF-RI), растворимый рецептор фактора некроза опухоли II (sTNF-RII), растворимый рецептор VEGF Flt-1, растворимый человеческий рецептор Fc γ IIb, стафилокиназу, стрептокиназу, сульфамидазу, лиганд T-клеточного рецептора, тенектеплазу, белок, стимулирующий тромбоциты (AMG-531), тромбопоэтин, тромбоспондин, тироидный гормон, аналог тиротропин-релизинг-гормона (TRH), такой как тальтирелин, тканевый активатор плазминогена, активатор плазминогена тканевого типа, такой как памитеплаза, трипептидилпептидазу I, фактор некроза опухоли (TNF α), антагонист фактора некроза опухоли α , уриказу, такую как расбуриказа или пегадриказа (pegadricase), уродилатин, урофоллитропин, урокиназу, утероглобин, антагонист VEGF, такой как ранибизумаб или бевацизумаб, антагонист VEGF/PDGF, антагонист VEGF/PDGF, такой как мульти-VEGF/PDGF DARPin или слитый белок, вискумин, факторы фон Виллебранда, такие как вониког альфа (vonicog alfa). Здесь предпочтительны антагонист рецептора интерлейкинов, особенно антагонисты рецептора интерлейкина-1, такие как EBI-005 или анакинра, и лептин, особенно человеческий лептин или мутантный человеческий лептин (huLeptin(W100Q), мутантный человеческий лептин с заменой триптофана на глутамин в

положении 100 зрелой полипептидной цепи). Аминокислотная последовательность человеческого лептина раскрыта, например, в UniProtKB под номером доступа P41159. Мутантный huLeptin(W1100Q) описан Ricci (2006) Mutational approach to improve physical stability of protein therapeutics susceptible to aggregation, In Murphy (ed.) Misbehaving proteins: protein (mis)folding, aggregation, and stability, Springer, 1st edition, New York.

Типичные пептиды и пептидомиметики включают, без ограничения, аденокортикотропный гормон (АКТГ), афамеланотид, аларелин (alarelin), ингибитор альфа-4-интегрин, ингибитор слияния HIV (такой как энфувиртид, V2o, SC34EK, SC35EK, IQN17 или IZN17), агонист рецептора ангиотензина II 2 (AT2) (такой как LT2), антиидиотипический пептид p53, амилин, аналог амилина, астрессин, атозибан, фрагмент бактериального пептида, активный против рака и HIV (такой как ATR-01), бициклический пептид (такой как TG-758), бивалирудин, антагонист брадикинина (такой как икатибант), бремеланотид, натрийуретический пептид В-типа, кальцитонин, карбетоцин, карфилзомиб, хризалин, циленгитид, натрийуретический пептид С-типа, колостринин, кортикотропин-рилизинг-фактор (такой как ксерсепт (Xercept), косинтропин), опухолевый хоминг-пептид CNGRCG, пептид ω -коноксина (такой как зиконотид), С-пептид, данегапид (daneaptide), дефензин, экаллангид, элкатонин, элдеоисин, эксендин-4, аналог эксендина-4 (такой как эксендин 9-39), эзриновый пептид 1, фрагменты человеческого фосфогликопротеина внеклеточного матрикса (такие как AC-100), галанин, желудочный ингибирующий пептид (GIP), аналог GIP, глатирамер, глюкагон, аналог глюкагона, глюкагоноподобный пептид 1 (GLP-1), аналог GLP-1 (такой как ликсисенатид, лираглутид или семаглутид), глюкагоноподобный пептид 2 (GLP-2), аналог GLP-2 (такой как тедуглутид), гонадорелин, агонист гонадотропин-рилизинг-гормона (такой как госсерелин, бусерелин, трипторелин, леупролид, протирелин, лецирелин, фертирелин (fertirelin) или деслорелин), антагонист гонадотропин-рилизинг-гормона (такой как абареликс, цетрореликс, дегареликс, ганиреликс или тевереликс), грелин, аналог грелина (такой как AZP-531), рилизинг-гормон гормона роста, аналог рилизинг-гормона гормона роста (такой как серморелин или тезаморелин), гематид, пептидный миметик гепсидина, гистрелин, индолицидин, аналог индолицидина (такой как омиганан), пептид понижающей регуляции IgE (такой как SC-01), пептид INGAP (эксулин), инсулиноподобный фактор роста 1, инсулиноподобный фактор роста 2, антагонист ионных каналов Kv1.3 (такой как cgtxA, cgtxE или cgtxF), ланреотид, лектин-связывающий пептид (такой как sv6B, sv6D, svC2, svH1C, svH1D или svL4), лантипептид, ларазотид, линаклотид, лусупултид (lusupultide), агонист рецептора меланокортина-4 (такой как AZD2820), пептид из 23 аминокислот, имеющий происхождение от МЕРР, митохондриальный пептид (такой как MOTS-с, гуманин, SHLP-6 или SHLP-2), мутант белка-2, связывающегося с инсулиноподобным фактором роста (такой как I-HBD1), модуляторы ионных каналов Nav (такие как GT \times 1-15 или VST \times 3), октреотид, пептид, ингибирующий пропротеинконвертазу субтилизин/кексин 9 (PCSK9), пептидный фрагмент азурина, филомер, пептидный антагонист инвариантного пептида, ассоциированного с МНС класса II (CLIP) (такой как VG1177), пептид, имеющий происхождение от белка теплового шока (такой как энкастим (enkastim)), пексиганан, пловамер, прамлинтид, пептид направленного действия на прохибитин-1, проостровковый пептид, тирозинтирозиновый пептид (PYY 3-36), RGD-пептид или пептидомиметик, рамоплагин, секретин, синапультид (sinapultide), соматостатин, аналог соматостатина (такой как пасиреотид или CAP-232), антимицробный пептид специфического направленного действия (specifically targeted antimicrobial peptide, STAMP) (такой как C16G2), агонист рецептора костного морфогенетического белка (такой как THR-184 или THR-575), стресскопин, сурфаксин, Тс99m апцитид, терипаратид (PTH 1-34), тетракозактид, тимозин альфа-1, TLR2-ингибирующий пептид, TLR3-ингибирующий пептид, TLR4-ингибирующий пептид, тимозин В4, тимозин В15, вазоактивный кишечный пептид, вазопрессин, аналог вазопрессина (такой как десмопрессин, фелипрессин или терлипрессин).

Типичные биологически активные белки, которые представляют интерес и могут быть применены в контексте вакцинации, включают, без ограничения, пептид AE37, гастрин-рилизинг пептид бомбезин, карциноэмбриональный антиген (CEA), открытую рамку считывания 2 (ORF2) белка капсида, белок вируса гепатита Е, холерный токсин В, агглютинин A Staphylococcus aureus, дифтерийный токсин, мутантный дифтерийный токсин (такой как CRM197), термолабильный энтеротоксин E. coli, экзотоксин A Pseudomonas aeruginosa, белок F вируса кори, гликопротеин E вируса японского энцефалита (JE), пептид, имеющий происхождение от GPC3, полипротеин вируса гепатита А, пептид GP2, имеющий происхождение от HER2, геррегулин (herregulin), пептид Her2neu, поверхностный антиген вируса гепатита В (HbSAg), изоформу 65 кДа белка человеческой декарбоксилазы глутаминовой кислоты (rhGAD65), гемагглютининовые антигены вируса гриппа (HA), нейраминидазу вируса гриппа (NA), белок L1 вируса папилломы человека, гибридный пептид li-Key/HER2/neu, наружный поверхностный липопротеин Vogelia burgdorferi (OspA), главный наружный белок капсида ротавируса, пептид муцин-1 (MUC-1), белок капсида вируса Норволк (rNVP), VLP парвовируса В19, пептид, имеющий происхождение от гранулоцитарно-макрофагального колониестимулирующего фактора, белок капсида цирковируса свиней 2 (PCV2 ORF2), белок С вируса клещевого энцефалита, белок Е вируса клещевого энцефалита, белок Е вируса желтой лихорадки, белок Е-1 вируса краснухи, белок G вируса бешенства, белок Н вируса кори, белок Н парамиксовируса, белок NS вируса желтой лихорадки, белок N парамиксовируса, простатоспецифиче-

ский антиген, белок E2 вируса лихорадки свиней, белок VP6 ротавируса, белок VP7 ротавируса, белок шипов вируса SARS (D3252), белок VP1 вируса полиомиелита, белок VP4 вируса полиомиелита, онкобелок Ras, пептиды семенной жидкости (такие как YLP12, P10G, A9D, mFA-12-19, SP56 и/или mFA-1117-136), столбнячный токсин, туберкулин, опухолеассоциированные пептиды (tumor-associated peptides, TUMAP) (такие как IMA901, IMA910 или IMA950) и тому подобное.

В одном аспекте настоящее изобретение относится к молекуле нуклеиновой кислоты, раскрытой здесь, например, молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов; или в одном аспекте настоящее изобретение относится к молекуле нуклеиновой кислоты, раскрытой здесь, функционально связанной в одной и той же рамке считывания, с нуклеиновой кислотой, кодирующей биологически активный белок, например молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов, где молекула нуклеиновой кислоты функционально связана, в одной и той же рамке считывания, с нуклеиновой кислотой, кодирующей биологически активный белок;

где указанная нуклеотидная последовательность не является

```

ATGGGCAGCAGCCATCATCATCACCATCATGGTAGCCTGGTCCGGCTAGCTCTTCTGCA
AGTCCGGCAGCACCAGCCAGCTTCCACAGCTGCACCAGCACCAGCCAGCCAGCCAGCA
TCTCCAGCAGCCCTGCACCGGCAAGCCCTGCAGCTCCAGCACCAGCCAGCCAGCAGCA
AGCCAGCTGCTCCTGCTCCAGCGAGCCAGCAGCCAGCTCCTAGTCCCTGCTGCTGCT
TCTCCTGCTGCTCCGGCAGCAGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TCCCGAGCCGAGCAGCTCAGCTTCTCCGCGAGCCGAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCT
TCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGTCCAGCAGCACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC
CCATTTGTGAACAAGCAGTTAACTATAAGGACCCGGTGAACGGTGTGGATATCGCGTAT
ATCAAAATCCCGAATGCGGGCCAGATGCAACCAGTCAAGGCGTTCAAGATTCATAACAAG
ATTGGGTATTCGGGAACGTGATACCTTCACCAATCCGGAAGAAGGGCAGCTAAACCCG
CCGCCAGAAGCCAAACAAGTCCCGGTGAGCTACTATGATAGCACGTATCTTAGCACCGAT
AATGAAAAGACAATTACCTGAAGGGCTGACCAAGTGTTCGAGCGCATCTACAGTACC
GACTTAGGCCGATGTTGTGACGAGCATCGTTCGCGGTATCCCGTTCGGGGCCGCTCG
ACCAATGATACCGAGTTGAAAGTCAATGACACGAACTGTATCAATGTTATCCAAACCGGAC
GGCAGTTATCGCAGCGAGGAGTTAAATTTGGTCAATCCGGTCCAAAGCCAGATATTTATT
CAGTTCGAATGCAAGAGCTTCGGCCATGAGGCTTGAATTTGACCGCAACGGTTACGGC
AGCACCCAAATACATCCGCTTAGCCCGGATTTCACTTTGGCTCGAGGAGAGCTTGAGG
GTGGACACCAACCCGCTGTAGGTGCCGGCAAAATCCGCAACCCAGCCGCGAGTACGCTT
GCGCAGAAATGATTCATGCGGGTCAACGCTTATACGGTATCGCGATCAATCCGAATCCG
GCTTTAAACTCAATACCAACCGCTACTACGAAATGAGCGGCTTAGAGGTTAGCTTTGAA
GAATTAACCGACCTTCGGTGGCCAGCAGCCAAAGTTCATCGACAGCTGCAGGAAAATGAG
TTCGCTTGTACTATTACAATAAAATCAAGGACATCGCGAGCACCTTAAATAAAGCAAA
AGCATTTGGGGCACCACCGCAAGCTTGCAGTACATGAAGAAGCTATTAAGGAAAATAT
TTGTTGTGCGGAGGATACAGCGGAAATTCAGCGTCGATAAGCTGAAATTCGCAAAATG
TATAAAATGCTGACCGAGATTTACACCGAGGATAACTTCGTCAGTTTTTTAAAGGTGTTA
AATCGTAAGACCTATTAAACTTTGATAAAGCGGTGTTAAAAATTAATATCGTCCGGAAG
GTGAATTACACCATCTACGATGGTTCAATTTACGCAACACGAAATTCGGCGGCGAATTT
AATGGCCAAAACACCGAAATTAACAACATGAACTTTACGAAGTTAAAGAAATTTACGGGG
TTATTCGAATTTTACAAGTTATATGCTGCGCGGCAATATACAGCAAGCCAGGTGCG
GGCAAGTCTTTGGTTCGGGCTGGCAGCCCGCGCGCGCGCTCAATGATCTGTGTATT
AAAGTCAATAACTGGGACCTGTCTTCAGCCGAGCGGAGGATAACTTTACCAACGACTTA
AACAAAGCCGAGGAGATCACGAGCGATACGAACATCGAGGCGCGGAGGAAAATATTAGC
CTGGACCTCATTGACGAGTACTATCGAGTTCATTTGACAAATGAGCCGGAGAACATC
AGCATTGAAAATCTCAGCAGCGACATCATCGTCAAGTTGAACTGATGCGCAACATTTGAA
CGCTTTCGAAACGGCAAAAAATGAACTGGACAAAGTATACCATGTTCCATTACTTACGC
GCACAGGAAATTTAGCAGCGCAAGACCGCATTCGCTGACCAATAGCGTTAACAGGAGCC
TTGTTAAATCCGAGCGGTGCTACACGTTCTTCAGCAGCGATTTGTTCAAAAAAAGTGAAC
AAGGCGACCGAAGCCGATGTTTTGGGCTGGGTCGAGCAATTTGGTTTACGATTTTACC
GAGCAAAACAGCGAGGTGAGCAGCAGCCGACAAAATTCGAGATATCACCATCATCTCCG
TACATCGGTCGGCGCTCAATATCGGCAATATGTTATACAAGGACGACTTTGGGGCGCG
CTGATCTTTAGCGCGCGGTTATCTTATAGAAATTCATCCCGGAGATCGCAATCCCGGTC
TTGGGCACCTTTGCGTGTGAGCTATATCGCAATAAAGTGTCCAGGTTCAAAACCATC
GATAACGCGCTCAGCAAGCGTAATGAGAAATGGGACGAGGTTTATAAGTATATCTGTGACC
AATCGTTAGCAAAAATCAATACGAGATCGATCTCATCCGCAAAAAATGAAAGAGCC
AAAGATCGCTCCTCAAGTATATTTGACAAACCGCGGACCCCTCATTTGGCCAGGTGGAC
CGCTGAAGGATAAAGTGAACAATAGCTCAGCAGCGATATCCCGTTCAGCTGAGCAAG
TACGTCGACAAACAGCGCTTACTGAGCACCTTTACCGAGTATATCAAGAACATCATTAAT
ACCAGCATCTCAACTTGGCTGATGAGAGCAATCACCTGATCGACCTCAGCCGCTACGCC
AGCAAGATCAACATCGGCAGCAAGGTCAATTTCCAGCCGATCGATAAAGAAATCAGATCCAA
TTGTTTAACTGGAAAGCAGCAAGATCGAGGTTATCTTGAAGAACCGGATTTGTATCAAC
AGCATGTACGAGAACTTTAGCAGGAGCTTCTGGATTCGATCCCGAAGTATTTCAATAGC
ATTAGCCTGAATAACGAATAATACCATTAACAACGCTGCAATGGAATAAATAGCGGCTGGAAG
GTGAGCTTAAATACGGCGAGATCATTGGACCTTACAGGATACCCAAAGAAATCAACAG
CGGCTGCTTTTAAATAGCCAGATGATCAACATCAGCGATTAACATCAACCGCTGGAT
TTCGTGACCATCACCATAATCGCTTGAATAATAGCAAGATTTACATCAATGGTCCGCTTG
ATTGATCAAAAAACCGATCAGCAATCTCGGTAATATCCATGCCAGCAATAACATCATGTTT
AAGTTAGACGGTTGCGCGGATACCCAGCGCTATATCTGGATCAAGTATTTAACTTATTT
GATAAGGAACTCAACGAAAGGAAATTAAGAACTTATATGACAAATCAGAGCAATAGCGGC
ATCCTGAAGGATTTTCGGGGCACTACCTGCAAGTACGATAAGCCGCTACTATATGTTGAAC
TTGATGACCCGAACAAATATGTCGATGTGAACAATGTGGGTATTCCTGGCTATATGTCAC

```

TAAAGGGCCCGGTGGTAGCGTGATGACCACGAATATTTACTTAAACAGCAGCTTATAC
 CGCGGCACGAAGTTTATATCAAGAAGTATGCCAGCGGCAACAAGGACAATATCGTCCGC
 AACACACCGCTGTGTATATTAACGTGGTGGTGAAGAAATAAGAGTACCGCTTGGCCACG
 AATGCGAGCCAGCGGGCGTGGAAAAATCTTGAGCGCGTGGAGATCCCGGACGTCCGC
 AACCTCAGCCAGGTTGGTGTGATGAAGTCTAAAAACGACCGAGGCGATCACGAACAAGTC
 AAAATGAATTTGCAAGATAACACCGCAACGCATCGGCTTATTTGGTTTTCACCAGTTC
 AATAACATCGCCAACTCGTGGCCAGCAATTTGGTATAACCGCAATTTGAACGCGAGCAGC
 CGCACGCTCGGCTGTAGCTGGGAGTTTCATCCCGGTGGACGATGGCTGGGGCGAGCGCCG
 CTCGGAGATCTGGTGCACCGGTTCCGCGAATTCGAGCTCCGTCGACAAAGCTTTGGAGC
 CACCCGAGTTTCGAAAAATA

(SEQ ID NO: 196).

В одном аспекте настоящее изобретение относится к молекуле нуклеиновой кислоты, раскрытой здесь, например, молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов; или в одном аспекте настоящее изобретение относится к молекуле нуклеиновой кислоты, раскрытой здесь, функционально связанной, в одной и той же рамке считывания, с нуклеиновой кислотой, кодирующей биологически активный белок, например молекуле нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов, где молекула нуклеиновой кислоты функционально связана в одной и той же рамке считывания с нуклеиновой кислотой, кодирующей биологически активный белок;

где указанная нуклеотидная последовательность не является

ATGGGTAGCAGCCATCATCATCACCATCATGGTAGCCTGGTTCCGCGTAGCTCTTCTGCA
 AGTCCGGCAGCACCGGCACCGGCTTCAACAGCTGCACCAGCACCTAGCGCACCGGCAGCA
 TCTCCAGCAGCCCTGCACCGGCAAGCCCTGCAGCTCCAGCACCGCTCAGCACCGCAGCA
 AGCCAGCTGCTCCTGCTCCAGCGAGCCAGCAGCGCCAGCTCCTAGTGCCCTGTGCTGCC
 TCTCCTGCTGCTCCGCGCACCAGCAAGTCTGCTGCGCCTGCACCAGTGTCTCCGGCTGCT
 AGTCTGCGCAGCAGCTCCGGCTAGTCCAGCTGCTCCAGCCCTTCAGCCCTGCAGCA
 CCATTTGTGAACAAGCAGTTTAACTATAAAGGACCCGGTGAACGGTGTGGATATCGCGTAT
 ATCAAAATCCCGAATGCGGGCCAGATGCAACCAGTCAAGGCGTTCAGATTCATAACAAG
 ATTTGGGTTATTCGGAACTGATACCTTCACCAATCCGGAAGAAGCGGATTTAAATCCG
 CCGCCAGAAGCCAACAAGTGCCTGAGCTACTATGATAGCACGTATCTTAGCACCGAT
 AATGAAAAGACAATTACCTGAAGGCGTGACCAAGTTCAGAGCGCATCTACAGTACC
 GACTTAGCGCCATGTTGTTGACGAGCATCGTTCGCGGTATCCCGTCTCTGGGGGGGCTCG
 ACCATTGATACCGAGTTGAAAGTCAATGACACGAACTGATCAATGTTATCCAACCGGAC
 GGCAGTTATCGCAGCGAGGAGTTAATTTGGTCATCATCGGTCCAAGCGCAGATATATT
 CAGTTCGAATGCAAGAGCTTCGGCCATGAGGCTTGAATTTGACGCGCAACGGTTACGGC
 AGCACCAATACATCCGCTTTAGCCCGGATTCACCTTTGGCTTCGAGGAGAGCTTGGAG
 GTGGACACCAACCGCTGTTAGGTGCCGGCAATTCGCAACCGACCCGGCAGTGACGTTG
 GCGCACGAATGATTCATGCGGGTCAACCGCTTATACGGTATCGCGATCAATCCGAATCCG
 GTCTTTAAAGTCAATACCAACCGCTACTACGAAATGAGCGGCTTAGAGGTTAGCTTTGAA
 GAATTAGCCACCTTCGGTGGCCACGACGCAAGTTCATCGACAGCCTGCAGGAAATGAG
 TTCGCTTGACTATTACAATAAATCAAGGACATCGCGAGCACCTTAAATAAAGCAAAG
 AGCATTTGGGCACCACCGCAAGCTTGCAGTACATGAAGAAGTATTTAAGGAAAAATAT
 TTGTGTCCGGAGGATACCGAGCGGAAATTCAGCGTCGATAAGCTGAAATTCGACAAATG
 ТАТАААТGCTGACCGAGATTTACACCGAGGATAACTTCGTCAGTТТТТАAGGTGTTA
 ААТCGТАAGACCTATTTAAACTTTGATAAAGCGGTGTTAAATTAATATCGTGCCGAAG
 GTGAATTACACCATCTACGATGGTTTCAATTTACGCAACACGAACTTGGCGGGCAATTTT
 ААТGGCCAAAACACCGAAATTAACAACATGAACTTTACGAAGTТААГААТТТCACCGGC
 ТТАТTCGAATТCTACAAGTTATTAAGCGTGGCGGCATCATACCAGCAAGGCAGGTGCG
 GGCAGTCCTTGGTTCCGCGTGGCAGCGCCGCGCGCGCGCTCAATGATCTGTGTATT
 AAAGTCAATAACTGGGACCTGTCTCTCAGCCGAGCGAGGATAACTTTACCAACGACTTA
 AACAAAGGCGAGGAGATCAGAGCGATACGAACATCGAGGCGCGGAGGAAAATATTAGC
 CTGGACCTCAATCAGCAGTACTATCTGACGTTCAATTTGACAATGAGCCGGAGAATC
 AGCATTTGAAAATTCAGCAGCGACATCATCGGTGAGTTGAACTGATGCCGAACATTTGAA
 CGCTTCCGAACGGCAAAAAATATGAACGACAGTATACCATGTTCCATTACTTACGC
 GCACAGGAATTTGAGCAGCGCAAGAGCCGATTTGCGCTGACCAATAGCGTTAACGAGGCC
 TTGTTAAATCCGAGCCGTGTCTACACGTTCTCAGCAGCGATATGTCAAAAAAGTGAAC
 AAGCGCACGAAGCCGATGTTTGGGCTGGTCCAGCAATTTGTTTACGATTTTACC

GACGAAACCAGCGAGGTGAGCAGCACCGACAAAAATTGCAGATATCACCATCATCATCCG
TACATCGTCCGGCGCTCAATATCGGCAATATGTTATACAAGGACGACTTTGTGGGCGG
CTGATCCTTAGCGGCGGGTTATCTTATAGAATTCATCCCGGAGATCGCAATCCCGGTC
T'GGGCACCTTTGGCTGGTGAAGCTATATCGCGAATAAAGTCTCACGGTCCAAACCAAT
GATAACGGCTCAGCAAGCGTAATGAGAAATGGGACGAGGTTATAAGTATATCGTGACC
ACTGGTTAGCAAAAGTCAATACGAGATCGATCTCATCCGCAAAAAATGAAAGAGGC
TTGAAAAATCAAGCGGAGGCAACCAAGCCATCAATTAATACCAGTATAACCAATATACC
GAAGAAGAAAAAACAATATCAACTCAATATCGATGATTTGAGCAGCAAACTGAACGAG
AGCATTAACAAAGCGATGATTAACATCAACAGTCTTGAATCAATGCAGCGTGAAGCTAT
CTCATGAACAGCATGATCCCGTATGGCTCAACCGCTTGGAAAGATTTGACGCCAGCCTG
AAAGATGCGCTCCTCAAGTATATTTATGACAACCGCGCACCTCATTTGGCCAGGTGGAC
CGCTTGAAGGATAAAGTGAACAATACGCTCAGCAGCGATATCCCGTCCAGCTGAGCAAG
TACGTCGACAACCAGCGCTTACTGAGCACCTTACCAGTATATCAAGAACATCATTAAT
ACCAGCATCCTCAACTTGGCTATGAGAGCAATCACCTGATCGACCTCAGCCGCTACGCC
AGCAAGATCAACATCGGAGCAAGGTCAATTTGACCCGATCGATAAAGATCAGATCCAA
TTGTTAACCTGGAAAGCAGCAAGATCGAGGTTATCTTGAAGAAGCGGATTTGTACAAC
AGCATGTACGAGAAGTTAGCAGCGCTTCTGGATCGTATCCCGAAGTATTTCAATAGC
ATTAGCCTGAATAACGAATATACCATATCAACTGCATGGAAAAATAATAGCGGCTGGAAG
GTGAGCTTAAATACGGCGGAGATCATTGGACCTTACAGGATACCAAGAAATCAACAG
CGCGTCTGTTAAGTATAGCCAGATGATCAACATCAGCGATACATCAACCGCTGGAT
TTCGTGACCATCACCATAATCGCTTGAATAATAGCAAGATTTACATCAATGGTCGCTTG
ATTGATCAAAAACCGATCAGCAATCTCGGTAATATCCATGCCAGCAATAACATCATGTTT
AAGTTAGACGGTTGGCGGATACCCACCGCTATATCTGGATCAAGTATTTAACTATTT
GATAAGGAATCAACGAAAAGGAAATTAAGACTTATATGACAAATCAGAGCAATAGCGGC
ATCCTGAAGGATTTTGGGGCAGTACCTGCAGTACGATAAGCCGTACTATATGTTGAAC
TTGATGACCCGAAACAATATGTCGATGTGAACAATGTGGGATTTCTGGCTATATGTAC
TTAAAGGGCCCGTGGTAGCGTGATGACCCGAAATTTACTTAAACAGCAGCTTATAC
CGCGGACGAAAGTTTATATCAAGAAGTATGCCAGCGGCAACGAAGCAATATCGTCCGC
AACAAACCGTGTGTATATTAACGTGGTGGTGAAGAATAAAGAGTACCGCTTGGCCAG
AATGCGAGCCAGGCGGGCTGGAAAAATCTTGAAGCGCTTGGAGATCCCGGACGTCGGC
AACCTCAGCCAGGTTGTGGTGAAGTCAAAAACGACCGGGCATCAGCAACAGTGC
AAAATGAATTTGCAAGATAACAACGGCAACGACATCGGCTTTATTGGTTTTACCAAGTTC
AATAACATCGCCAACTCGTGGCCAGCAATTTGGTATAACCGCAATTAACGAGCAGC
CGCAGCTCGGCTGTAGTGGGAGTTCATCCCGTGGAGGATGGCTGGGGCAGGCGCCG
CTCGGAGATCTGGTCCACCGGTTCCCGCAATTCGAGCTCCGTCGACAAGCTTTGGAGC
CACCGCAGTTCGAAAAATAA

(SEQ ID NO: 197).

В одном аспекте настоящее изобретение относится к молекуле нуклеиновой кислоты, состоящей из нуклеотидной последовательности, кодирующей полипептид, состоящий из пролина, аланина и возможно серина,

где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов,

где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (NRS) менее 50000,

где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле:

$$NRS = \frac{\sum_{n=1}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности,

n представляет собой длину повтора в указанной нуклеотидной последовательности и

$f_i(n)$ представляет собой частоту указанного повтора длиной n ,

где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных различных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1.

Кроме того, настоящее изобретение также относится к вектору, содержащему молекулу нуклеиновой кислоты или нуклеотидную последовательность по изобретению. Вектор может также быть использован для обеспечения молекулы нуклеиновой кислоты, содержащей (1) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, и (2) нуклеотидную последовательность, кодирующую биологически активный белок (см., например, фиг. 4 и пример 7).

Такой вектор может быть использован как вектор экспрессии для экспрессии слитых белков, содержащих кодируемые полипептиды, образующие случайный клубок, и биологически активные белки. Соответственно, кодируемый слитый белок содержит (1) полипептид, образующий случайный клубок, кодируемый нуклеотидной последовательностью с низкой повторяемостью, в сочетании с (2) биологически активным белком. Предпочтительно, полипептид, образующий случайный клубок, состоит из пролина, аланина и возможно серина. Типичный вектор представлен в SEQ ID NO: 56. В прилагаемых примерах продемонстрирован типичный способ обеспечения такого вектора или молекулы нуклеиновой кислоты (см., например, фиг. 6 и пример 7).

В данном способе по изобретению вектор, предложенный здесь, содержит нуклеотидную последовательность, кодирующую биологически активный белок, и на второй стадии в вектор вводят нуклеотидную последовательность, кодирующую полипептид с высоким содержанием РА. Для введения нуклеотидной последовательности, кодирующей последовательность с высоким содержанием пролина/аланина, в вектор, содержащий нуклеотидную последовательность, кодирующую биологически активный белок, могут быть использованы "липкие" концы, содержащие по меньшей мере один нуклеотидный триплет/кодон, кодирующий, например, аланин, пролин и/или возможно серин (см. выше). Соот-

ветственно, такой триплет или кодон может кодировать аминокислоту, рассматриваемую как часть последовательности с высоким содержанием пролина/аланина, в частности аланин.

В результате, способ и вектор, предложенные здесь, позволяют избежать введения дополнительных аминокислотных линкеров, которые могут быть введены при использовании обычных сайтов рестрикции. Таким образом, средства и способы, предложенные здесь, позволяют проводить "бесшовное" клонирование молекулы нуклеиновой кислоты по изобретению, содержащей (1) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, и (2) нуклеотидную последовательность, кодирующую биологически активный белок.

Специалистам в области молекулярной биологии известно множество подходящих векторов. Выбор подходящего вектора зависит от желаемой функции и включает плазмиды, космиды, вирусы, бактериофаги и другие векторы, обычно используемые в генной инженерии.

Предпочтительно вектор представляет собой плазмиду, более предпочтительно плазмиду на основе универсального вектора экспрессии для *E. coli* pASK37, pASK75 или pXL2.

Для конструирования различных плазмид могут быть использованы способы, хорошо известные специалистам в данной области, см., например, методики, описанные в Sambrook (2001), приведенной выше, и Ausubel (1989), приведенной выше. Типичные плазмидные векторы включают, например, pQE-12, плазмиды серии pUC, pBluescript (Stratagene), векторы экспрессии серии pET (Novagen) или pCRIPTOPO (Invitrogen), lambda gt11, pJOE, серию pBBR1-MCS, pJB861, pBSMuL, pBC2, pUCPKS, pTACT1. Типичные векторы, совместимые с экспрессией в клетках млекопитающих, включают векторную систему E-027 pCAG Kosak-Cherry (L45a), pREP (Invitrogen), pCEP4 (Invitrogen), pMCIneo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2neo, pBPV-1, pdBPVMMTneo, pRSVgpt, pRSVneo, pSV2-dhfr, pIZD35, кДНК вектор экспрессии Окаямы-Берга (Okayama-Berg) pcDV1 (Pharmacia), pRc/CMV, pcDNA1, pcDNA3 (Invitrogen), pcDNA3.1, pSPORT1 (GIBCO BRL), pGEMHE (Promega), pLXIN, pSIR (Clontech), pIRES-EGFP (Clontech), pEAK-10 (Edge Biosystems) pTriEx-Hygro (Novagen) и pCIneo (Promega). Неограничивающие примеры плазмидных векторов, подходящих для *Pichia pastoris*, включают, например, плазмиды pAO815, pPIC9K и pPIC3.5K (все от Invitrogen).

Обычно векторы могут содержать один или более чем один репликатор (*ori*) с системой наследования для клонирования или экспрессии, один или более чем один маркер для отбора у хозяина, например, антибиотикорезистентность, и одну или более чем одну экспрессионную кассету. Примеры подходящих репликаторов включают, например, полноразмерный ColE1, его усеченные варианты, такие как присутствующие в плаزمиде pUC, репликаторы вируса SV40 и фага M13. Неограничивающие примеры селективируемых маркеров включают ампициллин, хлорамфеникол, тетрациклин, канамицин, dhfr, gpt, неомицин, гигромицин, бластицидин или генетицин.

Кроме того, указанный вектор содержит регуляторную последовательность, функционально связанную с указанной нуклеотидной последовательностью или молекулой нуклеиновой кислоты, определенной здесь.

Кодирующая последовательность (последовательности), например, указанная нуклеотидная последовательность, кодирующая полипептид с высоким содержанием РА, присутствующая в векторе, может быть связана с элементом (элементами) контроля транскрипции и/или с последовательностями, кодирующими другие аминокислоты, с применением общепринятых способов. Такие регуляторные последовательности хорошо известны специалистам в данной области и включают, без ограничения, регуляторные последовательности, обеспечивающие инициацию транскрипции, внутренние сайты связывания рибосом (internal ribosomal entry site, IRES) и возможно регуляторные элементы, обеспечивающие терминацию транскрипции и стабилизацию транскрипта. Неограничивающие примеры таких регуляторных последовательностей, обеспечивающих инициацию транскрипции, включают промоторы, кодон инициации трансляции, энхансеры, инсуляторы и/или регуляторные элементы, обеспечивающие терминацию транскрипции. Другие примеры включают последовательности Козак и вставочные последовательности, фланкированные донорными и акцепторными сайтами для сплайсинга РНК, нуклеотидные последовательности, кодирующие секреторные сигналы, или, в зависимости от используемой системы экспрессии, сигнальные последовательности, способные направлять экспрессированный белок в клеточный компартмент или в культуральную среду.

Примеры подходящих промоторов включают, без ограничения, промотор цитомегаловируса (CMV), промотор SV40, промотор RSV (вируса саркомы Рауса), промотор lacZ, промотор β -актина цыпленка, промотор CAG (сочетание промотора β -актина цыпленка и предраннего энхансера цитомегаловируса), промотор человеческого фактора элонгации 1 α , промотор AOX1, промотор GAL1, промотор СаМ-киназы, промотор lac, trp или tac, промотор lacUV5, промотор T7 или T5, полиэдральный промотор вируса множественного ядерного полиэдроза *Autographa californica* (AcMNPV) или глобиновый интрон клеток млекопитающих и других животных. Одним примером энхансера является, например, энхансер SV40. Неограничивающие дополнительные примеры регуляторных элементов/последовательностей, обеспечивающих терминацию транскрипции, включают сайт поли-А SV40, сайт поли-А tk или полиэдральные сигналы полиаденилирования AcMNPV.

Кроме того, в зависимости от системы экспрессии, в кодирующую последовательность молекулы нуклеиновой кислоты, предложенной здесь, могут быть введены лидерные последовательности, способные направлять полипептид в клеточный компартмент или обеспечивать его секрецию в среду. Лидерную последовательность (последовательности) вводят в рамку с последовательностями инициации и терминации трансляции, и, предпочтительно, лидерная последовательность способна обеспечивать секрецию транслированного белка или его части в периплазматическое пространство или во внеклеточную среду. Подходящими лидерными последовательностями являются, например, сигнальные последовательности VAP (бактериальной щелочной фосфатазы), CTB (субъединицы В холерного токсина), DsbA, ENX, OmpA, PhoA, stII, OmpT, PelB, Tat (Twin-arginine translocation) у *E. coli* и сигнальные последовательности бычьего гормона роста, человеческого химотрипсина, человеческого фактора VIII, человеческого иммуноглобулина-каппа, человеческого инсулина, человеческого интерлейкина-2, люциферазы *Metridia* или *Vargula*, человеческого трипсиногена-2, инулиназы *Kluuyveromyces marxianus*, фактора конъюгации альфа-1 *Saccharomyces cerevisiae*, меллитина, человеческого азуроцидина и тому подобное у эукариотических клеток.

Векторы могут также содержать дополнительную экспрессируемую последовательность нуклеиновой кислоты, кодирующую один или более чем один шаперон, способствующий правильному сворачиванию белка.

Предпочтительно, вектор по настоящему изобретению представляет собой вектор экспрессии. Вектор экспрессии по настоящему изобретению способен обеспечивать репликацию и экспрессию молекулы нуклеиновой кислоты по изобретению, например, молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид с высоким содержанием пролина/аланина, и нуклеотидную последовательность, кодирующую биологически активный белок. В прилагаемых примерах был сконструирован вектор экспрессии, содержащий (1) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, и (2) биологически активный белок, такой как IL-1Ra; см. пример 6. Типичный вектор экспрессии, содержащий молекулу нуклеиновой кислоты, кодирующую полипептид, состоящий из пролина и аланина, показан в примере 10.

Подходящие бактериальные хозяева для экспрессии включают, например, штаммы, имеющие происхождение от *Escherichia coli*, JM83, W3110, KS272, TGI, BL21 (такой как BL21(DE3), BL21(DE3)PlysS, BL21(DE3)RIL, BL21(DE3)PRARE), Origami (K-12), Origami B or Rosetta. Методики модификации, PCR-амплификации и лигирования векторов описаны в Sambrook (2001), приведенной выше.

Кроме того, в качестве векторов могут также быть использованы бакуловирусные системы для экспрессии молекул нуклеиновой кислоты по изобретению в эукариотических системах экспрессии. В этих аспектах в качестве вектора экспрессии может быть использован вектор pFBDM. При трансформации клеток *E. coli* DH10 с использованием MultiBac введение бакуловирусной ДНК MultiBac опосредовано транспозиционной последовательностью Tn7 (Berger (2013) *J. Vis. Exp.* 77:50159, Fitzgerald (2006) *Nat. Methods.* 2006 3:1021-1032.). Амплификация и экспрессия вируса может быть проведена в клетках Sf21 (*Spodoptera frugiperda*) или High Five (*Trichoplusia ni*).

Молекулы нуклеиновой кислоты и/или векторы по изобретению, описанные выше, могут быть предназначены для введения в клетки, например, нехимическими способами (электропорация, сонопорация, оптическая трансфекция, электрический перенос генов, гидродинамическая доставка или естественная трансформация при приведении клеток в контакт с молекулой нуклеиновой кислоты по изобретению), химическими способами (фосфат кальция, DMSO, PEG, липосомы, DEAE-декстран, полиэтиленмин, нуклеофекция и так далее), способами на основе частиц (генная пушка, магнитофекция, импалефекция (impalefection)), способами на основе фаговых или фагмидных векторов и вирусными способами. Например, для доставки молекул нуклеиновой кислоты в целевую популяцию клеток могут быть использованы векторы экспрессии, имеющие происхождение от вирусов, таких как ретровирусы, вирус коровьей оспы, аденоассоциированный вирус, вирусы герпеса, вирус лихорадки леса Семлики или вирус папилломы крупного рогатого скота.

Предпочтительно молекулы нуклеиновой кислоты и/или векторы по изобретению предназначены для трансформации электрокомпетентных *E. coli* электропорацией или для стабильной трансфекции клеток CHO с использованием фосфата кальция, полиэтиленмина или липофектамина (Pham (2006) *Mol. Biotechnol.* 34:225-237; Geisse(2012) *Methods Mol. Biol.* 899:203-219; Hacker (2013) *Protein Expr. Purif.* 92:67-76).

Настоящее изобретение также относится к клетке-хозяину или хозяину, не являющемуся человеком, трансформированным вектором или молекулой нуклеиновой кислоты по настоящему изобретению. Следует понимать, что согласно настоящему изобретению "клетка-хозяин или хозяин, не являющийся человеком, трансформированный вектором по изобретению", относятся к клетке-хозяину или хозяину, не являющемуся человеком, содержащим вектор или молекулу нуклеиновой кислоты по изобретению. Клетки-хозяева для экспрессии полипептидов хорошо известны в данной области и включают прокариотические клетки, а также эукариотические клетки. Хозяин может быть выбран из группы, состоящей из бактерии, клетки млекопитающего, клетки водоросли, ресничной клетки, клетки дрожжей и растительной клетки.

Типичные бактерии включают *Escherichia*, *Corynebacterium* (*glutamicum*), *Pseudomonas* (*fluorescens*), *Lactobacillus*, *Streptomyces*, *Salmonella*, *Bacillus* (такие как *Bacillus megaterium* или *Bacillus subtilis*), или *Corynebacterium* (такие как *Corynebacterium glutamicum*). Наиболее предпочтительным бактериальным хозяином здесь является *E. coli*. Примером ресничного для использования здесь является *Tetrahymena*, например, *Tetrahymena thermophila*.

Типичные клетки млекопитающих включают клетки Hela, HEK293, HEK293T, H9, Per.C6 и Jurkat, мышинные клетки NIH3T3, NS0 и C127, клетки COS 1, COS 7 и CV1, перепелиные клетки QC1-3, мышинные L-клетки, клетки мышинной саркомы, клетки меланомы Боуэса (*Bowes*) и клетки яичника китайского хомячка (СНО). Наиболее предпочтительными клетками-хозяевами по настоящему изобретению, представляющими собой клетки млекопитающих, являются клетки СНО. Типичным хозяином для использования здесь является *Cricetulus*, например, *Cricetulus griseus* (китайский хомячок). Кроме того, предпочтительны клетки эмбриональной почки человека (НЕК).

Другими подходящими эукариотическими клетками-хозяевами являются, например, клетки дрожжей, таких как *Pichia pastoris*, *Kluuyveromyces lactis*, *Saccharomyces cerevisiae* и *Schizosaccharomyces pombe*, или куриные клетки, такие как клетки DT40. Клетками насекомых, подходящими для экспрессии, являются, например, клетки *Drosophila* S2, *Drosophila* Kc, *Spodoptera* Sf9 и Sf21 или *Trichoplusia* Hi5. Предпочтительными клетками водорослей являются клетки *Chlamydomonas reinhardtii* или *Synechococcus elongatus* и тому подобное. Типичным растением является *Physcomitrella*, например, *Physcomitrella patens*. Типичной растительной клеткой является растительная клетка *Physcomitrella*, например растительная клетка *Physcomitrella patens*.

Объем настоящего изобретения также включает первичные клетки или клеточные линии млекопитающих. Первичные клетки представляют собой клетки, полученные непосредственно из организма. Подходящими первичными клетками являются, например, мышинные эмбриональные фибробласты (MEF), первичные мышинные гепатоциты, кардиомиоциты и нервные клетки, а также мышинные мышечные стволовые клетки (сателлитные клетки), человеческие кожные и легочные фибробласты, человеческие эпителиальные клетки (клетки носового, трахеального, почечного, плацентарного, кишечного, бронхиального эпителия), человеческие секреторные клетки (из слюнных, сальных и потовых желез), человеческие эндокринные клетки (клетки щитовидной железы), человеческие жировые клетки, человеческие гладкомышечные клетки, человеческие поперечно-полосатые мышечные клетки, человеческие лейкоциты, такие как В-клетки, Т-клетки, НК-клетки или дендритные клетки и стабильные иммортализованные клеточные линии, имеющие происхождение от указанных клеток (например, клетки, иммортализованные hTERT или онкогенами). Подходящие среды и условия для культивирования описанных выше клеток-хозяев известны в данной области.

Клетки-хозяева могут, например, быть использованы для получения больших количеств молекулы нуклеиновой кислоты, предложенной здесь, кодируемого полипептида, предложенного здесь, и/или указанного конъюгата лекарственного средства, предложенного здесь. Следовательно, клетки-хозяева могут быть использованы для получения больших количеств молекулы нуклеиновой кислоты, предложенной здесь, полипептида, кодируемого молекулой нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид с высоким содержанием РА, и/или полипептида, кодируемого молекулой нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид с высоким содержанием РА и биологически активный белок.

Соответственно настоящее изобретение также относится к способу получения молекулы нуклеиновой кислоты или вектора, предложенных здесь, включающему культивирование хозяина или клетки-хозяина по изобретению в подходящих условиях и возможно выделение полученной молекулы нуклеиновой кислоты и/или вектора.

Кроме того, настоящее изобретение относится к способу получения полипептида, кодируемого молекулой нуклеиновой кислоты или нуклеотидной последовательностью, предложенной здесь, включающему культивирование хозяина или клетки-хозяина по изобретению в подходящих условиях и возможно выделение полученного полипептида.

Кроме того, настоящее изобретение относится к способу получения конъюгата лекарственного средства, включающему культивирование клетки-хозяина по изобретению в подходящих условиях и возможно выделение полученного конъюгата лекарственного средства. Предпочтительно настоящее изобретение относится к способу получения конъюгата лекарственного средства, кодируемого молекулой нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, и биологически активный белок.

Таким образом, в одном аспекте настоящее изобретение относится к способу получения конъюгата лекарственного средства, содержащего полипептид, кодируемый молекулой нуклеиновой кислоты, определенной здесь, и дополнительно содержащего (1) биологически активный белок и/или (2) низкомолекулярное соединение, и/или (3) углевод, дополнительно включающему культивирование хозяина или клетки-хозяина, предложенных здесь, и возможно выделение полученного полипептида и/или конъюгата лекарственного средства. Например, если конъюгат лекарственного средства представляет собой слитый белок, содержащий полипептид, кодируемый молекулой нуклеиновой кислоты, определенной здесь, и

дополнительно содержащий биологически активный белок, способ может дополнительно включать культивирование хозяина или клетки-хозяина, предложенных здесь (то есть хозяина или клетки-хозяина, содержащих нуклеиновую кислоту, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, предложенный здесь, и нуклеиновую кислоту, кодирующую биологически активный белок, определенный здесь, в частности терапевтически активный белок), и возможно выделение полученного слитого белка (конъюгата лекарственного средства). Разумеется, если конъюгат лекарственного средства представляет собой слитый белок, способ может также включать культивирование хозяина или клетки-хозяина, предложенных здесь (то есть хозяина или клетки-хозяина, содержащих нуклеиновую кислоту, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, предложенный здесь), и/или культивирование хозяина или клетки-хозяина, содержащих нуклеиновую кислоту, кодирующую биологически активный белок, определенный здесь, в частности терапевтически активный белок, и возможно выделение полученного полипептида, состоящего из пролина, аланина и возможно серина, и/или выделение полученного биологически активного белка, и возможно дальнейшую конъюгацию полипептида и биологически активного белка (например, химическим сочетанием) с получением конъюгата лекарственного средства.

Например, если конъюгат лекарственного средства представляет конъюгат полипептида, кодируемого молекулой нуклеиновой кислоты, определенной здесь (то есть нуклеиновой кислотой, кодирующей полипептид, состоящий из пролина, аланина и возможно серина, предложенный здесь), с низкомолекулярным соединением и/или углеводом, способ может дополнительно включать культивирование хозяина или клетки-хозяина, предложенных здесь (то есть хозяина или клетки-хозяина, содержащих нуклеиновую кислоту, кодирующую полипептид, состоящий из пролина, аланина и возможно серина, предложенный здесь), и возможно выделение полученного полипептида, и возможно дальнейшую конъюгацию полипептида с низкомолекулярным соединением и/или углеводом (например, химическим сочетанием).

В данном контексте "культивирование хозяина или клетки-хозяина" включает экспрессию полипептида, определенного здесь, и/или биологически активного белка в хозяине или клетке-хозяине.

В прилагаемых примерах продемонстрировано, что такая молекула нуклеиновой кислоты, содержащая (1) нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина, и (2) биологически активный белок, такой как IL-1Ra, может экспрессирована в бактериях с последующей очисткой (см. пример 8 и фиг. 7). Кроме того, здесь показана возможность экспрессии и очистки полипептида, состоящего из пролина и аланина, кодируемого молекулой нуклеиновой кислоты, предложенной здесь (см., например, пример 11 и фиг. 8). Конъюгация кодируемого полипептида, состоящего из пролина, аланина и возможно серина, с низкомолекулярным лекарственным средством, углеводом и/или биологически активным белком, позволяет увеличить период полувыведения из плазмы и/или растворимость низкомолекулярного соединения/низкомолекулярного лекарственного средства и уменьшить неспецифическую токсичность, а длительное воздействие активного лекарственного средства на целевые клетки или структуры в организме может привести к улучшению фармакодинамики.

Вектор, присутствующий в хозяине по изобретению, представляет собой вектор экспрессии или опосредует стабильную интеграцию молекулы нуклеиновой кислоты по настоящему изобретению в геном клетки-хозяина, обеспечивающую экспрессию белка. Средства и способы отбора клетки-хозяина, в которую была успешно введена молекула нуклеиновой кислоты по настоящему изобретению, обеспечивая экспрессию белка, хорошо известны в данной области и были описаны ранее (Browne (2007) *Trends Biotechnol.* 25:425-432; Matasci (2008) *Drug Discov. Today: Technol.* 5:e37-e42; Wurm (2004) *Nat. Biotechnol.* 22:1393-1398).

Подходящие условия культивирования прокариотических или эукариотических клеток-хозяев хорошо известны специалисту в данной области. Например, бактерий, таких как, например, *E. coli*, можно культивировать с аэрацией в среде Лурия-Бертани (LB), обычно при температуре от 4 до примерно 37°C. Для повышения выхода и растворимости продукта экспрессии среда может быть забуферена или дополнена подходящими добавками, о которых известно, что они повышают выход и растворимость. В тех случаях, когда молекулу нуклеиновой кислоты по изобретению в векторе, присутствующем в клетке-хозяине, контролирует индуцируемый промотор, экспрессию полипептида можно индуцировать добавлением подходящего индуктора, такого как, например, изопропил-β-D-тиогаляктопиранозид (IPTG) или ангидротетрациклин (аТс), использованные в прилагаемых примерах. Подходящие протоколы и способы экспрессии описаны в данной области, например, в Sambrook (2001), приведенной выше (Gebauer (2012) *Meth. Enzymol.* 503:157-188), и могут, по необходимости, быть адаптированы к потребностям конкретных клеток-хозяев и требованиям по экспрессируемому белку.

В зависимости от типа клеток и конкретных требований по ним, культивирование клеток млекопитающих может, например, быть проведено в среде RPMI, Williams' E или среде DMEM, содержащей 10% (об./об.) FCS, 2 mM L-глутамин и 100 ед/мл пенициллина/стрептомицина. Клетки можно держать, например, при 37°C, или при 41°C для куриных клеток DT40, в атмосфере с 5%-м CO₂ и высокой влажностью. Подходящей средой для культивирования клеток насекомых является, например TNM + 10% FCS, SF900 или среда NuClone SFX-Insect. Клетки насекомых обычно выращивают при 27°C в виде адгезионных или суспензионных культур. Подходящие протоколы экспрессии для эукариотических клеток или

клеток позвоночных хорошо известны специалисту в данной области и могут быть взяты, например, из Sambrook (2001) (см. выше).

Предпочтительно способ получения молекулы нуклеиновой кислоты, вектора, полипептида и/или конъюгата лекарственного средства по изобретению осуществляют с использованием бактериальных клеток, таких как, например, клетки *E. coli*, или клеток млекопитающих, таких как, например, клетки СНО. Более предпочтительно, способ осуществляют с использованием клеток *E. coli* или клеток СНО, и, наиболее предпочтительно, способ осуществляют с использованием клеток *E. coli*.

Способы выделения полученных кодируемых полипептидов включают, без ограничения, такие стадии очистки, как аффинная хроматография (предпочтительно с использованием слитой метки, такой как Strep-метка II или His₆-метка), гель-фильтрация (эксклюзионная хроматография), анионообменная хроматография, катионообменная хроматография, хроматография гидрофобного взаимодействия, высокоэффективная жидкостная хроматография (ВЭЖХ), ВЭЖХ с обращенной фазой, осаждение сульфатом аммония или иммунопреципитация. Эти способы хорошо известны в данной области и описаны в общих чертах, например, в Sambrook (2001), приведенной выше, а также описаны в прилагаемых примерах, см., например, примеры 8 и 11. Такие способы обеспечивают получение по существу чистых полипептидов. Однородность таких чистых полипептидов предпочтительно составляет по меньшей мере примерно 90-95% (на уровне белка), более предпочтительно по меньшей мере примерно 98-99%. Наиболее предпочтительно, эти чистые полипептиды являются подходящими для фармацевтического использования/применения. Кроме того, здесь предусмотрено применение в пищевой или косметической промышленности. В зависимости от клетки-хозяина/организма, используемых в способе получения, кодируемые полипептиды по настоящему изобретению могут быть гликозилированными или могут быть негликозилированными. Предпочтительно, полипептид, состоящий из пролина, аланина и возможно серина, кодируемый молекулой нуклеиновой кислоты по изобретению, не гликозилирован. Более предпочтительно боковые цепи полипептида, состоящего из пролина, аланина и возможно серина, кодируемого молекулой нуклеиновой кислоты по изобретению, не подвергаются посттрансляционной модификации, такой как, например, гидроксильное гликозилирование пролина.

Кодируемый полипептид, образующий случайный клубок, состоит преимущественно из аланиновых, пролиновых и возможно сериновых остатков, в то время как серин, треонин или аспарагин, необходимые для O- или N-гликозилирования предпочтительно отсутствуют. Таким образом, получение самого полипептида или биологически активного белка, содержащего кодируемый полипептид, образующий случайный клубок, может приводить к монодисперсному продукту, предпочтительно лишенному посттрансляционных модификаций в Pro/Ala/Ser или Pro/Ala последовательности. Это является преимуществом в случае получения рекомбинантного белка в эукариотических клетках, таких как клетки яичника китайского хомячка (СНО), клетки НЕК или дрожжи, которых часто выбирают для биосинтеза сложных белков.

Изобретение также относится к способу получения конъюгата лекарственного средства, содержащего полипептид, кодируемый молекулой нуклеиновой кислоты по изобретению, предложенный здесь, и дополнительно содержащего (1) биологически активный белок, и/или (2) низкомолекулярное соединение, и/или (3) углевод. Такие конъюгаты с углеводами могут быть особенно полезны в качестве вакцин.

Как описано выше, конъюгат лекарственного средства, содержащий полипептид с высоким содержанием РА и биологически активный белок, может быть получен экспрессией молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, кодирующую полипептид с высоким содержанием РА, и нуклеотидную последовательность, кодирующую биологически активный белок. Экспрессированный конъюгат лекарственного средства может быть выделен. Альтернативно, конъюгат лекарственного средства может быть получен культивированием/выращиванием хозяина, содержащего нуклеотидную последовательность или молекулу нуклеиновой кислоты, кодирующую указанный полипептид, состоящий из пролина, аланина и возможно серина. Таким образом, проводят экспрессию молекулы нуклеиновой кислоты у хозяина. Возможно, указанный пептид выделяют. Полученный полипептид, состоящий из пролина, аланина и возможно серина, может быть конъюгирован с биологически активным белком, например, через пептидную связь или непептидную связь.

В иллюстративных примерах продемонстрировано, что полипептид с высоким содержанием РА, кодируемый молекулой нуклеиновой кислоты, предложенной здесь, может быть экспрессирован в бактериях и очищен от них (см., например, пример 11 и фиг. 8). В частности, было показано, что в полученном полипептиде неожиданным образом происходит отщепление начального метионина (N-концевого метионина) и, таким образом, в полученном полипептиде он отсутствует (см., например, пример 12 и фиг. 8D). Отсутствие начального метионина в полученном полипептиде позволяет проводить "бесшовную" конъюгацию первичной аминокислоты последующей N-концевой аминокислоты (после начального метионина) с биологически активным белком, низкомолекулярным соединением и/или углеводом. Таким образом, молекулы нуклеиновой кислоты по изобретению и полученные кодируемые ими полипептиды особенно предпочтительны для конъюгации, например, с биологически активными белками.

В частности, возможна сайт-специфичная конъюгация биологически активного белка, например, в присутствии активатора, такого как N-(3-диметиламинопропил)-N'-этилкарбодимид (EDC) или N-

гидроксисукцинимидный (NHS) эфир (Hermanson (1996) *Bioconjugate Techniques*, 1st edition, Academic Press, San Diego, CA), с N-концом полученного полипептида, образующего случайный клубок. Альтернативно, биологически активный белок может быть конъюгирован сайт-специфичным образом с C-концом полученного полипептида, образующего случайный клубок, состоящего из пролина, аланина и возможно серина, например, в присутствии активатора, такого как EDC, или после активации, например, NHS-эфиром.

Кроме того, полученный полипептид, состоящий из пролина, аланина и возможно серина, может быть конъюгирован с низкомолекулярным соединением и/или с углеводом через непептидную связь. Непептидные связи, которые могут быть использованы для поперечного сшивания белков, известны в данной области и могут включать дисульфидные связи, например, между боковыми цепями и/или тиоловыми группами двух Cys, тиоэфирные связи и амидные связи между карбоксильными группами и аминокетто-группами. Непептидные ковалентные связи могут также быть образованы с использованием химических поперечно сшивающих агентов, таких как дисукцинимидилсуберат (DSS), N-β-малеимидопропилкарбонилсукцинимидный эфир (BMPS) или сульфосукцинимидил-4-[n-малеимидо-фенил]бутират (сульфо-SMPB), металлохелатирующих/комплексообразующих групп, а также нековалентных белок-белковых или белок-пептидных взаимодействий.

Кроме того, с полипептидом, образующим случайный клубок, может быть сайт-специфичным образом конъюгировано низкомолекулярное лекарственное средство. Возможно, N-конец полипептида может быть модифицирован с использованием подходящей защитной группы, например ацетильной группы или пироглутамильной группы, и после активации C-концевой карбоксильной группы, например, с использованием обычных реагентов EDC и NHS (Hermanson (1996), см. выше), может быть проведено сайт-специфичное сочетание лекарственного средства с C-концом полипептида, образующего случайный клубок. Таким образом могут быть легко получены однородные конъюгаты лекарственных средств.

В качестве альтернативы сайт-специфичной конъюгации по единственному сайту, полипептид, образующий случайный клубок, состоящий из пролина, аланина и возможно серина, может быть снабжен дополнительными боковыми цепями, на N- или C-конце или во внутренней части, подходящими для химической модификации, такими как лизиновые остатки с их ε-аминогруппами, цистеиновые остатки с их тиоловыми группами или даже искусственные аминокислоты, что позволит конъюгировать с ним одну, две или множество молекул низкомолекулярного соединения, используя, например, NHS или малеимидные активные группы.

Помимо стабильной конъюгации возможно временное связывание пролекарства с полипептидом, образующим случайный клубок. Связь может быть разработана таким образом, чтобы обеспечить ее предсказуемое расщепление *in vivo*, либо посредством ферментативного механизма, либо посредством медленного гидролиза, начинающегося при физиологическом pH, аналогично тому, как, например, малорастворимый противоопухолевый агент камптотецин был конъюгирован с полимером ПЭГ, что улучшило биораспределение, снизило токсичность, повысило эффективность и обеспечило его накопление в опухоли (Conover (1998) *Cancer Chemother. Pharmacol.* 42:407-414). Примерами других пролекарств являются такие химиотерапевтические агенты, как доцетаксел (Liu (2008) *J. Pharm. Sci.* 97:3274-3290), доксорубин (Veronese (2005) *Bioconjugate Chem.* 16: 775-784) или паклитаксел (Greenwald (2001) *J. Control Release* 74:159-171).

Здесь также предполагают, что низкомолекулярное соединение может быть конъюгировано со слитым белком, например полипептидом, образующим случайный клубок, состоящим из пролина, аланина и возможно серина, который генетически слит с доменом, обеспечивающим направленное действие, например фрагментом антитела, обеспечивая посредством этого специфичную доставку низкомолекулярного соединения. Например, иммунотоксин, полученный в таком случае конъюгацией с цитотоксическим низкомолекулярным соединением, особенно полезен, если домен, обеспечивающий направленное действие, направлен против рецептора клеточной поверхности, который подвергается интернализации.

При использовании здесь термин "лекарственное средство" относится к низкомолекулярному соединению, биологически активному белку, пептиду или углеводу. При использовании здесь термин "низкомолекулярное соединение" может относиться к (органическому) соединению с низкой молекулярной массой (менее 900 дальтон). Низкомолекулярные соединения способствуют регуляции биологического процесса и обычно имеют размер нанометрового порядка. Здесь предполагают, что низкомолекулярное соединение используется в способе лечения или диагностики или в пищевой или косметической промышленности. Например, конъюгат лекарственного средства с полученным полипептидом, кодируемым нуклеотидной последовательностью или молекулой нуклеиновой кислоты, предложенной здесь, может содержать низкомолекулярное соединение (соединения), выбранное из группы, состоящей из ингибиторов ангиогенеза, противоаллергических лекарственных средств, противорвотных лекарственных средств, антидепрессантов, антигипертензивных лекарственных средств, противовоспалительных лекарственных средств, противомикробных лекарственных средств, антипсихотических лекарственных средств, антипролиферативных (цитотоксических и цитостатических) лекарственных средств, антагонистов кальция и других лекарственных средств, действующих на органы кровообращения, холинергических агонистов, лекарственных средств, действующих на центральную нервную систему, лекарственных

средств, действующих на дыхательную систему, стероидов, антисмысловых нуклеиновых кислот, малых интерферирующих РНК (siRNA), ингибиторов микро-РНК (miR), миметиков микро-РНК, ДНК-аптамеров и РНК-аптамеров.

Типичные ингибиторы ангиогенеза включают, без ограничения, ингибиторы MetAP2 (такие как фумагиллин, производные фумагилина, 2-{3-[3,5-бис[4-нитробензилиден]-4-оксопиперидин-1-ил]-3-оксопропилсульфанил}-этансульфоновая кислота), ингибиторы VEGFR (такие как акситиниб, бриваниб, кабозантиниб, тивозаниб и мотезаниб), ингибиторы плацентарного фактора роста (PlGF), ингибиторы рецептора тромбоцитарного фактора роста (такие как AC 710, сорафениб, сунитиниб и ваталаниб) и тому подобное.

Типичные противоаллергические лекарственные средства включают, без ограничения, антигистаминные средства (такие как дифенгидрамин (бенадрил), дименгидринат (драмамин, дриминат (dominate)), гидрохлорид гидроксизина (ресталл (restall), вистакот (vistacot)), прометазин (фенерган)) и тому подобное.

Типичные антидепрессанты включают, без ограничения, гранисетрон, палоносетрон и тому подобное.

Типичные антидепрессанты включают, без ограничения, цис-флупентиксол, гидрохлорид имипрамина, миансерин и тому подобное.

Типичные антигипертензивные лекарственные средства включают, без ограничения, алпростадил, diaзоксид, нитроглицерин и тому подобное.

Типичные противовоспалительные лекарственные средства включают, без ограничения, кортизон, гиалуроновую кислоту, кеторолак и тому подобное.

Типичные противомикробные лекарственные средства включают, без ограничения, аминогликозиды, амадовир (amadovir), амоксициллин, ампициллин, бензилпенициллин, карбапенемы, цефалоспорины, цефтиофул, хлорамфеникол, цефепим, цефтазидим, цефтобипрол, клиндамицин, драксин, далбаванцин, даптомицин, дигидрострептомицин, эритромицин, флорфеникол, фторхинолоны, флуниксин меглумин, линезолид, марбофлоксацин, микафунгин, нитрофуразон, оритаваксин, окситетрациклин, пенициллин, пиперациллин, прокаин, рупинтривир, спирамицин, стрептограмин, сульфадиметоксин, сульфаметазин, тедизолид, телаванцин, тикарциллин, тилмикозин, тигециклин, тилдипирозин, тилозин, ванкомицин и тому подобное.

Типичные антипсихотические лекарственные средства включают, без ограничения, амисульприд, арипразол, бенперидол, бромперидол, клопромазин, хлорпротиксен, клопентиксол, клозапин, флупентиксол, флуфеназин, флуспирилен, галоперидол, левомепромазин, мелперон, оланзапин, перазин, перфеназин, пимозид, пипамперон, промазин, прометазин, протипендил, кветиапин, рисперидон, сульпирид, тиоридазин, трифлуоперазин, трифлупромазин, зуклопентиксол и тому подобное.

Типичные противоопухолевые лекарственные средства включают, без ограничения, антрациклины (такие как доксорубин, эпирубин, идарубин и даунорубин), алкилирующие агенты (такие как калихеамицины, дактиномицины, митромицины и пиролобензодиазепины), ингибиторы АКТ (такие как АТ7867), аматоксины (такие как α -аманитины, β -аманитины, γ -аманитины, ϵ -аманитины, амануллин, амануллиновая кислота, аманинамид, аманин и проаманулин, SN-38 и камптотецин), ингибиторы АТМ, ауристатины (такие как ауристатин ЕВ (АЕВ), ауристатины ЕFP (АЕFP), монометилауристатин Е (ММАЕ), монометилауристатин F (ММАF), ауристатин F и доластатин), криптофицины, ингибиторы циклинзависимых киназ (такие как BMS-387032, PD0332991, GSK429286, AZD7762; AZD 1152, MLN8054 и MLN8237; BI2536, BI6727, GSK461364, ON-01910, SB 743921, SB 715992, MK-0731, AZD8477, AZ3146 и ARRY-520), дуокармицины, ингибиторы DNA-РК, эпотилоны (такие как эпотилон А, В, С, D, Е или F и их производные), ингибиторы GSK-3, ингибиторы FIDAC (такие как белинонат, CUDC-101, дроксинонат (droxinostat), ITF2357, JNJ-26481585, LAQ824 и панобинонат, MC1568, моцетинонат, энтинонат, PCI-24781, пироксамид, трихостатин А и воринонат), ингибиторы hsp70, ингибиторы hsp90 (такие как производные 17AAG, BI1021, BI1028, SNX-5422, NVP-AUY-922, KW-2478 и гелданамицин), ингибиторы сигнального пути MAPK (такие как MEK, Racs, INK, B-Raf), майтанзиноиды, аналоги майтанзиноидов (такие как майтанзинол, аналоги майтанзинола, майтанзин, DM-1 и DM-4), ингибиторы p38MAPK (такие как GDC-0973, GSK1 120212, MSC1936369B, AS703026, R05126766 и R04987655, PD0325901, AZD6244, AZD8330, GDC-0973, CDC-0879, PLX-4032, SB590885, BIRB796, LY2228820, SB 202190, AEE788, BIBW2992, афатиниб, лапатиниб, эрлотиниб и гефитиниб), ингибиторы PARP (такие как инипариб, олапариб, велипариб, AG014699, CEP 9722, MK 4827, KU-0059436, LT-673, 3-аминобензамид, А-966492 и AZD2461), ингибиторы PDK-1, соединения платины (такие как цисплатин, карбоплатин, оксалиплатин, ипроплатин, ормаплатин (ormaplatin) или тетраплатин), таксаны (такие как паклитаксел и доцетаксел), тубулизины (такие как тубулизин А, тубулизин В и производные тубулизинов), алкалоиды барвинка (такие как винбластин, виндезин и навельбин), ингибиторы сигнального пути Wnt/Hedgehog (такие как висмодегид, GDC-0449, циклопамин и XAV-939) и тому подобное.

Типичные лекарственные средства, действующие на центральную нервную систему, включают, без ограничения, бупренорфин, криостатин (cryostatин), налтрексон, налоксон и тому подобное.

Типичные витамины включают, без ограничения, витамин В12 (цианкобаламин), витамин А и тому

подобное.

Типичные стероиды включают, без ограничения, андрогенные стероиды (такие как флуоксиместерон, метилтестостерон, тестостерон, тренболон), эстрогены (такие как бета-эстрадиол, диэтилстилбестрол, эстрон, эстриол, эквилин, эстроипат, эквилин, местранол), прогестагенные соединения (такие как 19-норпрогестерон, альфапростол, хлормадинон, демегестон, дигидрогестерон, диметистерон, этистерон, диацетат этинодиола, норэтиндрон, ацетат норэтиндрона, медроксипрогестерон, меленгестрол, прогестерон, норгестрел, промегестон, зеранол) и тому подобное.

Типичные антисмысловые нуклеиновые кислоты включают, без ограничения, антисмысловые нуклеиновые кислоты, оказывающие направленное действие на рецептор андрогенов (такие как ISIS-AR, AZD5312), ангиопоэтин-подобный белок 3 (такие как ISIS-ANGPTL3), аполипопротеин B100 (такие как мипомерсен), аполипопротеин СIII (такие как ISIS-АРОСIII, воланесорсен), фактор роста соединительной ткани (CTGF) (такие как EXC 001, PF-06473871), кластерин (такие как кустирсен, OGX-011), С-реактивный белок (такие как ISIS-CRP), диацилглицеролацилтрансферазу (такие как ISIS-DGAT2), фактор VII (такие как ISIS-FVII), рецептор фактора роста фибробластов 4 (такие как ISIS-FGFR4), гепсидин (такие как XEN701), Hsp27 (такие как апаторсен (apatorsen), OGX-427), ген HTT (такие как ISIS-HTT), ICAM-1 (такие как аликафорсен), прекалликреин (такие как ISIS-PKK), SMN2 (такие как ISIS-SMN), STAT3 (такие как ISIS-STAT3-2.5, AZD9150), ген транстиретина (такие как ISIS-TTR) и тому подобное.

Типичные малые интерферирующие РНК (siRNA) включают, без ограничения, siRNA, оказывающие направленное действие на Z-ААТ-мутант альфа-1-антитрипсина (такие как ALN-ААТ), аминолевулинатсинтазу-1 (ALAS-1) (такие как ALN-AS1, ALN-AS2), антитромбин III (такие как ALN-AT3), компонент комплемента C5 (такие как ALN-CC5), компонент комплемента C6 (такие как ALN-CC6), фактор роста соединительной ткани (такие как RXI-109), экзон 8 гена дистрофина (такие как SRP-4008), экзон 44 гена дистрофина (такие как SRP-4044), экзон 45 гена дистрофина (такие как SRP-4045), экзон 50 гена дистрофина (такие как SRP-4050), вирус лихорадки Эбола (такие как AVI-7537), экзон 51 гена дистрофина (такие как этеплирсен, AVI-4658), экзон 52 гена дистрофина (такие как SRP-4052), экзон 53 гена дистрофина (такие как SRP-4053), вирус гриппа (такие как AVI-7100), белок веретена деления кинезин (KSP), болезни легких (такие как Atu111), вирус Марбург (такие как AVI-7288), смеси малых интерферирующих РНК (siRNA), оказывающие направленное действие на множество мишеней (такие как STP503, STP523, STP601, STP702, STP705, STP801, STP805, STP900, STP902, STP911, STP916, siPOOLS), нуклеокапсид N вирусного генома (такие как ALN-RSV01), PCSK9 (такие как ALN-PCS01, ALN-PCSsc), проапоптотический белок каспазу-2 (такие как QPI-1007), проапоптотический белок p53 (такие как QPI-1002), RTP801 (такие как PF-655), SERPINC1 (такие как ALN-AT4), трансмембранную сериновую протеазу-6 (Tmprss6) (такие как ALN-TMP), транстиретин (такие как ALN-TTRsc, ALN-TTR02), PCSK10 (такие как ALN-PCS02), PKN3 (такие как Atu027), фактор роста эндотелия сосудов (VEGF) (такие как ALN-VSP) и тому подобное.

Типичные ингибиторы микро-РНК включают, без ограничения, ингибиторы miR-10b, miR-15, miR-21, miR-29, miR-33, miR-92, miR-145, miR-195, miR-208, miR-221, miR-451, miR-499 и тому подобное.

Типичные миметики микро-РНК включают, без ограничения, аналог miR-34 (такой как MRX34), miR-Rx06, miR-Rx07, miR-Rx16, аналог let7 (такой как miR-Rxlet-7) и тому подобное.

Типичные ДНК-аптамеры включают, без ограничения, ингибитор нуклеолина (такой как AS1411), ингибитор pGDF (такой как E10030), ингибитор тромбина (такой как NU172), ингибитор vWF (такой как ARC1779) и тому подобное.

Типичные РНК-аптамеры включают, без ограничения, ингибитор C5a (такой как NOX-D21 или ARC1905), ингибитор пептида, связанного с геном кальцитонина (такой как NOX-L41), ингибитор лиганда С-С-хемокина 2 (такой как NOX-E36), ингибитор CXCL12 (такой как NOX-A12), ингибитор глюкагона (такой как NOX-G16), антагонист гепсидина (такой как NOX-H94), агонист рецептора распознавания патогенов (такой как агонист RIG-I), ингибитор сфингозин-1-фосфата (такой как NOX-S93), антагонист VEGF (такой как NX1838) и тому подобное.

Типичные углеводы, потенциально применимые для изготовления вакцин, включают, без ограничения, углеводные эпитопы, с которыми специфично связываются лектины, О-антиген E. coli 0121, производные О-антигена E. coli 0121, Man9 из gp120 HIV-1, О-полисахариды Shigella flexneri 2a, капсульный полисахарид 5 Staphylococcus aureus, капсульный полисахарид 8 Staphylococcus aureus, опухолеассоциированные углеводные антигены (ТАСА) (такие как Tn-антигены (например, α -GalNAc-O-Ser/Thr), сиалиловые Tn-антигены (например, NeuAc α (2,6)-GalNAc α -O-Ser/Thr), антиген Томсена-Фриденрайха (Gal β 1-3GalNAc α 1), LewisY (например, Fuc α (1,2)-Gal β (1,4)-[Fuc α (1,3)]-GalNAc), сиалил-LewisX и сиалил-LewisA, LewisX (стадиеспецифический эмбриональный антиген-1/SSEA-1), антиген Globo H (например, Fuc α (1,2)-Gal β (1,3)-GalNAc β (1,3)-Gal α (1,4)-Gal β (1,4)-Glc), Т-антиген (например, Gal β (1,3)-GalNAc α -O-Ser/Thr), гликофинголипидный стадиеспецифический эмбриональный антиген-3 (SSEA-3), сиалосодержащие гликофинголипиды, ганглиозид GD2, GD3, ганглиозид GM2, ганглиозид фукозил-GM и ганглиозид Neu5GcGM3) и тому подобное.

Конъюгат лекарственного средства, содержащий полипептид, кодируемый молекулой нуклеиновой

кислоты по изобретению, предложенной здесь, содержащий биологически активный белок, и/или низкомолекулярное соединение, и/или углеводов, может быть использован для лечения воспалительных заболеваний, инфекционных заболеваний, заболеваний дыхательной системы, эндокринных расстройств, заболеваний центральной нервной системы, заболеваний опорно-двигательной системы, заболеваний сердечно-сосудистой системы, онкологических заболеваний, заболеваний мочеполовой системы и метаболических заболеваний.

Типичные воспалительные заболевания включают, без ограничения, анкилозирующий спондилит, артрит, атеросклероз, атипичный гемолитико-уремический синдром (аГУС), фибромиалгию, синдром Гийена-Барре (СГБ), синдром раздраженного кишечника (СРК), болезнь Крона, колит, дерматит, дивертикулит, остеоартрит, псориаз, псориатический артрит, миастенический синдром Ламберта-Итона, системную красную волчанку (СКВ), нефрит, болезнь Паркинсона, рассеянный склероз, пароксизмальную ночную гемоглобинурию (ПНГ), ревматоидный артрит (РА), синдром Шегрена, язвенный колит и тому подобное.

Типичные инфекционные заболевания включают, без ограничения, африканский трипаносомоз, боррелиоз, холеру, криптоспоридиоз, лихорадку денге, гепатит А, гепатит В, гепатит С, HIV-инфекцию/СПИД, грипп, японский энцефалит, лейшманиоз, малярию, корь, менингит, онхоцеркоз, пневмонию, ротавирусную инфекцию, шистосомоз, сепсис, шигеллез, стрептококковый тонзиллит, туберкулез, брюшной тиф, желтую лихорадку и тому подобное.

Типичные заболевания дыхательной системы включают, без ограничения, астму, хроническую обструктивную болезнь легких (ХОБЛ), муковисцидоз и тому подобное.

Типичные эндокринные расстройства включают, без ограничения, акромегалию, диабет I, диабет II, гестационный диабет, болезнь Грейвса, дефицит гормона роста, гипергликемию, гиперпаратиреоз, гипертиреоз, гипогликемию, бесплодие, ожирение, заболевания паращитовидных желез, синдром Морquio А, мукополисахаридоз и тому подобное.

Типичные заболевания центральной нервной системы включают, без ограничения, болезнь Альцгеймера, каталепсию, болезнь Гентингтона, болезнь Паркинсона и тому подобное.

Типичные заболевания опорно-двигательной системы включают, без ограничения, остеопороз, мышечную дистрофию и тому подобное.

Типичные заболевания сердечно-сосудистой системы включают, без ограничения, острую сердечную недостаточность, цереброваскулярную болезнь (инсульт), ишемическую болезнь сердца и тому подобное.

Типичные онкологические заболевания включают, без ограничения, рак надпочечников, рак мочевого пузыря, рак молочной железы, колоректальный рак, эндометриальный рак, рак почки, острый лимфобластный лейкоз (ОЛЛ) и другие типы лейкозов, рак легкого, меланому, неходжкинскую лимфому, рак поджелудочной железы, рак предстательной железы, рак щитовидной железы и тому подобное.

Типичные заболевания мочеполовой системы включают, без ограничения, доброкачественную гиперплазию предстательной железы (ДГПЖ), гематурию, нейрогенный мочевой пузырь, болезнь Пейрони и тому подобное.

Типичные метаболические заболевания включают, без ограничения, болезнь Гоше, болезнь Фабри, дефицит гормона роста, синдром Гурлера, синдром Хантера, гипероксалурию, нейрональный цероид-липофуциноз, синдром Марото-Лами, синдром Морquio, синдром Нунан, галлонедостаточность гена SHOX, синдром Турнера, синдром Прадера-Вилли, фенилкетонурию, синдром Санфилиппо и тому подобное.

Как описано выше, молекула нуклеиновой кислоты, предложенная здесь, может также быть использована сама по себе или как часть вектора в целях генной терапии. Генная терапия, основанная на введении терапевтических генов в клетки методами *ex vivo* или *in vivo*, является одним из самых важных применений переноса генов. Подходящие векторы, способы или системы доставки генов для генной терапии *in vivo* описаны в литературе и известны специалисту в данной области.; см., например, Giordano (1996) *Nat. Med.* 2:534-539; Schaper (1996) *Circ. Res.* 79:911-919; Anderson (1992) *Science* 256:808-813; Verma (1997) *Nature* 389:239-249; Isner (1996) *Lancet* 348:370-374; Muhlhauser (1995) *Circ. Res.* 77:1077-1086; Onodera (1998) *Blood* 91:30-36; Verma (1998) *Gene Ther.* 5:692-699; Nabel (1997) *Ann. N.Y. Acad. Sci.* 811:289-292; Verzeletti (1998) *Hum. Gene Ther.* 9:2243-2251; Wang (1996) *Nat. Med.* 2:714-716; WO 94/29469; WO 97/00957, US 5580859; US 5589466; или Schaper (1996) *Curr. Opin. Biotechnol.* 7:635-640. Молекулы нуклеиновой кислоты и векторы, предложенные здесь, могут быть предназначены для введения в клетки непосредственно или посредством липосом или вирусных (например, аденовирусных, ретровирусных) векторов. Например, вектор может представлять собой аденоассоциированный вирусный (AAV) вектор, в частности вектор AAV8. AAV-векторы привлекательны с точки зрения генной терапии. Система AAV имеет ряд преимуществ, включая длительную экспрессию генов, неспособность к автономной репликации без вируса-помощника, трансдукцию делящихся и неделящихся клеток и отсутствие патогенности от инфекций дикого типа. Предпочтительно указанная клетка, в которую вводят молекулу нуклеиновой кислоты или вектор, представляет собой клетку зародышевой линии, эмбриональную клетку или яйцеклетку или имеет происхождение от указанных клеток, наиболее предпочтительно указанная клетка представляет собой стволовую клетку. Примером эмбриональной стволовой клетки может быть,

среди прочего, стволовая клетка, описанная в Nagy (1993) Proc. Natl. Acad. Sci. USA 90:8424-8428.

Термины "полипептид", "пептид" и "белок" использованы здесь взаимозаменяемо и относятся к полимеру из двух или более аминокислот, связанных амидными или пептидными связями, образованными между аминогруппой одной аминокислоты и карбоксильной группой другой аминокислоты. Предпочтительно пептидная связь образована между α -аминогруппой одной аминокислоты и α -карбоксильной группой другой аминокислоты. Аминокислоты, входящие в состав пептида или белка, которые также называют аминокислотными остатками, могут быть выбраны из 20 стандартных протеиногенных α -аминокислот (то есть Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr и Val), а также из непротеиногенных и/или нестандартных α -аминокислот (таких как, например, орнитин, цитруллин, гомолизин, пирролизин или 4-гидроксипролин) и, кроме того, β -аминокислот (например, β -аланина), γ -аминокислот и δ -аминокислот. Предпочтительно аминокислотные остатки, входящие в состав пептида или белка, выбраны из α -аминокислот, более предпочтительно из 20 стандартных протеиногенных α -аминокислот (которые могут присутствовать в форме L-изомеров или D-изомеров и предпочтительно все, за исключением Gly, присутствуют в форме L-изомеров).

Кодируемый полипептид или белок может быть немодифицированным или может быть модифицированным, например, на его N-конце, на его C-конце и/или по боковой цепи любого из аминокислотных остатков (в частности по функциональной группе боковой цепи одного или более Lys, His, Ser, Thr, Tyr, Cys, Asp, Glu и/или Arg остатков).

Такие модификации могут включать, например, присоединение любой из защитных групп, описанных для соответствующих функциональных групп в Wuts (2006) *Greene's protective groups in organic synthesis*, John Wiley & Sons, 4th edition, Hoboken, N.J. Такие модификации могут также включать гликозилирование и/или ацилирование одной или более чем одной жирной кислотой (например, одной или более чем одной C₈₋₃₀ насыщенной или мононенасыщенной жирной кислотой; с образованием пептида или белка, ацилированного жирной кислотой). Кодируемый полипептид предпочтительно не гидроксипролин, в частности не содержит гидроксипролин.

Аминокислотные остатки, входящие в состав кодируемого пептида или белка, могут, например, присутствовать в виде линейной молекулярной цепи (образуя линейный пептид или белок) или могут образовывать одно или более чем одно кольцо (что соответствует циклическому пептиду или белку), например, через пептидную, изопептидную или дисульфидную связь. Пептид или белок может также образовывать олигомеры, состоящие из двух или более идентичных или разных молекул. При использовании здесь термин "домен" относится к любой области/части аминокислотной последовательности, способной самостоятельно принимать определенную структуру и/или выполнять определенную функцию. Соответственно в контексте настоящего изобретения "домен" может представлять собой функциональный домен или структурный домен, который может, например, образовывать часть более крупного полипептида.

При использовании здесь термины "содержащий" и "включающий" или их грамматические варианты следует понимать как указывающие на заявленные признаки, объекты, стадии или компоненты, но не исключающие добавление одного или более чем одного дополнительного признака, объекта, стадии, компонента или их группы. Объем этих терминов включает термины "состоящий из" и "состоящий по существу из".

Таким образом, термины "содержащий"/"включающий"/"имеющий" указывают на возможность присутствия любого другого компонента (или, сходным образом, признаков, объектов, стадий и тому подобного). Таким образом, при использовании здесь терминов "содержащий"/"включающий"/"имеющий" они могут быть заменены на "состоящий по существу из" или предпочтительно на "состоящий из".

Термин "состоящий из" указывает на невозможность присутствия другого компонента (или, сходным образом, признаков, объектов, стадий и тому подобного).

При использовании здесь термин "состоящий по существу из" и его грамматические варианты следует понимать как указывающие на заявленные признаки, объекты, стадии или компоненты, но не исключающие добавление одного или более чем одного дополнительного признака, объекта, стадии, компонента или их группы, но только если дополнительные признаки, объекты, стадии, компоненты или их группы не приводят к существенным изменениям основных и новых характеристик заявленной композиции, устройства или способа.

Таким образом, термин "состоящий по существу из" указывает на возможность присутствия определенных других компонентов (или, сходным образом, признаков, объектов, стадий и тому подобного), а именно тех, которые не оказывают существенного влияния на основные характеристики композиции, устройства или способа. Иными словами, термин "состоящий по существу из" (который может быть использован здесь взаимозаменяемо с термином "по существу содержащий/включающий") допускает присутствие других компонентов в композиции, устройстве или способе, в дополнение к обязательным компонентам (или, сходным образом, присутствие других признаков, объектов, стадий и тому подобного), при условии, что присутствие других компонентов не оказывает существенного влияния на основные

характеристики устройства или способа.

Термин "способ/метод" относится к образу действий, средствам, методикам и процедурам выполнения заданной задачи, включая, без ограничения, образ действий, средства, методики и процедуры, известные специалистам в области химии, биологии и биофизики или легко разрабатываемые ими на основе известного образа действий, средств, методик и процедур.

Если не указано иное, при использовании здесь термин "примерно" предпочтительно относится к $\pm 10\%$ от указанного числового значения, более предпочтительно $\pm 5\%$ от указанного числового значения и, в частности, к точному указанному числовому значению.

Далее настоящее изобретение описано со ссылкой на следующие неограничивающие графические материалы и примеры. Если не указано иное, применяли общепринятые методы технологии генетической рекомбинации, описанные, например, в Sambrook (2001), приведенной выше и полностью включенной сюда посредством ссылки.

Далее настоящее изобретение описано со ссылкой на следующие неограничивающие графические материалы и примеры. На графических материалах представлено следующее.

Фиг. 1. Сборка нуклеиновых кислот с низкой повторяемостью, кодирующих последовательности с высоким содержанием пролина/аланина с использованием плазмиды pXL2

(а) Карта плазмиды pXL2 (SEQ ID NO: 48). Сайт распознавания SapI (5'-GCTCTTC-3') и сайт распознавания EarI (5'-CTCTTC-3') вводят в селективируемый маркерный ген lacZ, находящийся под транскрипционным контролем промотора/оператора lac (lac^{P₀}). Сайт рестрикции SapI позволяет проводить последовательное введение одного или более чем одного элемента (структурного блока) молекул нуклеиновой кислоты с низкой повторяемостью, кодирующих последовательности с высоким содержанием пролина/аланина. Разрезание с использованием EarI, рестриктазы, сайт распознавания которой перекрывается/является частью последовательности распознавания SapI, позволяет вырезать и/или выделять собранную молекулу нуклеиновой кислоты с низкой повторяемостью (генную кассету), кодирующую последовательность с высоким содержанием пролина/аланина. Основа плазмиды идентична основе универсального клонирующего вектора pUC19 (Yanisch-Perron (1985) Gene 33:103-119), за исключением удаления дополнительных сайтов рестрикции SapI и EarI посредством "молчащей" мутации. (В) Карта плазмиды pXL1 (SEQ ID NO: 55). Два сайта распознавания SapI (5'-GCTCTTC-3') в противоположной ориентации вводят в селективируемый маркерный ген lacZ, находящийся под транскрипционным контролем промотора/оператора lac (lac^{P₀}). Сайт рестрикции SapI обеспечивает возможность введения/клонирования, а также воспроизведения/амплификации и последующего вырезания и/или выделения молекулы нуклеиновой кислоты с низкой повторяемостью (генной кассеты), кодирующей последовательность с высоким содержанием пролина/аланина. Основа плазмиды идентична основе pUC19, за исключением удаления дополнительных сайтов рестрикции SapI посредством "молчащей" мутации. Следует отметить, что вводимая генная кассета имеет обратную, по сравнению с pXL2, ориентацию. (С) Участок нуклеотидной последовательности, расположенный внутри сайта клонирования pXL2 и рядом с ним, используемый при клонировании и сборке молекул нуклеиновой кислоты с низкой повторяемостью. Нуклеотидная последовательность содержит сайт рестрикции SapI и сайт рестрикции EarI в обратной ориентации. Поскольку последовательность распознавания EarI является также частью последовательности распознавания SapI, EarI разрезает оба сайта распознавания. Независимо от того, какая рестриктаза используется, SapI или EarI, расщепление приводит к образованию выступающих концов GCC/CGG (кодирующих Ala), совместимых с "липкими" концами GCC/CGG молекул нуклеиновой кислоты с низкой повторяемостью, кодирующих последовательности с высоким содержанием пролина/аланина, обеспечивая возможность эффективного лигирования.

Последовательности распознавания подчеркнуты. (D) Нуклеотидная последовательность и кодируемая аминокислотная последовательность сайта клонирования pXL2 после введения одного элемента/структурного блока нуклеиновой кислоты с низкой повторяемостью PAS#1b(200) (SEQ ID NO: 19). Отмечены сайты рестрикции SapI и EarI, фланкирующие клонированный элемент/структурный блок нуклеиновой кислоты с низкой повторяемостью (последовательности распознавания подчеркнуты). (E) Сборка элементов последовательности нуклеиновой кислоты с низкой повторяемостью с получением клонированных нуклеотидных последовательностей (молекул нуклеиновой кислоты) большей длины, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина. На первой стадии pXL2 расщепляют с использованием SapI, дефосфорилируют и лигируют с первым элементом последовательности, PAS#1b(200). На второй стадии полученную плазмиду раскрывают/линеаризуют выше от клонированного элемента последовательности однократным расщеплением рестриктазой SapI с последующим дефосфорилированием и лигированием со вторым элементом нуклеотидной последовательности, PAS#1c(200). Полученная плазида pXL2-PAS#1c/1b(400) содержит введенную генную/ДНК-кассету длиной в общей сложности 1200 пар оснований. В целом, полученная клонированная нуклеотидная последовательность с низкой повторяемостью, кодирующая повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, содержит только нуклеотидные повторы максимальной длиной 14 нуклеотидов (SEQ ID NO: 52). Вся собранная генная кассета/молекула нуклеиновой кислоты может быть легко вырезана расщеплением с использованием EarI и

затем использована для субклонирования в векторы экспрессии, например, в одной и той же рамке считывания с нуклеотидной последовательностью, кодирующей биологически активный белок и тому подобное. Следует отметить, что повторение второй стадии позволяет систематическим образом собирать генные кассеты все большей длины и клонировать их в рXL2. При использовании разных подходящих элементов нуклеотидной последовательности полученная длинная молекула нуклеиновой кислоты с низкой повторяемостью, кодирующая повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, содержит лишь небольшое число коротких нуклеотидных повторов.

Фиг. 2. Анализ повторов в последовательностях с высоким содержанием пролина/аланина

(а) Точечные диаграммы последовательности с высоким содержанием пролина/аланина по предшествующему уровню техники PA#3a(200) (SEQ ID NO: 15) в сравнении с точечными диаграммами нуклеотидной последовательности с низкой повторяемостью по настоящему изобретению PA#3b(200) (SEQ ID NO: 36), обе диаграммы получены с использованием инструмента для построения точечных диаграмм "dotup", включенного в пакет программного обеспечения Geneious V8.1 (Biomatters, Окленд, Новая Зеландия), с применением окна повтора 14 или 15. Сравнивая анализируемую нуклеотидную последовательность саму с собой по осям x и y, соответственно, двумерной диаграммы, инструмент "dotup" определяет все области, где присутствуют идентичные последовательности указанной длины (окно повтора), и отмечает их диагональными линиями, указывающими на положение повторяющихся последовательностей по осям x и y (кроме центральной диагональной линии, отражающей идентичность последовательности самой себе). Расположение повторов друг за другом приводит к продлению диагональной линии. Нуклеотидная последовательность PA#3a(200), проанализированная здесь, продемонстрировала высокую повторяемость, на что указывают многочисленные и/или длинные диагональные линии. В этом случае в нуклеотидной последовательности присутствует множество повторов по 60 пар оснований каждый. Применение окна повтора 14 позволяет выявить внутри повтора длиной 60 пар оснований (п.о.) еще более короткий повтор длиной 14 п.о. В отличие от этого нуклеотидная последовательность PA#3b(200) демонстрирует лишь небольшое число редких повторов длиной 14 п.о. на протяжении всей анализируемой нуклеотидной последовательности длиной 600 п.о., а применение несколько большего окна повтора 15 не позволяет выявить никаких повторов. (Б) Точечные диаграммы нуклеиновой кислоты PAS#1a(600) (SEQ ID NO: 12), раскрытой в WO 2008155134, кодирующей последовательность с высоким содержанием пролина/аланина, в сравнении с точечными диаграммами собранных нуклеотидных последовательностей с низкой повторяемостью, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению PAS#1d/1f/1c/1b(800) (SEQ ID NO: 39) и PAS#1f/1c/1b(600) (SEQ ID NO: 38) с применением окон повтора 14 и 15 пар оснований. В то время как нуклеотидная последовательность PAS#1a(600) демонстрирует повторы длиной 60 пар оснований, нуклеотидные последовательности PAS#1d/1f/1c/1b(800) и PAS#1f/1c/1b(600) не демонстрируют повторов или, в случае окна повтора 14 нуклеотидов, только один повтор длиной 14 нуклеотидов (диагональная линия) на всем протяжении анализируемых нуклеотидных последовательностей длиной 2400 или 1800 п.о., соответственно. (С) Анализ точечных диаграмм синтетической последовательности ДНК, кодирующей мультимер [(AP)₅]_n (SEQ ID NO: 16), раскрытой в US 2006/0252120, и естественной последовательности ДНК, кодирующей повторяющуюся область сверхкрупного оболочечного белка Масасине herpesvirus 1 (GenBank AAP41454.1) с высоким содержанием пролина/аланина (SEQ ID NO: 18), в сравнении с нуклеотидной последовательностью с низкой повторяемостью по настоящему изобретению PA#3a(200) (SEQ ID NO: 15) с применением окон повтора 14 и 15 нуклеотидов. Точечные диаграммы последовательностей ДНК, кодирующих мультимер [(AP)₅]_n и область сверхкрупного оболочечного белка Масасине herpesvirus 1 с высоким содержанием пролина/аланина, указывают на высокую повторяемость этих нуклеотидных последовательностей. В отличие от этого, точечная диаграмма нуклеотидной последовательности PA#1b(200) демонстрирует лишь небольшое число редких повторов длиной 14 нуклеотидов (окно повтора 14 нуклеотидов) или полное отсутствие повторов (окно повтора 15 нуклеотидов) на протяжении всей анализируемой нуклеотидной последовательности длиной 600 п.о.

Фиг. 3. Электрофореграмма ДНК-секвенирования молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующей аминокислотные последовательности с высоким содержанием пролина/аланина

Электрофореграмма ДНК-секвенирования рXL2-PAS#1f/1c/1b(600) (SEQ ID NO: 38), имеющей нуклеотидную последовательность с низкой повторяемостью, кодирующую последовательность с высоким содержанием пролина/аланина, по настоящему изобретению. Двупечечную плазмидную ДНК секвенировали с использованием праймера XLP-1 (SEQ ID NO: 3), специфично гибридизирующегося с кодирующей областью элемента нуклеотидной последовательности PAS#1b(200) (SEQ ID NO: 19). На данной электрофореграмме определено более 900 оснований, соответствующих известной нуклеотидной последовательности рXL2-PAS#1f/1c/1b(600). Следует отметить, что на электрофореграмме нет признаков неспецифической или множественной гибридизации праймера.

Фиг. 4. рASK75-PAS#1f/1c/1b(600)-IL1Ra, генетически стабильный вектор экспрессии для получения терапевтически значимого слитого белка PAS#1(600)-IL1Ra в бактериях

Карта плазмиды (а) рASK75-IL1Ra (SEQ ID NO: 49) и (Б) ее производного рASK75-PAS#1f/1c/1b(600)-IL1Ra (SEQ ID NO: 50) после введения генной кассеты PAS#1f/1c/1b(600). Структур-

ный ген биологически/фармакологически активного (пре)белка PAS#1(600)-IL1Ra, содержащий нуклеотидную последовательность с низкой повторяемостью, кодирующую полипептид PAS#1 из 601 аминокислотного остатка, и структурный ген человеческого IL-1Ra, а также кодирующие области бактериальной сигнальной последовательности OmpA и His₆-метки, клонируют, помещая под транскрипционный контроль промотора/оператора tet ($tet^{p/o}$). Основа плазмиды вне экспрессионной кассеты, фланкированной сайтами рестрикции XbaI и HindIII, идентична основе универсального вектора экспрессии pASK75 (Skerra (1994) Gene 151:131-135). Показан единственный сайт рестрикции SapI, оставшийся после введения молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, по настоящему изобретению.

Фиг. 5. Анализ генетической стабильности молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, PAS#1f/1c/1b(600) в сравнении с нуклеотидной последовательностью по предшествующему уровню техники PAS#1a(600)

Электрофорез в агарозном геле после XbaI/HindIII-рестрикционного анализа 10 препаратов плазмид из клонов *E. coli* JM83, трансформированных плаزمидами pASK75-PAS#1a(600)-IL1Ra (дорожки 1-5) (SEQ ID NO: 51) или pASK75-PAS#1f/1c/1b(600)-IL1Ra (дорожки 1-10) (SEQ ID NO: 50), культивированных на протяжении 7 суток, что соответствует примерно 70 поколениям бактериального клеточного деления. 4 из 5 проанализированных клонов pASK75-PAS#1a(600)-IL1Ra продемонстрировали укорочение фрагментов ДНК, содержащих введенную нуклеиновую кислоту, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина (фиг. 5А), что явно указывает на генетическую нестабильность. В отличие от этого все 5 клонов pASK75-PAS#1f/1c/1b(600)-IL1Ra продемонстрировали только ожидаемые полосы, соответствующие 3093 п.о. и 2377 п.о. соответственно, что указывает на интактность введенной нуклеиновой кислоты, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, и высокую генетическую стабильность плазмиды. Таким образом, нуклеотидные последовательности с низкой повторяемостью, кодирующие повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению имеют явное преимущество перед повторяющимися нуклеотидными последовательностями по предшествующему уровню техники.

Фиг. 6. "Бесшовное" клонирование нуклеотидной последовательности с низкой повторяемостью, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, в экспрессионную плазмиду, кодирующую биологически активный белок IL-1Ra

(а) Схематическое изображение кодирующей области слитого белка, содержащей ген сигнальной последовательности OmpA с последующим аланиновым кодоном GCC, первую последовательность распознавания SapI GCTCTTC на некодирующей цепи, динуклеотидный спейсер GC и вторую последовательность распознавания SapI в обратной ориентации с ее последовательностью распознавания GCTCTTC на кодирующей цепи и, в завершение, аланиновый кодон GCC и кодирующую последовательность зрелого IL-1Ra (UniProt ID P18510). Используя сайты рестрикции XbaI/HindIII, всю показанную последовательность клонировали в универсальный вектор экспрессии pASK75. (Б) Схематическое изображение ДНК-кассеты, описанной в (а), после расщепления с использованием SapI и вырезания короткой вставки длиной 24 п.о., фланкированной двумя сайтами рестрикции SapI. Следует отметить, что из-за двух "липких" 5'-концов только 18 нуклеотидов, которые расположены в середине, образуют двойную спираль ДНК и являются истинными парами оснований. (С) "Бесшовное" однонаправленное введение фрагмента ДНК, содержащего нуклеотидную последовательность с низкой повторяемостью PA#1b(200), например, вырезанную из pXL2-PA#1b(200) (SEQ ID NO: 54), с использованием совместимых "липких" концов GCC/CGG, полученных при расщеплении рестриктазой EarI (см. фиг. 1). Полученная экспрессионная кассета, содержащая нуклеотидную последовательность с низкой повторяемостью, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, по настоящему изобретению, показана ниже как SEQ ID NO: 47.

Фиг. 7. Анализ слитого белка PAS#1(600)-IL1Ra, полученного в *E. coli* с использованием генетически стабильного вектора экспрессии pASK75-PAS#1f/1c/1b(600)-IL1Ra

(а) Анализ слитого белка PAS#1(600)-IL1Ra, очищенного с применением IMAC, AEX, CEX и SEC, посредством 10% SDS-PAGE. На геле представлены два образца PAS#1(600)-IL1Ra по 2 мкг с восстановлением 2-меркаптоэтанолом (дорожка 1) и без восстановления (дорожка 2). Размеры маркерных белков (М) указаны слева. Слитый белок PAS#1(600)-IL1Ra виден как одна однородная полоса с кажущейся молекулярной массой более 116 кДа. Из-за слабого связывания с SDS слитые PAS-белки обычно демонстрируют значительно большие размеры (Schlapschy (2013) Protein Eng Des Sel. 26:489-501), чем, например, расчетная масса 68 кДа для PAS#1(600)-IL1Ra. (Б) Анализ слитого белка PAS#1(600) масс-спектрометрией с электрораспылительной ионизацией (Electrospray Ionisation Mass Spectrometry, ESI-MS). Развернутый ESI-MS-спектр очищенного слитого белка PAS#1(600)-IL1Ra демонстрирует измеренную массу 67994,8 Да, что почти идеально соответствует расчетной массе 67994,9 Да.

Фиг. 8. Анализ чистого полипептида PA#1(600), полученного в *E. coli* с использованием генетически стабильного вектора экспрессии pASK37-MP-PA#1d/1c/1b(600)

(а) Карта плазмиды рASK37-MP-PA#1d/1c/1b(600) (SEQ ID NO: 53). Нуклеотидную последовательность с низкой повторяемостью, кодирующую полипептид PA#1(600), клонировали, помещая под транскрипционный контроль промотора/оператора lacUV5 (lacUV5^{p/o}) после кодонов начального остатка Met и остатка Pro. Основа плазмиды вне экспрессионной кассеты, фланкированной сайтами рестрикции XbaI и HindIII, идентична основе универсального вектора экспрессии рASK37 (Skerra (1991) Protein Eng. 4:971-979). (В) Анализ рекомбинантного полипептида PA#1(600) посредством 10% SDS-PAGE с последующим окрашиванием йодидом бария. Внесенные образцы представляют собой лизированные клетки после 3 ч экспрессии (дорожка 1), белок, осажденный 20%-м, мас./об., (NH₄)₂SO₄ (дорожка 2), супернатант (NH₄)₂SO₄-преципитата, растворенный в 20 мМ трис/HCl, после центрифугирования в течение 20 мин при 17000 об/мин и последующей обработки 1%-м, об./об., AcOH (дорожка 3) и элюат, полученный при последующей субтрактивной катионообменной хроматографии полипептида PA#1(600) (дорожка 4). Полипептид PA#1(600) плохо связывается с SDS; поэтому полипептид PA#1(600) виден как красная/желтая окрашенная йодом полоса в концентрирующем геле, что указывает на однородность состава. (С) Гель, показанный на (В), после обесцвечивания водой и окрашивания кумасси бриллиантовым голубым. Полипептид PA#1(600) плохо связывается с SDS и не окрашивается красителем кумасси, поэтому на окрашенном кумасси геле видны только примеси (белки клеток-хозяев). (D) Анализ чистого полипептида PA#1(600) масс-спектрометрией с электрораспылительной ионизацией (ESI-MS). Развернутый спектр демонстрирует измеренную массу 48301,78 Да, что почти идеально соответствует расчетной массе рекомбинантного полипептида PA#1(600), несущего дополнительный остаток Pro на N-конце, как объяснено выше, и дополнительный остаток Ala на C-конце из-за сайта рестрикции SapI, использованного для клонирования гена (48301,4 Да). Следует отметить, что данный рекомбинантный полипептид больше не несет начального остатка Met, вероятнее всего, в результате внутриклеточной активности метионинаминопептидазы (Giglione (2015) Biochimie 114:134-46).

Фиг. 9. Автоматизированный анализ повторов в нуклеотидных последовательностях, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина

Повторы в природных, а также синтетических нуклеотидных последовательностях, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина, по предшествующему уровню техники анализировали, как описано в примере 13, с применением Python-сценария калькулятора NRS (см. пример 14) и сравнивали с нуклеотидными последовательностями с низкой повторяемостью PAS#1b(200) и PA#1e/1d/1c/1b(800) по настоящему изобретению. Частоту (число копий) всех повторов определенной длины в анализируемой нуклеотидной последовательности наносили на гистограмму против длины повтора, (а) Гистограмма нуклеотидной последовательности по предшествующему уровню техники PAS#1a(200) (SEQ ID NO: 11). (В) Гистограмма нуклеотидной последовательности PA#1a(200) (SEQ ID NO: 14). (С) Гистограмма нуклеотидной последовательности, кодирующей гликомодуль [(Ala-Pro)₅]₂₀APA (SEQ ID NO: 16). (D) Гистограмма нуклеотидной последовательности, кодирующей гликомодуль [AAPAPAPAP]₁₀AS (SEQ ID NO: 17). (Е) Гистограмма нуклеотидной последовательности, кодирующей участок с высоким содержанием пролина/аланина крупного матричного белка Масасине herpesvirus 1 (SEQ ID NO: 18). (F) Гистограмма нуклеотидной последовательности с низкой повторяемостью PAS#1b(200) (SEQ ID NO: 19). (G) Данные по PAS#1b(200), представленные на (F), в более крупном масштабе, (а) Гистограмма нуклеотидной последовательности PA#1e/1d/1c/1b(800) по настоящему изобретению (SEQ ID NO: 44). (G) Данные по PA#1e/1d/1c/1b(800), представленные на (E), в более крупном масштабе. Сводные данные по общей длине нуклеотидной последовательности (N_{tot}) и ее показателю нуклеотидных повторов (NRS), являющемуся критерием оценки качества молекул нуклеиновой кислоты, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина, в отношении частоты и длины повторов, для различных нуклеотидных последовательностей представлены в табл. 1 и 2.

Фиг. 10. рASK75-PA#1d/1c/1b(600)-IL1Ra, генетически стабильный вектор экспрессии для получения терапевтически значимого слитого белка PA#1(600)-IL1Ra в бактериях

Карта плазмиды рASK75-PA#1d/1c/1b(600)-IL1Ra (SEQ ID NO: 77). Структурный ген биологически/фармакологически активного (пре)белка PA#1(600)-IL1Ra, содержащий нуклеотидную последовательность с низкой повторяемостью, кодирующую полипептид PA#1 из 601 аминокислотного остатка, и структурный ген человеческого IL-1Ra, а также кодирующие области бактериальной сигнальной последовательности OmpA и His₆-метки, клонируют, помещая под транскрипционный контроль промотора/оператора tet (tet^{p/o}). Основа плазмиды вне экспрессионной кассеты, фланкированной сайтами рестрикции XbaI и HindIII, идентична основе универсального вектора экспрессии рASK75 (Skerra (1994) Gene 151:131-135). Показан единственный сайт рестрикции SapI, оставшийся после введения молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, по настоящему изобретению.

Фиг. 11. Анализ генетической стабильности молекулы нуклеиновой кислоты с низкой повторяемостью PA#1d/1c/1b(600), кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, в сравнении с нуклеотидной последовательностью по предшествующему уровню техники PA#1a(600)

Электрофорез в агарозном геле после XbaI/HindIII-рестрикционного анализа 10 препаратов плазмид из клонов *E. coli* JM83, трансформированных плазмидой pASK75-PA#1d/1c/1b(600)-IL1Ra (фиг. 10) (SEQ ID NO: 77), или 10 препаратов плазмиды pASK75-PA#1a(600)-IL1Ra (SEQ ID NO: 78) после культивирования на протяжении 7 суток, что соответствует примерно 70 поколениям бактериального клеточного деления. Дорожки: M, стандарт молекулярной массы (GeneRuler 1 kb DNA Ladder: 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000 и 10000 п.о.); 1-10: образцы плазмид от отдельных клонов после расщепления рестриктазой. По меньшей мере 4 из 10 проанализированных клонов pASK75-PA#1a(600)-IL1Ra продемонстрировали укорочение фрагментов ДНК, содержащих введенную нуклеиновую кислоту, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина (фиг. 11A), что явно указывает на генетическую нестабильность. В отличие от этого, все 10 клонов pASK75-PA#1d/1c/1b(600)-IL1Ra (фиг. 11B) продемонстрировали только ожидаемые полосы, соответствующие 3093 п.о. и 2377 п.о. соответственно, что указывает на интактность введенной нуклеиновой кислоты, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, и высокую генетическую стабильность плазмиды. Таким образом, нуклеотидные последовательности с низкой повторяемостью, кодирующие повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению имеют явное преимущество перед повторяющимися нуклеотидными последовательностями по предшествующему уровню техники.

Фиг. 12. Конструирование генетически стабильных векторов экспрессии для получения человеческого лептина, слитого с аминокислотными последовательностями с высоким содержанием пролина/аланина, в бактериях

(А) Карта плазмиды pASK37-MP-huLeptin (SEQ ID NO: 81), содержащей кассету для "бесшовного" клонирования, фланкированную сайтами рестрикции SapI, позволяющую проводить "бесшовное" и прямое клонирование нуклеотидной последовательности с низкой повторяемостью, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, в рамке со структурным геном человеческого лептина. (В) Плазмидная карта pASK37-MP-huLeptin-PA#1d/1c/1b(600) (SEQ ID NO: 82), производного pASK37-MP-huLeptin с введенной генной кассетой PA#1d/1c/1b(600) (SEQ ID NO: 42). (С) Плазмидная карта pASK37-MP-huLeptin-PAS#1f/1c/1b(600) (SEQ ID NO: 83), производного pASK37-MP-huLeptin с введенной генной кассетой PAS#1f/1c/1b(600) (SEQ ID NO: 38). Структурные гены биологически/фармакологически активного человеческого белка лептина, человеческого лептина, слитого с нуклеотидной последовательностью с низкой повторяемостью, кодирующей PA#1(600), и человеческого лептина, слитого с нуклеотидной последовательностью с низкой повторяемостью, кодирующей полипептид PAS#1(600), клонировали, помещая под транскрипционный контроль промотора/оператора lacUV5 (lacUV5^{pro}) и располагая каждый ген после кодонов начального остатка Met и остатка Pro. Основа плазмиды вне экспрессионной кассеты, фланкированной сайтами рестрикции XbaI и HindIII, идентична основе универсального вектора экспрессии pASK37 (Skerra (1991) Protein Eng. 4:971-979).

Фиг. 13. Анализ варианта человеческого лептина, слитого с полипептидом PA#1(600), полученного в *E. coli* с использованием генетически стабильного вектора экспрессии pASK37-MP-PA#1d/1c/1b(600)-huLeptin(W100Q)

(А) SDS-PAGE-анализ слитого белка PA#1(600)-huLeptin(W100Q) с использованием 10%-го полиакриламидного геля и последующим окрашиванием кумасси бриллиантовым голубым R-250. В геле показаны маркер молекулярной массы белка (MW) (дорожка M; Thermo Fisher Scientific, Уолтем, штат Массачусетс, США), экстракт целых клеток *E. coli* после 19 ч экспрессии в восстанавливающих условиях (дорожка 1), белок, осажденный с использованием 1 М (NH₄)₂SO₄, восстановленный (дорожка 2) и невосстановленный (дорожка 5), белок после анионообменной хроматографии, восстановленный (дорожка 3) и невосстановленный (дорожка 6), и белок после эксклюзионной хроматографии, восстановленный (дорожка 4) и невосстановленный (дорожка 7). PA#1(600)-huLeptin(W100Q) виден как единственная полоса, что указывает на однородность состава. (В) Анализ слитого белка PA#1(600)-huLeptin(W100Q) масс-спектрометрией с электрораспылительной ионизацией (Electrospray Ionisation Mass Spectrometry, ESI-MS). Развернутый спектр демонстрирует измеренную массу 64249,5 Да, что соответствует расчетной массе рекомбинантного слитого белка (64249,8 Да), указывая на успешное отщепление начального остатка Met бактериальной метионинаминопептидазой.

Настоящее изобретение дополнительно описано посредством следующих иллюстративных неограничивающих примеров, обеспечивающих лучшее понимание настоящего изобретения и многих его преимуществ.

Пример 1. Синтез нуклеотидных последовательностей с низкой повторяемостью, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина

Несколько разных нуклеотидных последовательностей с высокой содержанием пролина/аланина из 200 остатков, оптимизировали, в том числе вручную, для обеспечения низкой повторяемости на нуклеотидном уровне, низкого содержания GC, малой склонности РНК к образованию вторичных структур, использования ко-

донов, предпочтительных для экспрессии в *E. coli*, и отсутствия антивирусных мотивов, а также cis-действующих элементов. Для этого применяли общеизвестные алгоритмы, такие как метод оптимизации кодонов с учетом конкретных условий (condition-specific codon optimization) (Lanza (2014) BMC Syst Biol 8:33) или алгоритм GeneOptimizer (Raab (2010) Syst Synth Biol 4:215-225). Полученные исходные последовательности корректировали вручную следующим образом.

Повторы, длина которых превышала пороговое значение (например, 14 нуклеотидов), выявляли с использованием программного обеспечения Visual Gene Developer версии 1.2, которое доступно на бесплатной основе на сайте <http://visualgenedeveloper.net>. Затем в выявленных повторах проводили постепенную замену кодонов. В частности, кодоны с высоким содержанием GC в выявленных повторах заменяли кодонами с высоким содержанием AT, преобладающими в генах, активно экспрессируемых у выбранного организма-хозяина (например, *E. coli*, *P. pastoris* или СНО). После каждой замены всю нуклеотидную последовательность еще раз анализировали на предмет повторов. В случае, если замена привела к новому повтору, длина которого превышала заданный порог, нуклеотидную замену (замены) отвергали и заменяли другой кодон в выявленном ранее повторе. Если это не приводило к устранению повтора, то два кодона в выявленном длинном повторе заменяли одновременно. Таким образом, все повторы, выявленные выше, длина которых превышала заданный порог, постепенно устранили, не изменяя при этом кодируемую аминокислотную последовательность с высоким содержанием пролина/аланина.

На второй стадии индекс адаптации кодонов (codon adaptation index, CAI), содержание GC и стабильные мРНК-структуры оптимизированной нуклеотидной последовательности анализировали с использованием программного обеспечения Visual Gene Developer и сравнивали с начальной последовательностью. Проводили дополнительную ручную корректировку, снова проводя замену кодонов/"молчащие" мутации до достижения оптимизированной нуклеотидной последовательностью CAI, содержания GC или мРНК-структур, таких же, как у начальной последовательности, или более благоприятных, чем у начальной последовательности. Снова проводили повторный анализ по стадии 1 и, по необходимости, заменяли другие кодоны для достижения целевых параметров, которыми были пороговая длина повторов, CAI, содержание GC и мРНК-структуры (вторичные структуры).

На третьей стадии разные оптимизированные по отдельности нуклеотидные последовательности, каждая из которых кодировала одну и ту же повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина из 200 остатков, объединяли, то есть присоединяли друг к другу, и полученную нуклеотидную последовательность большей длины оптимизировали таким же образом, как на стадиях 1 и 2. В завершение, полученную длинную последовательность нуклеиновой кислоты разделяли на более короткие, например, на ДНК-кассеты длиной 600 нуклеотидов. Например, последовательность из 2400 нуклеотидов PAS#1d/1f/1c/1b (SEQ ID NO: 39) была разделена на четыре более короткие кассеты (SEQ ID NO: 19, 20, 21, 23). Сходным образом, последовательность из 2400 нуклеотидов PA#1e/1d/1c/1b (SEQ ID NO: 44) была разделена на четыре более короткие кассеты (SEQ ID NO: 28, 29, 30, 31), каждая из которых содержит по 600 нуклеотидов.

Эти оптимизированные элементы нуклеотидной последовательности, фланкированные сайтами распознавания SapI (5'-GCTCTTC-3') в обратно-комплементарной ориентации, приводящими к образованию "липких" нуклеотидных концов 5'-GCC/5'-GGC после расщепления рестриктазой, были синтезированы по отдельности разными коммерческими поставщиками. Следует отметить, что из-за присутствия двух "липких" нуклеотидных концов GCC/GGC только 597 нуклеотидов, которые расположены в середине, образуют двойную спираль ДНК после вырезания и, таким образом, составляют пары оснований (п.о.). Кроме того, оптимизированная последовательность из 600 нуклеотидов удлинена дополнительным кодоном Ala из-за присутствия второго сайта рестрикции SapI, из-за которого клонированная ДНК-кассета, кодирующая аминокислотную последовательность с высоким содержанием пролина/аланина, имеет общую длину 603 нуклеотида. Присутствие двух фланкирующих сайтов рестрикции SapI позволяет проводить точное вырезание и субклонирование всей ДНК-касеты по изобретению, например, в rXL2.

Аналогичным образом разрабатывали и синтезировали другие наборы элементов нуклеотидной последовательности, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, кодоны которых были оптимизированы для экспрессии в *Escherichia coli*, *Pichia pastoris*, клетках эмбриональной почки человека (HEK), *Pseudomonas fluorescens*, *Corynebacterium glutamicum*, *Bacillus subtilis*, *Tetrahymena thermophila*, *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Physcomitrella patens* или *Cricetulus griseus*. Таблицы предпочтительных кодонов для этих организмов можно загрузить с сайта <http://www.kazusa.or.jp/codon>. Сводные данные по синтезированным молекулам нуклеиновой кислоты по изобретению и характеристикам их нуклеотидных последовательностей представлены в табл. 1.

Пример 2. Сборка нуклеотидных последовательностей большей длины, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, из элементов нуклеотидной последовательности с низкой повторяемостью

Плазмиды, полученные от коммерческих поставщиков, каждая из которых несла клонированный фрагмент синтезированной ДНК, обрабатывали SapI, и полученные фрагменты ДНК длиной 600 нуклео-

тидов очищали электрофорезом в агарозном геле, следуя стандартным методикам (Sambrook (2001), см. выше). Из отдельных элементов нуклеотидной последовательности собирали нуклеотидные последовательности большей длины с использованием плазмиды рXL2 (SEQ ID NO: 48), производного рUC19 (Yanisch-Perron (1985). *Gene*. 33, 103-119), показанной на фиг. 1A. рXL2 содержит один сайт рестрикции SapI, и ее расщепление этой рестриктазой Типа IIS приводит к образованию "липкого" конца 5'-GCC/5'-GGC, кодирующего аланин и совместимого с "липкими" концами очищенных синтезированных фрагментов ДНК (фиг. 1C). После введения/лигирования одного элемента нуклеотидной последовательности плазида может быть раскрыта на одном конце, в данном случае верхнем, клонированного элемента последовательности повторной обработкой рестриктазой SapI (фиг. 1D). Структура данного вектора позволяет проводить постепенное введение идентичных или разных элементов нуклеотидной последовательности с низкой повторяемостью с получением клонированных генных кассет большей длины, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина (фиг. 1E).

В качестве примера первый элемент нуклеотидной последовательности PAS#1b(200) (SEQ ID NO: 19), затем элемент последовательности PAS#1c(200) (SEQ ID NO: 20) и после него элемент последовательности PAS#1f(200) (SEQ ID NO: 23) вводили в рXL2 описанным образом с использованием сайта рестрикции SapI, получая плазмиду рXL2-PAS#1f/1c/1b(600) (SEQ ID NO: 38). На следующей стадии таким же образом с использованием сайта рестрикции SapI дополнительно вводили элемент последовательности PAS#1d(200) (SEQ ID NO: 19). Полученная плазида содержала собранную ДНК-кассету длиной 2400 п.о. PAS#1d/1f/1c/1b(800), и максимальная длина повторов, выявленных во всей этой кассете, составила 14 нуклеотидов (SEQ ID NO: 39). Поскольку последовательность распознавания EarI (5'-CTCTTC-3'), расположенная ниже ДНК-кассеты с низкой повторяемостью, клонированной в рXL2, является также частью последовательности распознавания SapI, вся собранная ДНК-кассета может быть легко вырезана при расщеплении рестриктазой EarI, разрезающей, таким образом, плазмиду в двух местах, что позволяет использовать ДНК-кассету для дальнейшего субклонирования.

Таким же образом нуклеотидная последовательность с низкой повторяемостью PA#1e/1d/1c/1b(800) (SEQ ID NO: 44) была собрана из единиц нуклеотидной последовательности PA#1b(200) (SEQ ID NO: 28), PA#1c(200) (SEQ ID NO: 29), PA#1d(200) (SEQ ID NO: 30) и PA#1e(200) (SEQ ID NO: 31) в указанном порядке. Описанная собранная нуклеотидная последовательность, а также другие типичные молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующие повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению, кодоны которых также оптимизированы для организмов-хозяев, отличных от *E. coli*, представлены в табл. 1. Раскрытая методика клонирования позволяет проводить простую постадийную сборку сложных генных кассет, содержащих длинные молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующие повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, которые нельзя получить напрямую обычными методами генного синтеза.

Пример 3. Анализ повторяемости нуклеотидных последовательностей, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина

Проводили анализ точечных диаграмм различных нуклеотидных последовательностей, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, PA#3 (SEQ ID NO: 15) (фиг. 2A), раскрытой WO 2011144756, PAS#1 (SEQ ID NO: 11) (фиг. 2B), раскрытой в WO 2008155134, мультимера [(AP)₅]_n (SEQ ID NO: 16), раскрытого в WO 2004094590, и гена повторяющейся области аминокислотной последовательности с высоким содержанием пролина/аланина сверхкрупного оболочечного белка Macacine herpesvirus 1, опубликованного в GenBank под номером доступа AAP41454.1 (SEQ ID NO: 18) (фиг. 2C). Анализ проводили, выравнивая каждую нуклеотидную последовательность относительно самой себя с использованием инструмента для построения точечных диаграмм "dottup" из пакета программного обеспечения Geneious версии 8.1 (Biomatters, Окленд, Новая Зеландия), применяя окно повтора 14 или 15 нуклеотидов. Алгоритм данного программного обеспечения основан на находящемся в свободном доступе инструменте "dottup" EMBOSS 6.5.7 (институт Сэнгера (Sanger Institute), Кембридж, Великобритания). Полученные двумерные точечные диаграммы нуклеотидных последовательностей по предшествующему уровню техники сравнивали с точечными диаграммами элементов нуклеотидной последовательности с низкой повторяемостью PA#3b(200) (SEQ ID NO: 36), PA#1b(200) (SEQ ID NO: 28) и собранных нуклеотидных последовательностей PAS#1f/1c/1b(600) (SEQ ID NO: 38) и PAS#1d/1f/1c/1b(800) (SEQ ID NO: 39), кодирующих повторяющиеся аминокислотные последовательности большей длины с высоким содержанием пролина/аланина. В то время как все проанализированные нуклеотидные последовательности по предшествующему уровню техники продемонстрировали высокую повторяемость на уровне нуклеотидной последовательности, на что указывают черные диагональные линии (фиг. 2A, B, C), точечные диаграммы оптимизированных нуклеотидных последовательностей, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению продемонстрировали лишь небольшое число редких или коротких повторов длиной 14 нуклеотидов (черные линии) на протяжении всей проанализированной нуклеотидной последовательности длиной 600 нуклеотидов в случае кассет PA#3b(200) и PA#1b(200)

(фиг. 2А, С), 1800 нуклеотидов в случае кассеты PAS#1f/1c/1b(600) (фиг. 2В) или 2400 нуклеотидов в случае кассеты PAS#1d/1f/1c/1b(800) (фиг. 2В).

Пример 4. ДНК-секвенирование молекул нуклеиновой кислоты с низкой повторяемостью, кодирующих длинные повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина

ДНК-кассету с низкой повторяемостью PAS#1f/1c/1b(600) (SEQ ID NO: 38), клонированную в плазмиду pXL2 и описанную в примере 2, секвенировали в центре, предоставляющем услуги по секвенированию ДНК (Eurofins Genomics, Эберсберг, Германия), с применением циклического секвенирования по Сэнгеру на приборе ABI 3730XL (Thermo Fisher Scientific, Уолтем, штат Массачусетс, США). Для этого 8 мкл (150 нг/мкл) плазмидной ДНК pXL2-PAS#1f/1c/1b(600), выделенной из трансформированных клеток *E. coli* XL1-blue с использованием набора QIAprep Spin Miniprep (Qiagen, Хильден, Германия), смешивали с 5 мкл дважды дистиллированной H₂O и 2 мкл праймера XLP-1 (10 мкМ) (SEQ ID NO: 3), гибридизующегося в кодирующей области элемента нуклеотидной последовательности PAS#1b(200), и передавали в центр, предоставляющий услуги по секвенированию ДНК. В результате, была получена безошибочная электрофограмма, включавшая более 900 определенных нуклеотидов (фиг. 3), без признаков неспецифического или множественного связывания праймеров. Таким образом, в отличие от длинных повторяющихся нуклеотидных последовательностей, которые можно секвенировать лишь частично с использованием праймеров, гибридизующихся с нуклеотидными последовательностями вектора выше или ниже клонированной ДНК, молекулы нуклеиновой кислоты с низкой повторяемостью по настоящему изобретению, кодирующие повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, можно легко секвенировать, используя также праймеры, специфично гибридизующиеся внутри клонированной нуклеотидной последовательности. Это позволяет считывать множество перекрывающихся последовательностей с использованием разных подходящих праймеров, охватывая всю длину последовательности даже в случае очень длинных молекул нуклеиновой кислоты по изобретению.

Пример 5. Конструирование pASK75-PAS#1f/1c/1b(600), генетически стабильного вектора экспрессии для получения терапевтического слитого белка PAS#1(600)-IL1Ra в бактериях

Для конструирования экспрессионной плазмиды, кодирующей антагонист рецептора интерлейкина-1 (IL-1Ra), слитый с повторяющейся аминокислотной последовательностью из 600 остатков PAS#1 (SEQ ID NO: 38), вектор pASK75-IL1Ra (фиг. 4А) (SEQ ID NO: 49) разрезали рестриктазой SapI дефосфорилировали щелочной фосфатазой креветки (Thermo Fisher Scientific, Уолтем, штат Массачусетс, США) и лигировали с фрагментом ДНК, соответствующим кассете нуклеотидной последовательности с низкой повторяемостью, кодирующей полипептид из 600 остатков PAS#1, вырезанным из плазмиды pXL2-PAS#1f/1c/1b(600) посредством расщепления рестриктазой EarI. После трансформации *E. coli* XL1-Blue (Bullock (1987) *Biotechniques* 5:376-378) получали плазмидную ДНК и присутствие введенного фрагмента ДНК подтверждали рестрикционным анализом и секвенированием ДНК. Полученная плазида была обозначена pASK75-PAS#1f/1c/1b(600)-IL1Ra (SEQ ID NO: 50) и показана на фиг. 4В

Пример 6. Анализ долгосрочной генетической стабильности плазмиды, содержащей молекулу нуклеиновой кислоты с низкой повторяемостью, кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина

Генетическую стабильность плазмиды pASK75-PAS#1f/1c/1b(600)-IL1Ra (SEQ ID NO: 50) сравнивали с генетической стабильностью pASK75-PAS#1a(600)-IL1Ra (SEQ ID NO: 51), производного, где ДНК-кассета PAS#1f/1c/1b(600) DNA была заменена повторяющейся нуклеиновой кислотой PAS#1a(600) (SEQ ID NO: 12). Для этого *E. coli* KS272 (Strauch (1988) *Proc. Natl. Acad. Sci. USA* 85:1576-1580) трансформировали соответствующей плазмидой, применяя метод с хлоридом кальция (Sambrook (2001), см. выше), и культивировали на протяжении 7 суток при 37°C, 170 об/мин в 50 мл среды Лурия-Бертани (LB), дополненной ампициллином, 100 мг/мл, во встряхиваемой колбе объемом 100 мл без индукции экспрессии гена. На протяжении этого периода бактериальные клетки два раза в сутки (утром и вечером) переносили в свежую среду, применяя разведение 1:1000. На 7 сутки, после непрерывного культивирования на протяжении примерно 70 поколений, культуру доводили до стационарной фазы и клетки высевали на агар LB/Amp. Затем отбирали отдельные клоны, которые использовали в качестве посевного материала для культур в среде LB объемом 50 мл, и, после выращивания до стационарной фазы в течение ночи, получали плазмидную ДНК от пяти клонов каждой из двух плазмид с использованием набора Qiagen Miniprep Kit (Qiagen, Хильден, Германия) и анализировали ее посредством расщепления рестриктазами XbaI/HindIII (фиг. 5).

Только 1 из 5 проанализированных клонов pASK75-PAS#1a(600)-IL1Ra продемонстрировал ожидаемые полосы, соответствующие 3093 п.о. и 2377 п.о. (фиг. 5, дорожка 1). Два клона (фиг. 5, дорожки 3 и 5) продемонстрировали полосу 573 п.о., примерно соответствующую размеру объединенных генных последовательностей OmpA и IL1Ra, указывая на более или менее полную потерю кассеты с повторяющейся последовательностью PAS#1a(600), вероятно, в результате рекомбинации. Два других клона продемонстрировали значительно укороченные фрагменты ДНК (фиг. 5, дорожки 2 и 4), также указывая на делеции в кассете повторяющейся последовательности PAS#1a(600) и, следовательно, генетическую не-

стабильность. В отличие от этого, все пять проанализированных клонов pASK75-PAS#1f/1c/1b(600)-IL1Ra продемонстрировали ожидаемые полосы 3093 п.о. и 2377 п.о. (фиг. 5, дорожки 6-10), что указывает на интактность генной кассеты, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, и, таким образом, высокую генетическую стабильность молекул нуклеиновой кислоты с низкой повторяемостью по настоящему изобретению в плазмиде.

Пример 7. "Бесшовное" и прямое клонирование нуклеотидной последовательности с низкой повторяемостью, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, в экспрессионную плазмиду, кодирующую биологически активный белок IL-1Ra

Для терапевтического применения желательны слитые белки, содержащие только биологически активный белок и повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина. Отсутствие дополнительных аминокислотных линкеров, например, вводимых для обеспечения или использования сайтов рестрикции для клонирования, может предотвратить возможные иммунные ответы при клиническом применении и/или позволит избежать нежелательных взаимодействий на белковом уровне. Поэтому была разработана методика "бесшовного" клонирования (фиг. 6) для прямого введения нуклеотидных последовательностей с низкой повторяемостью, в данном случае на примере фрагмента ДНК, содержащего PA#1b(200) (SEQ ID NO: 28), в плазмиду, являющуюся производным универсальной экспрессионной плазмиды pASK75 (Skerra (1994) loc. cit), кодирующую биологически активный белок IL1-Ra (Molto (2010) Joint Bone Spine. 77:102-107).

Сначала получали синтетический фрагмент ДНК, кодирующий зрелую аминокислотную последовательность IL1-Ra (UniProt ID P18510), от поставщика услуг по генному синтезу (Thermo Fisher Scientific, Регенсбург, Германия). Этот генный фрагмент (SEQ ID NO: 46) содержал сайт рестрикции XbaI с последующим сайтом связывания рибосом, нуклеотидную последовательность, кодирующую сигнальный пептид OmpA с последующим аланиновым кодоном GCC, первую последовательность распознавания SapI GCTCTTC на некодирующей цепи, динуклеотидный спейсер GC и вторую последовательность рестрикции SapI в обратно-комплементарной ориентации с ее последовательностью распознавания GCTCTTC на кодирующей цепи и последующим аланиновым кодоном GCC, напрямую связанным с кодирующей последовательностью зрелого IL-1Ra (UniProt ID P18510), и, в завершение, сайт рестрикции HindIII.

Этот генный фрагмент клонировали в pASK75 с использованием фланкирующих сайтов рестрикции XbaI и HindIII, следуя стандартным методикам (Sambrook (2001), см. выше). Полученную плазмиду (см. фиг. 6A) расщепляли с использованием SapI, что приводило к высвобождению небольшой (24 п.о.) ДНК-вставки, содержащей оба сайта распознавания SapI, и расщепленной основы вектора с совместимыми "липкими" концами 5'-GCC/5'-GGC, расположенными непосредственно напротив N-конца кодируемого зрелого IL-1Ra, что идеально подходит для введения молекулы нуклеиновой кислоты с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина (фиг. 6B). После выделения фрагмента вектора с использованием набора для выделения из геля QIAquick (Qiagen, Хильден, Германия) и дефосфорилирования термочувствительной щелочной фосфатазой FastAP (Thermo Fisher Scientific, Уолтем, штат Массачусетс, США), в обоих случаях следуя инструкциям изготовителей, его лигировали с генной кассетой PA#1b(200), вырезанной из pXL2-PA#1b(200) (SEQ ID NO: 54) посредством расщепления рестриктазой EarI (фиг. 6C). Полученная плазида (SEQ ID NO: 56) позволяет экспрессировать в бактериях слитый белок (SEQ ID NO: 10), состоящий только из повторяющейся аминокислотной последовательности с высоким содержанием пролина/аланина и биологически активного белка IL-1Ra (после *in vivo*-процессинга сигнального пептида OmpA в процессе периплазматической секреции у *E. coli*).

Пример 8. Получение слитого белка последовательности PAS#1(600) и IL-1Ra, кодируемого генетически стабильной плазмидой pASK75-PAS#1f/1c/1b(600)-IL1Ra, в бактериях и его очистка

Слитый белок PAS#1(600)-IL1-Ra (расчетная масса: 68 кДа) получали при 25°C в *E. coli* KS272, содержащих генетически стабильную экспрессионную плазмиду pASK75-PAS#1f/1c/1b(600)-IL1Ra по примеру 6 и плазмиду-помощника pTUM4, способствующую правильному сворачиванию белка (Schlarschu (2006) Protein Eng. Des. Sel. 20:273-284), с использованием стандового ферментера с синтетической глюкозоминеральной средой, дополненной ампициллином, 100 мг/л, и хлорамфениколом, 30 мг/л, в соответствии с опубликованной методикой (Schiweck (1995) Proteins 23:561-565). Экспрессию рекомбинантного гена индуцировали добавлением 500 мкг/л ангидротетрациклина (Skerra (1994), см. выше) по достижении культурой OD₅₅₀ 28. После периода индукции продолжительностью 2,5 ч клетки собирали центрифугированием и ресуспендировали в течение 10 мин в ледяном буфере для выделения периплазматической фракции (500 mM сахарозы, 1 mM EDTA, 100 mM трис/HCl, pH 8,0; 2 мл/л и OD₅₅₀). После добавления 15 mM EDTA и 250 мкг/л лизоцима суспензию клеток инкубировали в течение 20 мин на льду, центрифугировали несколько раз и отбирали осветленный супернатант, содержащий рекомбинантный белок.

Проводили четырехкратный диализ периплазматического экстракта при 4°C против 5 л 40 mM фосфата Na, pH 7,5, 500 mM NaCl, соответственно, и очистку посредством His₆-метки с использованием колонки HisTrap HP на 80 мл (GE Healthcare, Фрайбург, Германия). Белок элюировали с градиентом концентрации имидазола/HCl, pH 7,5, от 0 до 200 mM в 40 mM фосфата Na, pH 7,5, 0,5 M NaCl. Очищенный

белок объединяли и подвергали двукратному диализу против 5 л 20 мМ трис/HCl, pH 8,0, 1 мМ EDTA при 4°C на протяжении по меньшей мере 6 ч, соответственно. Полученный после диализа раствор белка подвергали анионообменной хроматографии с использованием колонки ХК на 60 мл (GE Healthcare, Фрайбург, Германия) со смолой Source15Q, соединенной с системой очистки Akta (GE Healthcare, Фрайбург, Германия), с использованием 20 мМ трис/HCl, pH 8,0, 1 мМ EDTA в качестве рабочего буфера. Белок элюировали с использованием градиента концентрации NaCl от 0 до 200 мМ в рабочем буфере.

Элюированные фракции подвергали двукратному диализу против 10 мМ MES/HCl, pH 6,0, 1 мМ EDTA при 4°C на протяжении по меньшей мере 6 ч, соответственно, и затем катионообменной хроматографии с использованием колонки ХК с 36 мл смолы Source15S (GE Healthcare, Фрайбург, Германия). Катионообменную хроматографию проводили в системе Akta с использованием 10 мМ MES/HCl, pH 6,0, 1 мМ EDTA в качестве рабочего буфера и градиента концентрации NaCl от 0 до 500 мМ в рабочем буфере за 4 объема колонки для элюирования белка. Элюированные белковые фракции, содержащие PAS#1(600)-IL1-Ra, снова объединяли, подвергали диализу против 5 л забуференного фосфатом физиологического раствора (PBS: 115 мМ NaCl, 4 мМ KH_2PO_4 и 16 мМ Na_2HPO_4 , pH 7,4) при 4°C в течение ночи, концентрировали до 5 мг/мл с использованием фильтрующей центрифуги Amicon Ultra (30000 MWCO; 15 мл; Millipore, Биллерика, штат Массачусетс, США) и дополнительно очищали эксклюзионной хроматографией с использованием колонки HiLoad 26/60 Superdex 200 Prep Grade (GE Healthcare, Фрайбург, Германия), уравновешенной с использованием PBS.

Был получен однородный препарат белка без признаков агрегации с конечным выходом 70 мг от одного ферментера объемом 8 л. Концентрацию белка определяли, измеряя поглощение при 280 нм с применением расчетного коэффициента экстинкции (Gill (1989) Anal. Biochem. 182:319-326) $15720 \text{ M}^{-1} \text{ cm}^{-1}$. SDS-PAGE проводили с использованием системы с трис-буфером высокой молярности (Fling (1986) Anal. Biochem. 155:83-88) (фиг. 7A).

Пример 9. ESI-MS-анализ слитого белка PAS#1 (600VIL1Ra PAS#1(600)-IL1Ra, полученный и очищенный как описано в примере 8, подвергали двукратному диализу против 1000-кратного объема 10 мМ ацетата аммония, pH 6,8, и анализировали ESI-масс-спектрометрией на приборе Q-ToF Ultima (Waters, Эшброн, Германия) в режиме положительных ионов. Развернутый спектр слитого белка PA#1(600)-IL1Ra продемонстрировал массу 67994,8 Да, что по существу совпадает с расчетной массой 67994,8 Да (фиг. 7B). Это ясно демонстрирует, что полноразмерный слитый белок PA#1(600)-IL1Ra можно эффективно получать в *E. coli* с использованием генетически стабильной экспрессионной плазмиды pASK75-PAS#1f/1c/1b(600)-IL1Ra.

Пример 10. Конструирование pASK37-MP-PA#1d/1c/1b(600), генетически стабильной плазмиды для получения полипептида с повторяющейся аминокислотной последовательностью с высоким содержанием пролина/аланина *E. coli*

Для конструирования стабильной экспрессионной плазмиды, кодирующей чистый полипептид PA#1(600), 100 пмоль праймеров NdeI-MP-SapI-HindIII_{fw} (SEQ ID NO: 4) и NdeI-MP-SapI-HindIII_{rev} (SEQ ID NO: 5) фосфорилировали, перемешивали и нагревали до 80°C на 10 мин с последующим медленным охлаждением до комнатной температуры в течение ночи для проведения гибридизации. Полученный фрагмент двуцепочечной ДНК имел "липкие" концы, совместимые с "липкими" концами NdeI и HindIII. Плазмиду pASK37 (Skerra (1991), см. выше) разрезали с использованием NdeI и HindIII и основной фрагмент лигировали с гибридизованными праймерами.

Полученную плазмиду расщепляли с использованием SapI, что приводило к высвобождению небольшой (24 п.о.) вставки, содержащей два сайта распознавания SapI, и расщепленной основы вектора с совместимыми "липкими" концами 5'-GCC/5'-GGC. Эти "липкие" концы идеально подходят для введения нуклеотидной последовательности с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, в положение непосредственно ниже кодона начального N-концевого метионина (ATG) с расположенным после него кодоном CCA, обеспечивающим эффективную инициацию трансляции. После выделения фрагмента вектора с использованием набора для выделения из геля QIAquick и дефосфорилирования термочувствительной щелочной фосфатазой FastAP, следуя инструкциям изготовителей, его лигировали с генной кассетой с низкой повторяемостью PA#1d/1c/1b(600) (SEQ ID NO: 42), вырезанной из pXL2-PA#1d/1c/1b(600) посредством расщепления рестриктазой EagI. Полученная плазида (SEQ ID NO: 53) позволяет экспрессировать полипептид содержащий только повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина (фиг. 8A)

Пример 11. Экспрессия полипептида PA#1(600), кодируемого генетически стабильной плазмидой pASK37-MP-PA#1d/1c/1b(600), в бактериях и его очистка

Полипептид PA#1(600) с дополнительным остатком Pro на N-конце и дополнительным остатком Ala на C-конце (расчетная масса: 48302 Да) получали в цитоплазме *E. coli* KS272, содержащих экспрессионную плазмиду pASK37-PA#1d/1c/1b(600), описанную в примере 10. В 4 мл среды LB в стерильной полипропиленовой пробирке объемом 13 мл (Sarstedt, Нюмбрехт, Германия), дополненной глюкозой, 1%, мас./об., и ампициллином, 100 мг/л, засевали колонию *E. coli* KS272, трансформированную pASK37-PA#1d/1c/1b(600), и проводили культивирование в течение ночи при 37°C, 170 об/мин. Получение белка

в бактериях проводили при 30°C во встряхиваемой колбе объемом 5 л с 2 л жидкой среды "terrific broth" (TB) (Sambrook (2001), см. выше), дополненной D-глюкозой, 2,5 г/л и ампициллином, 100 мг/л.

В культурах *E. coli*, полученных с использованием ночной культуры объемом 2 мл в качестве посевого материала, клетки выращивали в течение ночи и экспрессию рекомбинантного гена индуцировали при $OD_{550}=5$ добавлением изопропил- β -D-тиогалактопиранозид (IPTG) до конечной концентрации 0,5 mM. Бактерий собирали через 3 ч после индукции, ресуспендировали в 20 мл 40 mM фосфата Na, pH 7,2, 1 mM EDTA и лизировали с использованием пресса Френча (Thermo Scientific, Уолтем, штат Массачусетс, США). Телес включений после центрифугирования лизата (17000 об/мин, 1 ч, 4°C) не наблюдали. Супернатант, содержащий растворимый полипептид PA#1(600), подвергали осаждению сульфатом аммония, постепенно добавляя твердый $(NH_4)_2SO_4$ до конечной концентрации 20%, мас./об., с постоянным перемешиванием при комнатной температуре. Супернатант центрифугировали на скорости 17000 об/мин при комнатной температуре в течение 20 мин. Осадок, содержащий осажденный полипептид PA#1(600), растворяли в 20 mM трис/HCl, pH 8,0, и полученный раствор центрифугировали (13000 об/мин, 10 мин, комнатная температура) для удаления нерастворимых примесей.

Добавляли чистую уксусную кислоту (Sigma-Aldrich, Штайнхайм, Германия) до конечной концентрации 1% об./об. и примеси осаждали центрифугированием при 13000 об/мин в течение 10 мин. Супернатант, содержащий почти чистый полипептид PA#1(600), подвергали диализу против 100-кратного объема 1% об./об. уксусной кислоты в течение ночи при 4°C. Для удаления оставшихся примесей полученный после диализа белок подвергали субтрактивной катионообменной хроматографии, используя колонку с 1 мл Source15S (GE Healthcare, Фрайбург, Германия), соединенную с системой очистки Akta, и 1% об./об. уксусную кислоту в качестве рабочего буфера.

Образцы после каждой стадии очистки анализировали посредством SDS-PAGE с использованием системы с трис-буфером высокой молярности (Fling (1986), см. выше). После SDS-PAGE гель сначала окрашивали йодидом бария, как описано для анализа PEG (Kurfurst (1992) Anal. Biochem. 200:244-248). Кратко, полиакриламидный гель промывали водой и затем инкубировали в 2,5% мас./об. растворе BaI_2 (дигидрат йодида бария; Sigma-Aldrich, Штайнхайм, Германия) в воде в течение 5 мин. После промывания водой гель переносили в раствор Люголя (10%, мас./об., KI аналитической степени очистки (ArpliChem, Дармштадт, Германия), 5% I_2 аналитической степени очистки (Riedel de Haen AG, Зельце, Германия) в воде) на 5 мин. После обесцвечивания в 10% об./об. уксусной кислоте были видны оранжевые полосы полипептида PA#1(600) (фиг. 8B). Затем гель обесцвечивали водой и подвергали второму окрашиванию кумасси бриллиантовым голубым R250 (Applichem), растворенным в 10% уксусной кислоты (Honeywell Specialty Chemicals, Зельце, Германия), 65% H_2O и 25% изопропанола (CLN, Нидерхуммель, Германия). После обесцвечивания в 10% об./об. уксусной кислоте были видны голубые полосы (белки клеток-хозяев) (фиг. 8C).

Пример 12. ESI-MS-анализ чистого полипептида PA#1(600) 200 мкл выделенного полипептида PA#1(600) по примеру 11 в концентрации 5 мг/мл наносили на колонку Resource RPC объемом 1 мл (GE Healthcare, Фрайбург, Германия), соединенную с системой очистки Akta, с использованием 2% об./об. ацетонитрила, 1% об./об. муравьиной кислоты в качестве рабочего буфера. Белок элюировали с использованием градиента ацетонитрила от 2% об./об. ацетонитрила, 1% об./об. муравьиной кислоты до 80% об./об. ацетонитрила, 0,1% об./об. муравьиной кислоты за 20 объемов колонки. Анализировали непосредственно элюированный белок, проводя ESI-масс-спектрометрию на приборе Q-ToF Ultima в режиме положительных ионов. Развернутый спектр полипептида PA#1(600) продемонстрировал массу 48301,78 Да, что по существу совпадает с расчетной массой полипептида PA#1(600) с дополнительным остатком Pro на N-конце и дополнительным остатком Ala на C-конце, но без начального метионина (48301,4 Да) (фиг. 8D). Это ясно демонстрирует, что чистый полипептид PA#1(600) (без аффинной метки), кодируемый генетически стабильной нуклеотидной последовательностью, может быть получен в *E. coli* в его интактной форме.

Пример 13. Анализ повторов в нуклеотидных последовательностях, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина

В качестве критерия оценки качества молекул нуклеиновой кислоты, кодирующих последовательности с высоким содержанием пролина/аланина, в отношении частоты (числа копий) повторов нуклеотидной последовательности, авторы изобретения разработали показатель нуклеотидных повторов (Nucleotide Repeat Score, NRS), рассчитываемый по следующей формуле:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

В данной формуле N_{tot} представляет собой общую длину анализируемой нуклеотидной последовательности, n представляет собой длину повтора последовательности в анализируемой нуклеотидной последовательности, и частота $f_i(n)$ представляет собой число копий данного повтора последовательности. В случае нескольких разных повторов последовательности одинаковой длины n эти разные повторы последовательности различаются индексом i , а число разных повторов последовательности одинаковой

длины n представляет собой $k(n)$. При наличии только одного типа повтора последовательности длиной n $k(n)$ равно 1. NRS определяют как сумму квадратов длин повторов, умноженную на квадратный корень соответствующей общей частоты, разделенную на общую длину анализируемой нуклеотидной последовательности. Минимальная длина повтора, рассматриваемая при расчете NRS, составляет 4 нуклеотида, что включает все нуклеотидные последовательности, длина которых превышает длину одного кодона/триплета, и варьирует до $N_{\text{tot}}-1$, что является длиной наибольшего повтора нуклеотидной последовательности, который может встречаться в анализируемой нуклеотидной последовательности более одного раза.

В данном контексте термин "повтор" означает, что нуклеотидная последовательность встречается в анализируемой нуклеотидной последовательности по меньшей мере дважды. При расчете частот авторы изобретения учитывали как нуклеотидные участки с идентичной последовательностью, встречающиеся по меньшей мере дважды, так и разные последовательности одинаковой длины, также встречающиеся по меньшей мере дважды. Например, если общая частота 14-мерного повтора составляет пять, это может означать, либо что один и тот же 14-мерный нуклеотидный участок встречается 5 раз, либо что одна 14-мерная нуклеотидная последовательность встречается два раза, а другая последовательность из 14 нуклеотидов встречается в анализируемой нуклеотидной последовательности три раза.

Кроме того, каждый повтор меньшей длины, входящий в состав повтора нуклеотидной последовательности большей длины, при расчете учитывается отдельно. Например, если анализируемая нуклеотидная последовательность содержит два нуклеотидных участка (то есть повтора) GCACC, повторы GCAC и CACC также учитываются при расчете отдельно, независимо от того, входят ли они в состав указанного нуклеотидного участка GCACC или возможно какие-либо дополнительные участки анализируемой нуклеотидной последовательности. Следует отметить, что учитываются только повторы на кодирующей цепи молекулы нуклеиновой кислоты.

Специалист в данной области может определить повторы нуклеотидной последовательности вручную или с помощью универсального программного обеспечения, такого как Visual Gene Developer (Jung (2011), см. выше), которое можно загрузить с сайта <http://www.visualgenedeveloper.net>, или инструмент Repfind (Betley (2002), см. выше), доступный на сайте <http://zlab.bu.edu/repfind>. Тем не менее, не каждый алгоритм позволяет выявить все типы повторов, например, результаты Visual Gene Developer не включают перекрывающиеся повторы. Поэтому результаты, полученные при использовании программных инструментов, нужно проверять и, по необходимости, корректировать вручную. Альтернативно, для однозначного определения повторов нуклеотидной последовательности и автоматического расчета NRS можно применять алгоритм, названный калькулятором NRS (NRS-Calculator), описанный в примере 14.

В данной области известны природные, а также определенные синтетические нуклеиновые кислоты, кодирующие аминокислотные последовательности с высоким содержанием пролина/аланина. Тем не менее, все эти последовательности имеют высокую повторяемость на генетическом уровне, что становится очевидным при анализе NRS, описанном ниже, и поэтому их биотехнологическое и/или биофармацевтическое применение ограничено.

С применением калькулятора NRS, описанного в примере 14, было проведено сравнение нескольких нуклеотидных последовательностей по предшествующему уровню техники, кодирующих аминокислотные последовательности с высоким содержанием пролина/аланина, с молекулами нуклеиновой кислоты с низкой повторяемостью, кодирующими повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению: нуклеотидной последовательности PAS#1a(200) (SEQ ID NO: 11), раскрытой в WO 2008/155134 (фиг. 9A), нуклеотидной последовательности PA#1a(200) (SEQ ID NO: 14), раскрытой в WO 2011144756 (фиг. 9B), нуклеотидной последовательности, кодирующей гликомодуль [(AP)₅]₂₀APA (SEQ ID NO: 16), раскрытой в US 20060252120 (фиг. 9C), нуклеотидной последовательности синтетической генной конструкции, кодирующей гликомодуль [AAPAPAPAP]₁₀AS (SEQ ID NO: 17), опубликованной в GenBank под номером доступа DQ399411.1 (фиг. 9D), последовательности из 225 нуклеотидов, кодирующей последовательность с высоким содержанием пролина/аланина крупного оболочечного белка Macacine herpesvirus 1 (SEQ ID NO: 18), опубликованной в GenBank под номером доступа NP_851896 (фиг. 9E), нуклеотидной последовательности с низкой повторяемостью PAS#1b(200) (SEQ ID NO: 19) по настоящему изобретению (фиг. 9F,G) и нуклеотидной последовательности с низкой повторяемостью PA#1e/1d/1c/1b(800) (SEQ ID NO: 44) по настоящему изобретению (фиг. 9H,I).

Рассчитанные частоты повторов наносили на гистограммы против соответствующих длин повторов с использованием программного обеспечения Kaleidagraph V3.6 (Synergy Software, Реддинг, штат Пенсильвания, США) (фиг. 9). Все гистограммы нуклеотидных последовательностей по предшествующему уровню техники продемонстрировали высокую повторяемость, на что указывает высота гистограмм с широким распределением по длинам повторов вплоть до очень длинных повторов. Следует отметить, что в этих случаях при увеличении длины повторов их частота снижается лишь незначительно (фиг. 9A-E). В отличие от этого гистограммы нуклеотидных последовательностей с низкой повторяемостью PAS#1b(200) и PA#1e/1d/1c/1b(800) по настоящему изобретению демонстрируют лишь небольшое число повторов с максимальной длиной 14 нуклеотидов, частота которых быстро снижается до нуля при пере-

ходе от более коротких повторов к более длинным (фиг. 9F,G,H,I)

Различия повторяемости нуклеотидных последовательностей по предшествующему уровню техники и нуклеотидных последовательностей с низкой повторяемостью по изобретению становятся еще более очевидными при сравнении их показателей нуклеотидных повторов. В то время как у всех последовательностей по предшествующему уровню техники NRS превышает 80000 (табл. 2), последовательность длиной 600 нуклеотидов PAS#1b(200) и последовательность длиной 2400 нуклеотидов PA#1e/1d/1c/1b(800) демонстрируют значения NRS, составляющие лишь 13 и 14, соответственно (табл. 1). Это ясно демонстрирует, что, применительно к повторяемости, качество нуклеотидных последовательностей с низкой повторяемостью, кодирующих повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, по настоящему изобретению, намного выше, чем у последовательностей по предшествующему уровню техники, с меньшим числом и длиной повторов нуклеотидной последовательности.

Таблица 1. Характеристики молекул нуклеиновой кислоты по настоящему изобретению

	№ нуклеотидной последовательности с низкой повторяемостью	SEQ ID NO:	Кодон-оптимизирована для:	Кодируемый аминокислотный повтор	n _{max}	N _{tot}	NRS
A: Элементы (структурные блоки) нуклеотидной последовательности							
1	PAS#1b(200)	19	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	13
2	PAS#1c(200)	20	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	12
3	PAS#1d(200)	21	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	11
4	PAS#1e(200)	22	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	12
5	PAS#1f(200)	23	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	11
6	PAS#1g(200)	24	<i>Pichia pastoris</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	24
7	PAS#1h(200)	25	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	20
8	PAS#1i(200)	26	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP	14	600	17

				APSAPA (SEQ ID NO:1)			
9	PAS#1j(200)	27	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	16
10	PA#1b(200)	28	<i>E. coli</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	21
11	PA#1c(200)	29	<i>E. coli</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	18
13	PA#1d(200)	30	<i>E. coli</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	19
14	PA#1e(200)	31	<i>E. coli</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	22
15	PA#1f(200)	32	<i>CHO (C. griseus)</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	24
16	PA#1g(200)	33	<i>CHO (C. griseus)</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	14	600	24
17	PA#1h(200)	34	<i>CHO (C. griseus)</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	17	600	32
18	PA#1i(200)	35	<i>CHO (C. griseus)</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	17	600	17
19	PA#3b(200)	36	<i>E. coli</i>	AAAPAAAPAA AAPAAP (SEQ ID NO:57)	14	600	26
20	PA#5b(198)	37	<i>E. coli</i>	AAAAAPAAAA AAAAAP (SEQ ID NO:58)	14	594	27
101	PA#1j(200)	87	<i>P. pastoris</i>	AAPAAPAPAAP APAAPA (SEQ ID NO:2)	17	600	39
102	PA#1k(200)	88	<i>P. pastoris</i>	AAPAAPAPAAP APAAPA	17	600	29

				(SEQ ID NO:2)			
103	PA#1l(200)	89	<i>P. pastoris</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	31
104	PA#1m(200)	90	<i>P. pastoris</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	24
105	PA#1n(200)	91	<i>S. cerevisiae</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	38
106	PA#1o(200)	92	<i>S. cerevisiae</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	20
107	PA#1p(200)	93	<i>S. cerevisiae</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	19
108	PA#1q(200)	94	<i>K. lactis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	28
109	PA#1r(200)	95	<i>K. lactis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	23
110	PA#1s(200)	96	<i>K. lactis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	34
111	PA#1t(200)	97	<i>H. sapiens</i> (клетки HEK)	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	25
112	PA#1u(200)	98	<i>H. sapiens</i> (клетки HEK)	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	29
114	PA#1v(200)	99	<i>H. sapiens</i> (клетки HEK)	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	31
114	PA#1w(200)	100	<i>Bacillus subtilis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	23
115	PA#1x(200)	101	<i>Bacillus subtilis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	16	600	27

116	PA#1y(200)	102	<i>Bacillus subtilis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	32
117	PA#1z(200)	103	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	18	600	45
118	PA#1aa(200)	104	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	18
119	PA#1ab(200)	105	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	25
120	PA#1ac(200)	106	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	18
121	PA#1ad(200)	107	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	24
122	PA#1ae(100)	108	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	300	27
123	PA#1af(200)	109	<i>C. glutamicum</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	20
124	PA#1ag(200)	110	<i>C. glutamicum</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	24
125	PA#1ah(200)	111	<i>C. glutamicum</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	25
126	PA#1ai(200)	112	<i>C. glutamicum</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	16	600	21
127	PA#1aj(200)	113	<i>P. patens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	30
128	PA#1ak(200)	114	<i>P. patens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	31
129	PA#1al(200)	115	<i>P. patens</i>	AAPAAPAPAAPAAP	15	600	24

				APAAPA (SEQ ID NO:2)			
130	PA#1am(200)	116	<i>P. fluorescens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	32
131	PA#1an(200)	117	<i>P. fluorescens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	35
132	PA#1ao(200)	118	<i>P. fluorescens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	18	600	41
133	PA#1ap(200)	119	<i>T. thermophila</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	37
134	PA#1aq(200)	120	<i>T. thermophila</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	34
135	PA#1ar(200)	121	<i>T. thermophila</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	600	22
136	PA#1as(200)	122	<i>T. thermophila</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	600	35
137	PAS#1k(200)	123	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14
138	PAS#1l(200)	124	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	15	600	17
139	PAS#1m(200)	125	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	16
140	PAS#1n(100)	126	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	300	15
141	PAS#1o(200)	127	<i>P. pastoris</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	17
142	PAS#1p(200)	128	<i>P. pastoris</i>	ASPAAPAPASPAAP APSAPA	17	600	29

				(SEQ ID NO:1)			
143	PAS#1q(200)	129	<i>P. fluorescens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	600	25
144	PAS#1r(200)	130	<i>P. fluorescens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14
145	PAS#1s(200)	131	<i>P. fluorescens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	600	24
146	PAS#1t(200)	132	<i>C. glutamicum</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	15
147	PAS#1u(200)	133	<i>C. glutamicum</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	12
148	PAS#1v(200)	134	<i>C. glutamicum</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	11
149	PAS#1w(200)	135	<i>P. patens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	15
150	PAS#1x(200)	136	<i>P. patens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	12
151	PAS#1y(200)	137	<i>P. patens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	11	600	10
152	PAS#1z(200)	138	<i>K. lactis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	15
153	PAS#1aa(200)	139	<i>K. lactis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	15	600	17
154	PAS#1ab(200)	140	<i>K. lactis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	16
155	PAS#1ac(200)	141	<i>S. cerevisiae</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14

156	PAS#1ad(200)	142	<i>S. cerevisiae</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14
157	PAS#1ac(200)	143	<i>S. cerevisiae</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14
158	PAS#1af(200)	144	<i>T. thermophila</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	600	25
159	PAS#1ag(200)	145	<i>T. thermophila</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	600	25
160	PAS#1ah(200)	146	<i>T. thermophila</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	15	600	20
161	PAS#1ai(200)	147	<i>H. sapiens</i> (клетки HEK)	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	13
162	PAS#1aj(200)	148	<i>H. sapiens</i> (клетки HEK)	ASPAAPAPASPAAP APSAPA(SEQ ID NO: 1)	12	600	10
163	PAS#1ak(200)	149	<i>H. sapiens</i> (клетки HEK)	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	11
164	PAS#1al(200)	150	<i>B. subtilis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	12	600	11
165	PAS#1am(200)	151	<i>B. subtilis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	13
166	PAS#1an(200)	152	<i>B. subtilis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	600	14
167	PA#1at(200)	192	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	31	600	190
168	PA#1au(200)	193	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	26	600	105
169	PAS#1ao(200)	194	<i>E. coli</i>	ASPAAPAPASPAAP	32	600	211

				APSAPA (SEQ ID NO:1)			
170	PAS#1ap(200)	195	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	26	600	105
В: Собранные нуклеотидные последовательности с низкой повторяемостью							
21	PAS#1f/1c/1b(600)	38	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
22	PAS#1d/1f/1c/1b(800)	39	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	2400	8
23	PAS#1h/1c/1i(600)	40	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	14
24	PAS#1j/1h/1c/1i(800)	41	<i>CHO (C. griseus)</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	2400	13
25	PA#1d/1c/1b(600)	42	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	1800	15
26	PA#1i/1h/1g/1f(800)	43	<i>CHO (C. griseus)</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	2400	22
27	PA#1c/1d/1c/1b(800)	44	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	2400	14
28	PA#1i/1h/1g/1f/1c/1d/1c /1b(1600)	45	<i>E. coli / CHO (C. griseus)</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	27	4800	24
171	PA#1ae/1c(300)	153	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	900	18
172	PA#1ae/1d(300)	154	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	900	17
173	PA#1d/1c(400)	155	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	14	1200	17
174	PA#1b/1c/1d(600)	156	<i>E. coli</i>	AAPAAPAPAAPAAP	14	1800	15

				APAAPA (SEQ ID NO:2)			
175	PA#1d/1b/1c(600)	157	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	1800	17
176	PA#1c/1b/1d(600)	158	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	1800	16
177	PA#1c/1d/1b(600)	159	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	1800	17
178	PA#1b/1d/1c(600)	160	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	1800	16
179	PA#1aa/1c/1d/1c/1b(1000)	161	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	3000	17
180	PA#1ab/1aa/1c/1d/1c/1b(1200)	162	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	3600	17
181	PA#1ac/1ab/1aa/1c/1d/1c/1b(1400)	163	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	4200	16
182	PA#1ad/1ac/1ab/1aa/1c/1d/1c/1b(1600)	164	<i>E. coli</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	4800	16
183	PA#1ao/1an/1am(600)	165	<i>P. fluorescens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	19	1800	27
184	PA#1ai/1ah/1ag/1af(800)	166	<i>C. glutamicum</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	2400	17
185	PA#1y/1x/1w(600)	167	<i>B. subtilis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	1800	24
186	PA#1j/1k/1l/1m(800)	168	<i>P. pastoris</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	2400	23
187	PA#1p/1o/1n(600)	169	<i>S. cerevisiae</i>	AAPAAPAPAAPAAP APAAPA	18	1800	21

				(SEQ ID NO:2)			
188	PA#1s/1r/1q(600)	170	<i>K. lactis</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	17	1800	23
189	PA#1as/1ar/1aq/1ap (800)	171	<i>T. thermophila</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	20	2400	30
190	PA#1v/1u/1t(600)	172	<i>H. sapiens</i> (клетки HEK)	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	19	1800	28
191	PA#1al/1ak/1j(600)	173	<i>P. patens</i>	AAPAAPAPAAPAAP APAAPA (SEQ ID NO:2)	18	1800	24
192	PAS#1n/1b(300)	174	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	900	12
193	PAS#1n/1c(300)	175	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	900	13
194	PAS#1b/1f/1c(600)	176	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
195	PAS#1b/1c/1f(600)	177	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
196	PAS#1c/1b/1f(600)	178	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
197	PAS#1f/1b/1c(600)	179	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
198	PAS#1c/1f/1b(600)	180	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	9
199	PAS#1k/1d/1f/1c/1b (1000)	181	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	20	3000	11
200	PAS#1l/1k/1d/1f/1c/1b (1200)	182	<i>E. coli</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	20	3600	12

201	PAS#1s/1q/1r(600)	183	<i>P. fluorescens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	20	1800	21
202	PAS#1v/1u/1u(600)	184	<i>C. glutamicum</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	1800	13
203	PAS#1an/am/1l(600)	185	<i>B. subtilis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	11
204	PAS#1p/1o/1g(600)	186	<i>P. pastoris</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	1800	20
205	PAS#1ac/1ad/1ac(600)	187	<i>S. cerevisiae</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	15	1800	12
206	PAS#1ab/1aa/1z(600)	188	<i>K. lactis</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	1800	15
207	PAS#1ah/1ag/1af(600)	189	<i>T. thermophila</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	1800	19
208	PAS#1ak/aj/ah(600)	190	<i>H. sapiens</i> (клетки HEK)	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	14	1800	10
209	PAS#1y/1x/1w(600)	191	<i>P. patens</i>	ASPAAPAPASPAAP APSAPA (SEQ ID NO:1)	17	1800	14

Таблица 2. Характеристики нуклеотидных последовательностей по предшествующему уровню техники

	Название последовательности	Организм	SEQ ID NO:	Номер в GenBank / № патента	n_{max}	N_{tot}	NRS
1	PA#1a(200)	синтетическая	11	WO 2008155134	540	600	1127680
2	PA#1a(200)	синтетическая	14	WO 2011144756	540	600	1127680
3	PA#3a(200)	синтетическая	15	WO 2011144756	540	600	1127680
4	[(AP) ₅] ₂₀ APA	синтетическая	16	US 2006/0252120	579	609	1315159
5	Модуль [AAPAPAPAP] ₁₀ AS из pBI-SS-	синтетическая	17	DQ399411.1	243	276	150 961
	(Tom)(AP) ₅ 1-EGFP						
6	Крупный оболочечный белок	<i>Macacine herpesvirus 1</i>	18	NP_851896.1	197	225	81858

Пример 14. Калькулятор NRS (NRS-Calculator), алгоритм для однозначного определения повторов нуклеотидной последовательности и расчета показателя нуклеотидных повторов

Общедоступное программное обеспечение, такое как Visual Gene Developer (Jung (2011), см. выше) или инструмент Refind (Betley (2002), см. выше) не всегда надежны и могут требовать ручной корректировки для правильного расчета всех повторов последовательности в анализируемой нуклеотидной последовательности. Кроме того, возникает необходимость ручного подсчета повторов и отдельного расчета NRS по формуле, описанной в примере 13. Чтобы обеспечить алгоритм, позволяющий получать однозначные результаты, и облегчить расчет NRS, здесь описан простой Python-сценарий, названный калькулятором NRS (NRS-Calculator). Сценарий, выполняемый в среде Python 2.7.10 (<http://www.python.org>), основан на сравнении последовательностей с использованием точечной матрицы и позволяет определить все прямые повторы в потенциально длинной нуклеотидной последовательности, включая перекрывающиеся повторы, без учета разрывов. Сравнение последовательностей с использованием точечной матрицы является методом, который хорошо известен специалисту в данной области и описан в общепринятых руководствах по биоинформатике, таких как, например, Mount (2004) *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, 2nd edition, New York.

Калькулятор NRS проводит расчет частот повторов каждой длины и автоматически рассчитывает NRS по формуле, описанной в примере 13. Для выполнения сценария калькулятора NRS среду выполнения Python версии 2.7.10 загружали с сайта <https://www.python.org/downloads> и устанавливали на ноутбук ThinkPad L530 (Lenovo, Штутгарт, Германия) с установленной операционной системой Windows 7. Сценарий калькулятора NRS, приведенный ниже, сохраняли как открытый текст в файле, названном NRScalculator.py, с использованием Microsoft Windows Editor Version 6.1. Анализируемую нуклеотидную последовательность сохраняли как FASTA-файл, названный sequence.fas, в той же папке. Затем запускали ко-

мандную строку и выбирали директорию, содержащую файлы NRScalculator.py и sequence.fas. Для начала расчета выполняли следующую команду: c:\user\admin\NRSfolder> c:\Python27\python.exe NRScalculator.py sequence.fas.

Эта команда приводила к выводу на экран двух столбцов: в левом столбце были указаны длины повторов (Длина (Length)), а в правом (втором) столбце были указаны частоты соответствующих повторов (Частота (Frequency)). Кроме того, в начале и в конце выводимой информации были указаны N_{tot} и NRS (округленный до целого числа), соответственно.

Сценарий калькулятора NRS

```
import math
import sys
class NRScalculator:
    def __init__(self):
        self.repeats = dict()
        self.sums = dict()
        self.seq = None
        self.range_min = None
        self.range_max = None
    def _match_at(self, row, column):
        return self.seq[row] == self.seq[column]
    def _get_repeats_at(self, row, column):
        length = 1
        search_row = row
        search_column = column
        while True:
            if not 0 <= search_row < len(self.seq):
                break
            if not 0 <= search_column < search_row:
                break
            if length > self.range_max:
                break
            if not self._match_at(search_row, search_column):
                break
            if length >= self.range_min:
                repeats =
self.repeats.setdefault(self.seq[row:row + length], set())
                repeats.add(row)
                repeats.add(column)
                search_row += 1
                search_column += 1
                length += 1
    def _get_repeats(self):
        self.repeats = dict()
        for row in xrange(len(self.seq)):
            for column in xrange(row):
                self._get_repeats_at(row, column)
    def _get_sums(self):
        self.sums = dict()
        for (seq, repeats) in self.repeats.iteritems():
            length = len(seq)
```

```

        self.sums[length] = self.sums.get(length, 0) +
len(repeats)
    def set_range(self, range_min, range_max):
        self.range_min = range_min
        self.range_max = range_max
    def set_sequence(self, seq):
        self.seq = seq
    def work(self):
        if not self.seq and not self.range_min and not
self.range_max:
            raise RuntimeError('Can not work without
initialization')
        self._get_repeats()
        self._get_sums()
    def print_repeats(self):
        print('Sequence (Length bp) : NumRepeats (Positions)')
        for seq, repeats in sorted(self.repeats.iteritems(),
key=lambda t: len(t[0])):
            list = [seq, len(seq), len(repeats)]
            list.extend(map(lambda value: value + 1,
sorted(repeats)))
            print('%s Ntot = %u : %u (%s)' % (seq, len(seq),
len(repeats), ','.join(map(lambda value: str(value + 1),
sorted(repeats)))))
    def print_sums(self):
        print('Length\tFrequency')
        for item in self.sums.iteritems():
            print('%u\t%u' % item)
    def print_score(self):
        sum = 0
        for length, count in self.sums.iteritems():
            sum += (length ** 2) * math.sqrt(count)
        print('NRS = %.0f' % (sum / len(self.seq)))
def handle_sequence(finder, name, sequence):
    finder.set_range(4, len(sequence))
    finder.set_sequence(sequence)
    finder.work()
    print('%s: Ntot = %u' % (name, len(sequence)))
    #finder.print_repeats()
    finder.print_sums()
    finder.print_score()
if len(sys.argv) != 2:
    print('Usage: %s FILENAME' % sys.argv[0])
    sys.exit(1)
finder = NRSCalculator()
with open(sys.argv[1], 'r') as infile:
    name = 'Unnamed'
    seq = ''
    for line in infile:
        line = line.strip()
        if line.startswith('>'):
            if len(seq) > 0:
                handle_sequence(finder, name, seq)
                name = line
                seq = ''
                continue
            seq += line.upper()
    handle_sequence(finder, name, seq)

```

Пример информации, выводимой калькулятором NRS

```
>PAS#1b(200): Ntot = 600
```

```
Length Frequency
```

```
4 587
```

```
5 547
```

```
6 478
```

```
7 388
```

```
8 281
```

```
9 158
```

```
10 90
```

```
11 45
```

```
12 6
```

```
13 4
```

```
14 2
```

```
NRS = 13
```

Пример 15. Конструирование рASK75-PA#1d/1c/1b(600)-IL1Ra, генетически стабильного вектора экспрессии для получения терапевтического слитого белка PA#1(600)-IL1Ra в бактериях

Для конструирования экспрессионной плазмиды, кодирующей антагонист рецептора интерлейкина-1 (IL-1Ra), слитый с повторяющейся аминокислотной последовательностью из 600 остатков PA#1, вектор рASK75-IL1Ra (фиг. 4A) (SEQ ID NO: 49) разрезали рестриктазой SapI, дефосфорилировали щелочной фосфатазой креветки (Thermo Fisher Scientific, Уолтем, штат Массачусетс, США) и лигировали с фрагментом ДНК, соответствующим кассете нуклеотидной последовательности с низкой повторяемостью, кодирующей полипептид из 600 остатков PA#1, вырезанным из плазмиды рXL1-PA#1d/1c/1b(600) (SEQ ID NO: 79) посредством расщепления рестриктазой SapI. После трансформации E. coli XL1-Blue

(Bullock (1987), см. выше) получали плазмидную ДНК и присутствие введенного фрагмента ДНК подтверждали рестрикционным анализом и секвенированием ДНК. Полученная плазида была обозначена рASK75-PA#1d/1c/1b(600)-IL1Ra (SEQ ID NO: 77) и показана на фиг. 10.

Пример 16. Анализ долгосрочной генетической стабильности плазмиды рASK75-PA#1b/1c/1b(600)-IL1Ra, содержащей молекулу нуклеиновой кислоты с низкой повторяемостью PA#1d/1c/1b(600), кодирующую повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина

Генетическую стабильность плазмиды рASK75-PA#1d/1c/1b(600)-IL1Ra (SEQ ID NO: 77) сравнивали с генетической стабильностью рASK75-PA#1a(600)-IL1Ra (SEQ ID NO: 78), производного, где ДНК-кассета PA#1d/1c/1b(600) DNA была заменена повторяющейся нуклеиновой кислотой PA#1a(600) (SEQ ID NO: 80). Для этого *E. coli* JM83 (Yanisch-Perron C. (1985), см. выше) трансформировали соответствующей плазмидой, применяя метод с хлоридом кальция (Sambrook (2001), см. выше), и культивировали на протяжении 7 суток при 37°C, 170 об/мин в 50 мл среды Лурия-Бертани (LB), дополненной ампициллином, 100 мг/л, во встряхиваемой колбе объемом 100 мл без индукции экспрессии гена. На протяжении этого периода бактериальные клетки два раза в сутки (утром и вечером) переносили в свежую среду, применяя разведение 1:1000. На 7 сутки, после непрерывного культивирования на протяжении примерно 70 поколений, культуру доводили до стационарной фазы и клетки высевали на агар LB/Amp. Затем для каждой из двух плазмид отбирали по десять отдельных колоний, каждую из которых использовали в качестве посевного материала для культур в среде LB объемом 50 мл, и после выращивания до стационарной фазы в течение ночи получали плазмидную ДНК с использованием набора Qiagen Miniprep Kit (Qiagen, Хильден, Германия) и анализировали ее посредством расщепления рестриктазами XbaI/HindIII (фиг. 11)

Только 6 из 10 проанализированных клонов рASK75-PA#1a(600)-IL1Ra продемонстрировали ожидаемые полосы, соответствующие 3093 п.о. и 2377 п.о. (фиг. 11А, дорожки 1, 3, 4, 5, 7 и 8). Четыре клона продемонстрировали значительно укороченные фрагменты ДНК (фиг. 11А, дорожки 2, 6, 9 и 10), указывая на делеции в cassette повторяющейся последовательности PA#1a(600) и, следовательно, генетическую нестабильность. В отличие от этого все десять проанализированных клонов рASK75-PA#1d/1c/1b(600)-IL1Ra продемонстрировали ожидаемые полосы 3093 п.о. и 2377 п.о. (фиг. 11В, дорожки 1-10), что указывает на интактность генной cassette, кодирующей повторяющиеся аминокислотные последовательности с высоким содержанием пролина/аланина, и, таким образом, высокую генетическую стабильность молекул нуклеиновой кислоты с низкой повторяемостью по настоящему изобретению в плазмиде.

Пример 17. Конструирование генетически стабильных векторов экспрессии для получения человеческого лептина, слитого с повторяющимися аминокислотными последовательностями с высоким содержанием пролина/аланина, в бактериях

Для конструирования экспрессионной плазмиды, кодирующей человеческий лептин (huLeptin), слитый на N-конце с повторяющейся аминокислотной последовательностью из 600 остатков PA#1 (SEQ ID NO: 82), вектор рASK37-MP-huLeptin (фиг. 12А) (SEQ ID NO: 81) разрезали с использованием SaI, что приводило к вырезанию небольшой (24 п.о.) ДНК-вставки, содержащей оба сайта распознавания SapI, и расщепленной основы вектора с совместимыми "липкими" концами 5'-GCC/5'-GGC, расположенными непосредственно выше N-конца кодируемого зрелого человеческого лептина. Эти "липкие" концы идеально подходят для введения нуклеотидной последовательности с низкой повторяемостью, кодирующей повторяющуюся аминокислотную последовательность с высоким содержанием пролина/аланина, в положение непосредственно ниже кодона начального N-концевого метионина (ATG) с расположенным после него кодоном CCA, обеспечивающим эффективную инициацию трансляции. После выделения фрагмента вектора с использованием набора для выделения из геля QIAquick (Qiagen, Хильден, Германия) и дефосфорилирования термочувствительной щелочной фосфатазой FastAP (Thermo Fisher Scientific, Уолтем, штат Массачусетс, США), в обоих случаях следуя инструкциям изготовителей, плазмиду лигировали с фрагментом ДНК, соответствующим cassette нуклеотидной последовательности с низкой повторяемостью, кодирующей полипептид из 600 остатков PA#1, вырезанным из плазмиды рXL1-PA#1d/1c/1b(600) (SEQ ID NO: 79) посредством расщепления рестриктазой SapI. После трансформации *E. coli* XL1-Blue (Bullock (1987), см. выше) получали плазмидную ДНК и присутствие введенного фрагмента ДНК подтверждали рестрикционным анализом и секвенированием ДНК. Полученная плазида была обозначена рASK37-MP-PA#1d/1c/1b(600)-huLeptin (SEQ ID NO: 82) и показана на фиг. 12В. Аналогично рASK37-MP-PAS#1f/1c/1b(600)-huLeptin (SEQ ID NO: 83), экспрессионную плазмиду, кодирующую человеческий лептин (huLeptin), слитый на N-конце с повторяющейся аминокислотной последовательностью из 600 остатков PAS#1, показанную на фиг. 12С, конструировали введением нуклеотидной последовательности с низкой повторяемостью PAS#1f/1c/1b (SEQ ID NO: 38), вырезанной из рXL1-PAS#1f/1c/1b (SEQ ID NO: 84), в плазмиду рASK37-MP-huLeptin (фиг. 12А) (SEQ ID NO: 81). Сходную методику клонирования можно применять для конструирования вариантов лептина, слитых с повторяющимися аминокислотными последовательностями с высоким содержанием пролина/аланина на C-конце.

Пример 18. Получение слитого белка повторяющейся аминокислотной последовательности с высоким содержанием пролина/аланина и мутантного человеческого лептина, кодируемого генетически стабильной плазмидой pASK37-PA#1d/1c/1b(600)hu-Leptin(W100Q), в бактериях, его очистка и анализ PA#1(600)-huLeptin(W100Q), слитый белок мутантного человеческого лептина с заменой триптофана на глутамин в положении 100 зрелой аминокислотной последовательности (номер доступа в UniProtKB: P41159) и повторяющейся аминокислотной последовательности с высоким содержанием пролина/аланина PA#1(600) (SEQ ID NO: 85) (расчетная масса: 64,25 кДа), получали при 30°C в цитоплазме Origami B (Novagene/Merck Millipore, Биллерика, штат Массачусетс, США), штамма *E. coli*, цитоплазма которого обладает окислительными свойствами из-за мутаций *trxB*, *gor* и *ahpC* (Bessette (1999) Proc. Natl. Acad. Sci. USA 96:13703-13708). Для этого в 4 мл среды LB в стерильной полипропиленовой пробирке объемом 13 мл (Sarstedt, Нюмбрехт, Германия), дополненной D-глюкозой, 1% мас./об. и ампициллином, 100 мг/л, засеивали колонию *E. coli* Origami B, трансформированную генетически стабильной экспрессионной плазмидой pASK37-MP-PA#1d/1c/1b(600)-huLep(W100Q) (SEQ ID NO: 86). Бактериальные клетки выращивали в течение ночи при 30°C в шейкере при 170 об/мин.

Получение белка в бактериях проводили при 30°C в колбе с дефлекторами объемом 5 л с 2 л жидкой среды "terrific broth" (TB) (Sambrook (2001), см. выше), дополненной D-глюкозой, 2,5 г/л, и ампициллином, 100 мг/л, в которую вносили 2 мл ночной культуры *E. coli*. Бактериальные клетки выращивали при 30°C и экспрессию рекомбинантного гена индуцировали при $OD_{550}=0,85$ добавлением изопропил- β -D-тиогалактопиранозид (IPTG) до конечной концентрации 0,5 мМ. Бактерии собирали через 19 ч после индукции, ресуспендировали в 3 мл PBS/E (PBS, дополненный 10 мМ EDTA) на 1 г влажной массы бактериальных клеток и лизировали с использованием клеточного гомогенизатора Panda (GEA, Парма, Италия). Телес включения после центрифугирования лизата (20000 об/мин, 30 мин, 4°C) не наблюдали. В супернатант добавляли 1 мМ 2,2'-дитиопиридина для стимуляции образования дисульфидных связей в рекомбинантном лептине. Супернатант, содержащий растворимый слитый белок лептина, подвергали диализу в течение ночи при 4°C против 100-кратного объема PBS. Затем слитый белок осаждали при комнатной температуре, добавляя по каплям 4 М $(NH_4)_2SO_4$ (растворенный в воде) при постоянном перемешивании до достижения конечной концентрации $(NH_4)_2SO_4$ 1 М. После центрифугирования в течение 20 мин на скорости 17000 об/мин при комнатной температуре осадок, содержащий осажденный слитый белок PA#1(600)-hu-Leptin(W100/Q), растворяли в PBS и полученный раствор центрифугировали (13000 об/мин, 10 мин, комнатная температура) для удаления нерастворимых примесей.

Слитый белок PA#1(600)-hu-Leptin(W100Q) подвергали двукратному диализу против 5 л 20 мМ трис/HCl, pH 8,5, при 4°C, каждый раз на протяжении по меньшей мере 6 ч. Затем раствор белка подвергали анионообменной хроматографии с использованием колонки ResourceQ на 6 мл (GE Healthcare, Фрайбург, Германия), соединенной с системой очистки Akta (GE Healthcare, Фрайбург, Германия), с использованием 20 мМ трис/HCl, pH 8,5, в качестве рабочего буфера. Потом слитый белок элюировали с использованием градиента концентрации NaCl. Собирали элюированные фракции и проводили дальнейшую очистку посредством эксклюзионной хроматографии с использованием колонки Superdex 200 HR10/300 (GE Healthcare, Фрайбург, Германия), уравновешенной PBS.

Данная методика привела к получению однородного препарата белка без признаков агрегации с конечным выходом 0,8 мг на литр бактериальной культуры. Концентрацию белка определяли, измеряя поглощение при 280 нм с применением расчетного коэффициента экстинкции (Gill (1989), см. выше) $8605 \text{ M}^{-1} \text{ cm}^{-1}$. SDS-PAGE проводили с использованием системы с трис-буфером высокой молярности (Fling (1986), см. выше) (фиг. 13A). Элюированный белок подвергали двукратному диализу против 10000-кратного объема 10 мМ ацетата аммония, pH 5,5 и анализировали ESI-масс-спектрометрией на приборе maXis Q-TOF (Bruker Daltonics, Бремен, Германия) в режиме положительных ионов. Развернутый спектр PA#1(600)-hu-Leptin(W100Q) продемонстрировал массу 64249,53 Да (фиг. 13B), что по существу совпадает с расчетной массой данного слитого белка (64249,80 Да). Это ясно демонстрирует, что слитый белок PA#1(600)-hu-Leptin(W100Q), кодируемый генетически стабильной молекулой нуклеиновой кислоты с низкой повторяемостью по настоящему изобретению, может быть получен в *E. coli* в его интактной форме.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Молекула нуклеиновой кислоты, содержащая нуклеотидную последовательность, кодирующую полипептид, состоящий из пролина, аланина и серина или состоящий из пролина и аланина, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 300 нуклеотидов, где указанная нуклеотидная последовательность имеет показатель нуклеотидных повторов (Nucleotide Repeat Score, NRS) менее 1000, где указанный показатель нуклеотидных повторов (NRS) определен по следующей формуле:

$$NRS = \frac{\sum_{n=1}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}}$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности,

n представляет собой длину повтора в указанной нуклеотидной последовательности и

$f_i(n)$ представляет собой частоту указанного повтора длиной n ,

где при наличии более чем одного повтора длиной n $k(n)$ представляет собой число указанных разных последовательностей указанного повтора длиной n , в противном случае $k(n)$ для указанного повтора длиной n равен 1.

2. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина и аланина, предпочтительно где указанные пролиновые остатки составляют от 10 до 75% указанного кодируемого полипептида.

3. Молекула нуклеиновой кислоты по п.1, где указанный кодируемый полипептид состоит из пролина, аланина и серина, предпочтительно где указанные пролиновые остатки составляют от 4 до 40% указанного кодируемого полипептида.

4. Молекула нуклеиновой кислоты по любому из пп.1-3, где указанный показатель нуклеотидных повторов (NRS) составляет менее 100, менее 50 или менее 35.

5. Молекула нуклеиновой кислоты по любому из пп.1-4, где нуклеотидная последовательность указанной нуклеиновой кислоты имеет длину по меньшей мере 900 нуклеотидов.

6. Молекула нуклеиновой кислоты по любому из пп.1-5, имеющая повышенную генетическую стабильность.

7. Молекула нуклеиновой кислоты по любому из пп.1-6, где указанная нуклеотидная последовательность содержит указанные повторы, имеющие максимальную длину n_{max} , определенную по следующей формуле:

$$n_{max} \leq 17 + \frac{N_{tot}}{600},$$

где N_{tot} представляет собой длину указанной нуклеотидной последовательности.

8. Молекула нуклеиновой кислоты по любому из пп.1-7, где указанные повторы имеют максимальную длину от 14 до 55 нуклеотидов.

9. Молекула нуклеиновой кислоты по любому из пп.1-8, где указанный кодируемый полипептид содержит повторяющуюся аминокислотную последовательность с множеством аминокислотных повторов, где идентичны не более 9 расположенных друг за другом аминокислотных остатков и где указанный полипептид образует случайный клубок.

10. Молекула нуклеиновой кислоты по любому из пп.1, 2 и 4-9, выбранная из группы, состоящей из:

(а) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 192 и SEQ ID NO: 193;

(b) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, состоящую из SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172 и/или SEQ ID NO: 173;

(с) молекулы нуклеиновой кислоты, гибридизирующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (а) или (b);

(d) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 66,7% идентичную нуклеотидной последовательности, определенной в любом из (а), (b) и (с); и

(е) молекулы нуклеиновой кислоты, являющейся вырожденной в соответствии с генетическим кодом по отношению к нуклеотидной последовательности, определенной в (а) или (b).

11. Молекула нуклеиновой кислоты по любому из пп.1 и 3-9, выбранная из группы, состоящей из:

(а) молекулы нуклеиновой кислоты, содержащей по меньшей мере одну нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135,

SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 194 и SEQ ID NO: 195;

(b) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, выбранную из группы, состоящей из SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190 и SEQ ID NO: 191;

(c) молекулы нуклеиновой кислоты, гибридизирующейся в жестких условиях с комплементарной цепью нуклеотидной последовательности, определенной в (a) или (b);

(d) молекулы нуклеиновой кислоты, содержащей нуклеотидную последовательность, по меньшей мере на 56% идентичную нуклеотидной последовательности, определенной в любом из (a), (b) и (c);

(e) молекулы нуклеиновой кислоты, являющейся вырожденной в соответствии с генетическим кодом по отношению к нуклеотидной последовательности, определенной в (a) или (b).

12. Молекула нуклеиновой кислоты по любому из пп.1-11, функционально связанная в одной и той же рамке считывания с нуклеиновой кислотой, кодирующей биологически активный белок.

13. Молекула нуклеиновой кислоты по п.12, где указанный биологически активный белок представляет собой терапевтически эффективный белок.

14. Экспрессионный вектор, содержащий молекулу нуклеиновой кислоты по любому из пп.1-13.

15. Клетка-хозяин, содержащая молекулу нуклеиновой кислоты по любому из пп.1-13 или вектор по п.14.

16. Способ получения молекулы нуклеиновой кислоты по любому из пп.1-13, включающий культивирование клетки-хозяина по п.15 и выделение полученной молекулы нуклеиновой кислоты.

17. Способ получения вектора по п.14, включающий культивирование клетки-хозяина по п.15 и выделение полученного вектора.

18. Способ получения полипептида, кодируемого молекулой нуклеиновой кислоты по любому из пп.1-13, включающий культивирование клетки-хозяина по п.15.

19. Способ по п.18, дополнительно включающий выделение полученного полипептида.

20. Способ получения конъюгата лекарственного средства, содержащего полипептид, кодируемый молекулой нуклеиновой кислоты по любому из пп.1-11, и лекарственное средство, представляющее собой (1) биологически активный белок, и/или (2) низкомолекулярное соединение, и/или (3) углевод, включающий культивирование клетки-хозяина по п.15, выделение полученного полипептида и конъюгирование полученного пептида с (1) биологически активным белком, и/или (2) низкомолекулярным соединением, и/или (3) углеводом.

21. Способ по п.20, где указанный биологически активный белок представляет собой терапевтически эффективный белок.

22. Способ получения конъюгата по п.20 или 21, где указанный биологически активный белок выбран из группы, состоящей из связывающего белка, фрагмента антитела, цитокина, фактора роста, гормона, фермента, белковой вакцины, пептидной вакцины, пептида, состоящего из 50 или менее аминокислотных остатков, или пептидомиметика.

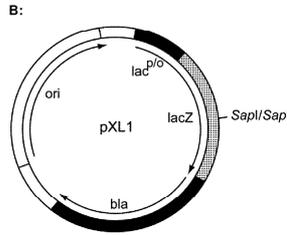
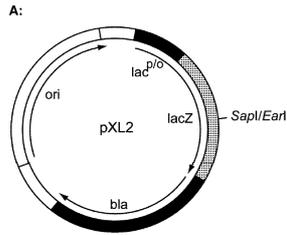
23. Способ получения конъюгата по п.22, где указанный связывающий белок выбран из группы, состоящей из антител, Fab-фрагментов, Fab'-фрагментов, F(ab')₂-фрагментов, одноцепочечных переменных фрагментов (scFv), (одно)доменных антител, выделенных переменных областей антител (областей VL и/или VH), CDR (гиперпеременных участков), иммуноглобулиновых доменов, пептидомиметиков, имеющих происхождение от CDR, лектинов, белковых каркасов, фибронектиновых доменов, тенасциновых доменов, доменов белка А, доменов SH3, доменов анкириновых повторов и липокалинов.

24. Способ получения конъюгата по любому из пп.20-22, где указанный биологически активный белок выбран из группы, состоящей из антагониста рецептора интерлейкина-1, лептина, кислой сфингомиелиназы, аденозиндезаминазы, агалсидазы-альфа, альфа-1-антитрипсина, предсердного натрийуретического пептида альфа, альфа-галактозидазы, альфа-глюкозидазы, альфа-N-ацетилглюкозаминидазы, алтеплазы, амедилазы, амилина, аналога амилина, пептидного ингибитора слияния ВИЧ, аргининдезаминазы, аспарагиназы, фактора VIII с делецией домена В, морфогенетического костного белка, антагониста брадикинина, натрийуретического пептида В-типа, буганина, гормона роста, хорионического гонадотропина, антагониста рецептора CD3, антагониста CD19, антагониста CD20, антагониста CD40, антагониста CD40L, церебролизидсульфатазы, фактора свертывания крови VIIa, фактора свертывания крови XIII, фактора свертывания крови IX, фактора свертывания крови X, ингибитора компонента комплемента C3, антагониста компонента комплемента 5a, С-пептида, антагониста CTLA-4 (цитотоксический Т-лимфоцитарный антиген 4), натрийуретического пептида С-типа, дефензина, дезоксирибонуклеазы I, антагониста рецептора EGFR (рецептор эпидермального фактора роста), эпидермального фактора роста, эритропоэтина, эксендина-4, эвринового пептида 1, антагониста рецептора FcγIIb, фактора роста фибробластов 21, фолликулстимулирующего гормона, желудочного ингибирующего полипептида (GIP),

аналога GIP, глюкагона, агониста рецепторов глюкагона, глюкагоноподобного пептида 1 (GLP-1), аналога GLP-1, глюкагоноподобного пептида 2 (GLP-2), аналога GLP-2, гонадорелина, агониста гонадотропин-рилизинг гормона, антагониста гонадотропин-рилизинг гормона, gp120, gp160, гранулоцитарного колониестимулирующего фактора (G-CSF), гранулоцитарно-макрофагального колониестимулирующего фактора (GM-CSF), грелина, аналога грелина, гормона роста, рилизинг-гормона гормона роста, гематида, фактора роста гепатоцитов, антагониста рецептора фактора роста гепатоцитов (HGFR), антагониста гепцидина, миметика гепцидина, антагониста рецептора Her2/neu, гистрелина, гирудина, антагониста hsp70, гуманина, гиалуронидазы, гидролитического лизосомального глюкоцереброзидспецифического фермента, идуронат-2-сульфатазы, антагонистов IgE, инсулина, аналога инсулина, инсулиноподобного фактора роста 1, инсулиноподобного фактора роста 2, интерферона-альфа, антагониста интерферона-альфа, суперагониста интерферона-альфа, интерферона-альфа-п3, интерферона-бета, интерферона-гамма, интерферона-лямбда, интерферона-тау, интерлейкина, слитого белка интерлейкина-2, антагониста альфа-субъединицы рецептора интерлейкина-22 (IL-22га), ирисина, белка, ассоциированного с островковым неогенезом (islet neogenesis associated protein), фактора роста кератиноцитов, антагонистов ионных каналов Kv1.3, лантипептида, липазы, лютеинизирующего гормона, лутропина альфа, лизоцифина, маннозидазы, N-ацетилгалактозамин-6-сульфатазы, N-ацетилглюкозаминидазы, нейтрофильного желатиназоассоциированного липокалина, октреотида, ω -конотоксина, ингибитора комплемента *Ornithodoros moubata*, остеогенного белка 1, остеопротегерина, оксалатдекарбоксилазы, P128, паратиреоидного гормона, филломера, антагониста PD-1 (белок программируемой клеточной гибели-1), антагониста PDGF (фактор роста тромбоцитов), фенилаланинаммиазидазы, тромбоцитарного фактора роста, проинсулина, белка C, релаксина, аналога релаксина, секретина, RGD-пептида (трипептид аргинин-глицин-аспарагиновая кислота), рибонуклеазы, сенреботазы, ингибитора сериновых протеаз, растворимого рецептора комплемента 1-го типа, растворимого рецептора DCC (рецептор, обнаруживаемый при колоректальном раке), растворимого рецептора TAC1 (трансмембранный активатор и партнер кальциевого модулятора и лиганда циклофилина), растворимого рецептора фактора некроза опухоли I (sTNF-RI), растворимого рецептора фактора некроза опухоли II (sTNF-II), растворимого рецептора VEGF (фактор роста эндотелия сосудов) Flt-1, растворимого рецептора Fc γ IIb, соматостатина, аналога соматостатина, стрептокиназы, лиганда T-клеточного рецептора, тенектеплазы, терипаратида, тромбомодулина-альфа, тимозина-альфа-1, ингибитора Toll-подобных рецепторов, фактора некроза опухоли (TNF α), антагониста фактора некроза опухоли α , уриказы, вазоактивного кишечного пептида, вазопрессина, аналога вазопрессина, антагониста VEGF, фактора фон Виллебранда.

25. Способ получения конъюгата по любому из пп.20-24, где указанное низкомолекулярное соединение выбрано из группы, состоящей из ингибиторов ангиогенеза, противоаллергических лекарственных средств, противорвотных лекарственных средств, антидепрессантов, антигипертензивных лекарственных средств, противовоспалительных лекарственных средств, противомикробных лекарственных средств, противомикробных лекарственных средств, антипсихотических лекарственных средств, антипролиферативных (цитотоксических и цитостатических) лекарственных средств, антагонистов кальция и других лекарственных средств, действующих на органы кровообращения, холинергических агонистов, лекарственных средств, действующих на центральную нервную систему, лекарственных средств, действующих на дыхательную систему, гормонов, стероидов, поликетидов, углеводов, олигосахаридов, нуклеиновых кислот, производных нуклеиновых кислот, анти-смысловых нуклеиновых кислот, малых интерферирующих РНК (siRNA), ингибиторов микро-РНК (miR), миметиков микро-РНК, ДНК-аптамеров и РНК-аптамеров.

26. Способ получения конъюгата по п.20, где указанный углевод выбран из группы, состоящей из следующего: О-антиген *E. coli* O 121, производные О-антигена *E. coli* O 121, Man9 из gp120 HIV-1, О-полисахариды *Shigella flexneri* 2a, капсульный полисахарид 5 *Staphylococcus aureus*, капсульный полисахарид 8 *Staphylococcus aureus*, опухлеассоциированные углеводные антигены (ТАСА) (такие как Tn-антигены (например, α -GalNAc-O-Ser/Thr), сиалиновые Tn-антигены (например, NeuAca(2,6)-GalNAc α -O-Ser/Thr), антиген Томсена-Фриденрайха (Gal β 1-3GalNAc α 1), LewisY (например, Fuc α (1,2)-Gal β (1,4)-[Fuc α (1,3)]-GalNAc), сиалил-LewisX и сиалил-LewisA, LewisX (стадиеспецифический эмбриональный антиген-1/SSEA-1), антиген GloboH (например, Fuc α (1,2)-Gal β (1,3)-GalNAc β (1,3)-Gal α (1,4)-Gal β (1,4)-Glc), T-антиген (например, Gal β (1,3)-GalNAc α -O-Ser/Thr), гликофинголипидный стадиеспецифический эмбриональный антиген-3 (SSEA-3), сиалосодержащие гликофинголипиды, ганглиозид GD2, GD3, ганглиозид GM2, ганглиозид фукозил-GM и ганглиозид Neu5GcGM3).



C:

```

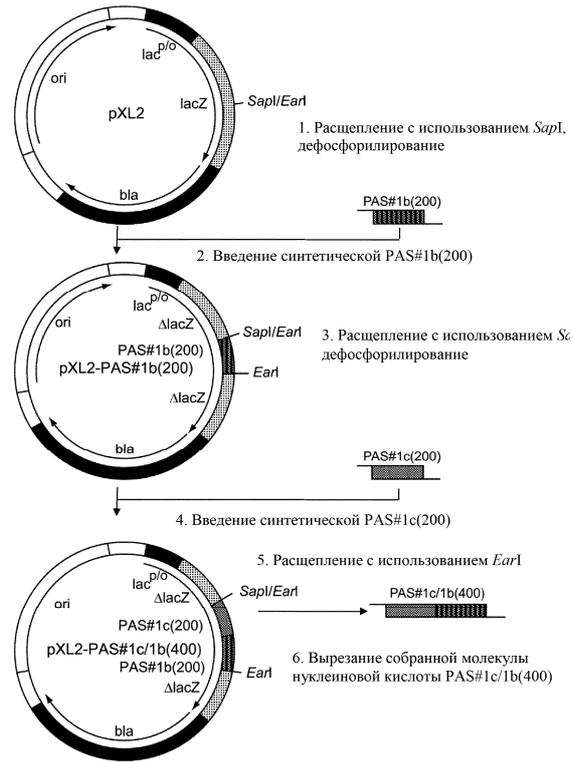
                    SapI
5'-TACGCCAAGCTTGGCTCTTCTCCAGCCCTGCCGCACCTGCCGCCGATCACCTGCCGCA
|||||
3'-ATCGGTTTCGAACCGAGAAGACGGTCCGGACGGCGTGGACCGGGCGTAGTGGACCGCGT
                    EarI
    
```

D:

```

                    SapI
TACGCCAAGCTTGGCTCTTCTCCAGCCCTGCCGCACCTGCCGCCGATCACCTGCCGCA
|||||
ATCGGTTTCGAACCGAGAAGACGGTCCGGACGGCGTGGACCGGGCGTAGTGGACCGCGT
AlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCTGCACCTTCCGCCCGGCTGCATCTCCTGCCGCACCCGCGCCTGCCAGCCAGCTGCA
GGACGTGGAAGGCGGGCCGACGTAGAGGACGGCGTGGCGCGGACGGTCCGGTCCGACGT
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCTGCCCAAGTGGCCAGCAGCATCCCTGCCGCGCCTGCCCGGCTAGTCCAGCGGCC
GGACGGGTTTCAAGCGTCTGTAGGGGACGGCGCGGACGGGGCGGATCAGGTCCGCCGG
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCAGTCCATCTGCACCAAGCTGTAGCCCTGCTGCACCAAGCTCCTGCTTCTCCCGCAGCC
GGTCGAGGTAGACGTGGTGCACGATCGGGACGACGTGGTGCAGGACGAAGAGGGCGTCGG
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCAGCGCCTTCTGCTCCCGCAGCCTCACCTGCCGGCCCGGCACCAAGCATCTCCAGCGGCA
GGTCGCGGAAGACGAGGGCGTCCGAGTGGACGCCGGGCGGTGGTCTGAGAGGTCCCGGT
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCAGCACCTTCCGCCCGTGTGCTAGCCAGCAGCACCTGCCCGAGCCTCACCAAGCTGTCT
GGTCGTGGAAGCCGGGACGACGATCGGGTCTGCTGAGCAGCGTCCGAGTGGTCCGACGA
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCCGCTCCTAGTCCCGCGGCGCCTGCCCTGCTGCTCCTGCACCAAGCTTCCCGCAGCGGCA
GGGCGAGGATCACGGGCGCCGGAGCGGACGACGAGGACGTGGTCCGAGCGGTCCCGGT
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCGGCTCCTTCCGGCGCGGCTGCTTACCAGCAGCACCTGCTCCAGCGTCCCGCAGCGGCC
GGCCGAGGAAGCCCGCGCCGACGAAGTGGTCTGCTGGACGAGGTCCGAGGGTCCCGCGG
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCTGCTCCAAGTGTCCGGTGCATCGCCTGCCGCTCCTGCTCCTGCATCCCGAGCTGCT
GGACGAGGTTTCAAGGCGGACGTAGCGGACGGCGAGGACGAGGACGTAGGGTCCGACGA
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCAGCACCAGCGCACCTGCCGCTCACCAAGCGGCGCCAGCACCCGCGAGCCCGCAGCGG
GGTCTGGTTCGCGTGGACGGCGGAGTGGTCCGCGCGGTCTGGGCGGTCCGGTCTGTCG
ProAlaProSerAlaProAlaAlaSerProAlaAlaProAlaProAlaSerProAlaAla
CCTGCTCCATCCGCACCGGCCGCCAGAAGAGTAGAATTCAGTGGCC
|||||
GGACGAGGTAGGCGTGGCGCGCGGCTCTTCTCATCTTAAGTGACCGG
ProAlaProSerAlaProAlaAla EarI
    
```

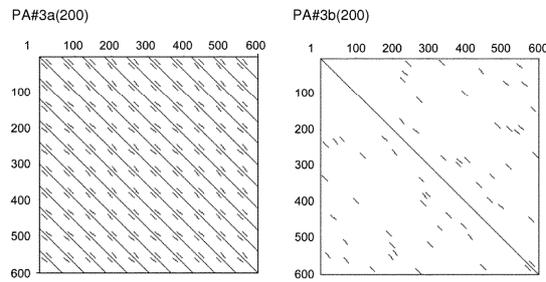
Е:



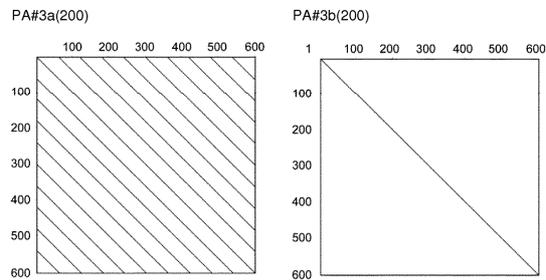
Фиг. 1

А:

Окно повтора 14:



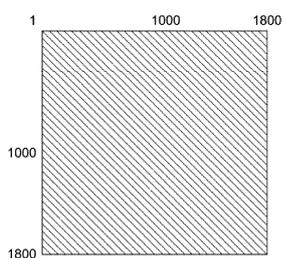
Окно повтора 15:



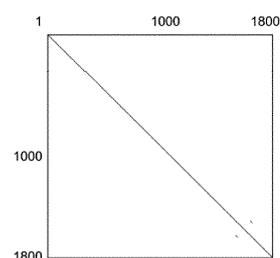
В:

Окно повтора 14:

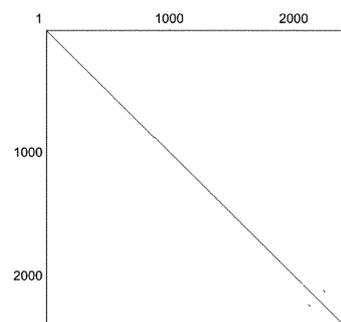
PAS#1a(600)



PAS#1f/1c/1b(600)

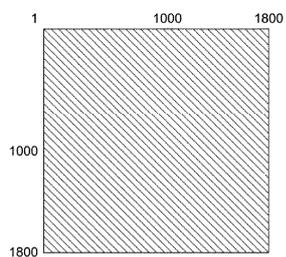


PAS#1d/1f/1c/1b(800)

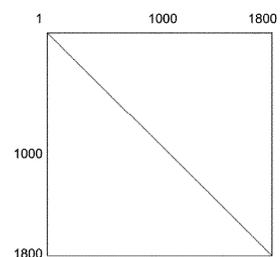


Окно повтора 15:

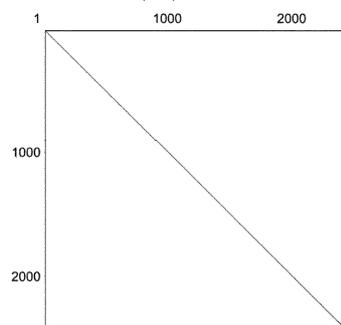
PAS#1a(600)



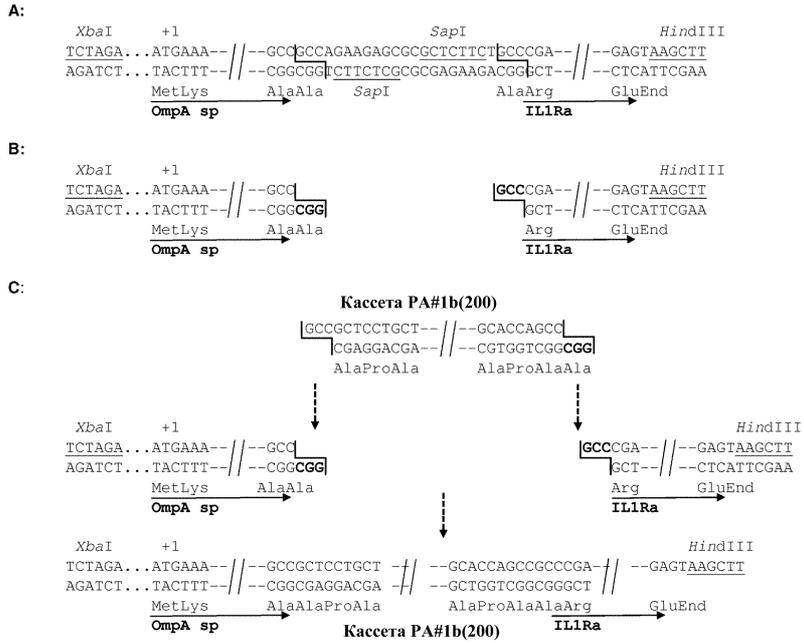
PAS1f/1c/1b(600)



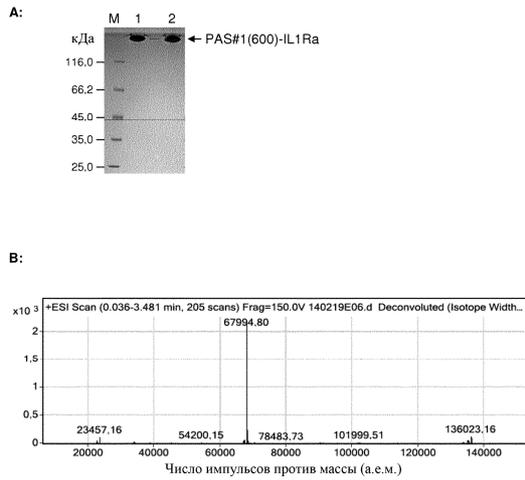
PAS#1d/1f/1c/1b(800)



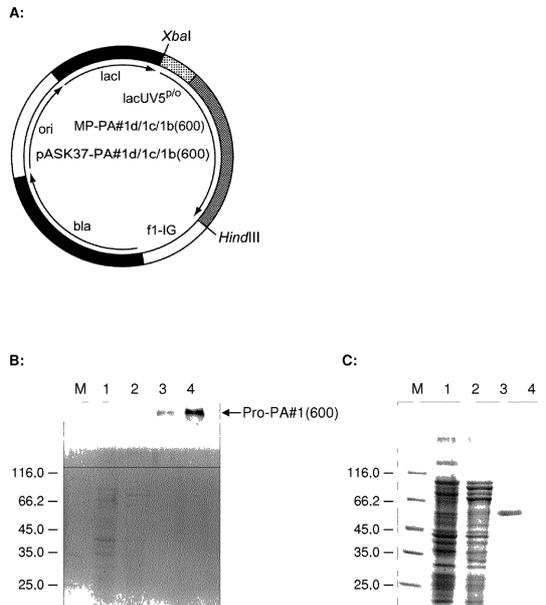
Фиг. 2



Фиг. 6



Фиг. 7

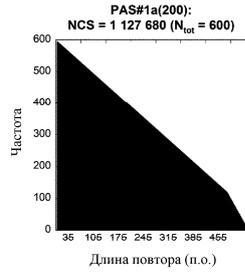


D:

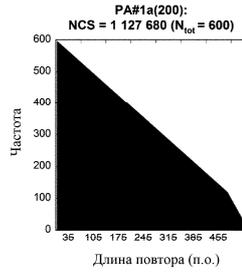


Фиг. 8

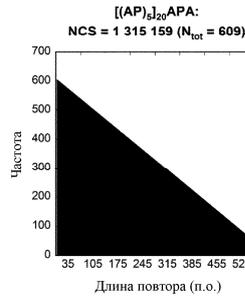
A:



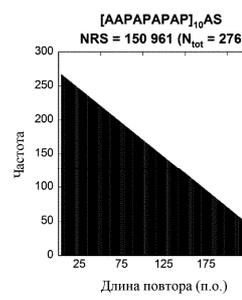
B:



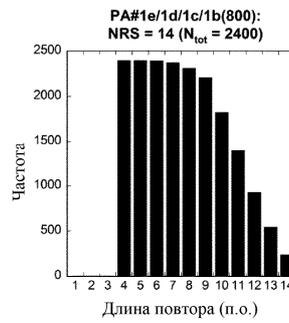
C:



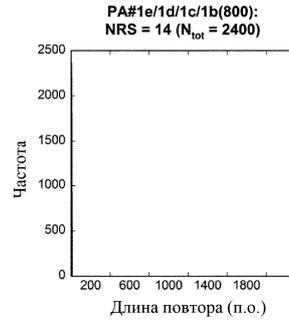
D:



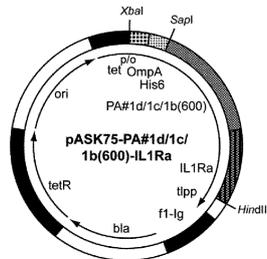
H:



I:



Фиг. 9

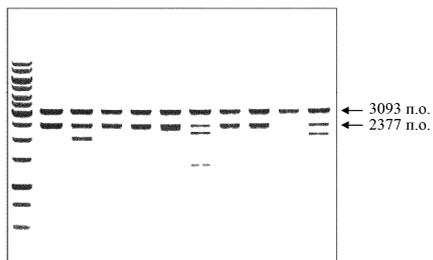


Фиг. 10

A:

pASK75-PA#1a(600)-IL1Ra

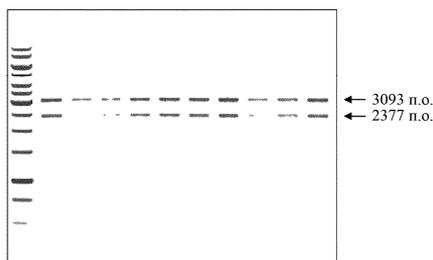
M 1 2 3 4 5 6 7 8 9 10



B:

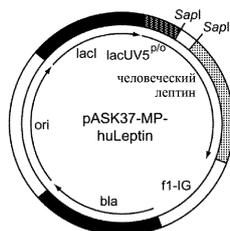
pASK75-PA#1d/1c/1b(600)-IL1Ra

M 1 2 3 4 5 6 7 8 9 10

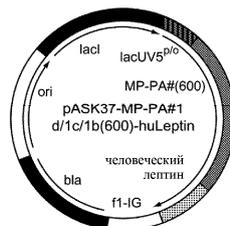


Фиг. 11

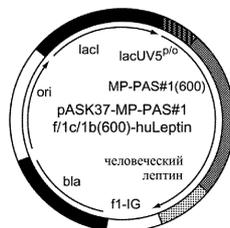
A:



B:

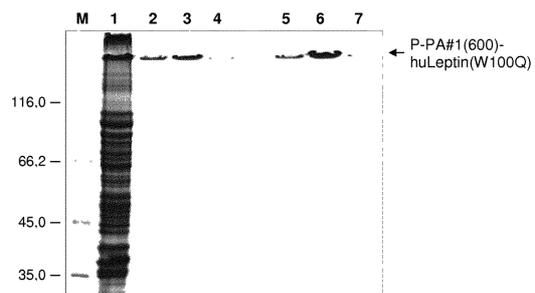


C:

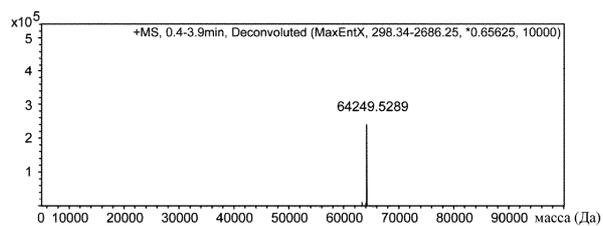


Фиг. 12

A:



B:



Фиг. 13

